Integrated Global Radiosonde Archive

Product Description

National Centers for Environmental Information (NCEI)

Prepared by

Imke Durre, Bruce Hundermark, Xungang Yin, Bryant Korzeniewski

January 2023

Contact: ncei.igra@noaa.gov

1. Intent of This Document and POC

1a) This document is intended to provide a basic understanding of the product (or dataset) including points of contact, sources used to develop the product, quality control practices, and steps to take to validate the accuracy of the product. A technical supplement and additional references are provided for users interested in further technical details.

1b) Technical point of contact for this dataset:

Names: Bruce Hundermark and Imke Durre

Email: ncei.igra@noaa.gov

2. Product Overview and Intended Uses

First released in 2004, the Integrated Global Radiosonde Archive (IGRA) is a collection of historical and near-real-time, land- and ship-based radiosonde observations from around the globe dating back to 1905. Version 2.2, released in February 2023, incorporates two new data streams of soundings received in the so-called Binary Universal Form for the Representation of meteorological data (BUFR), an international data format standard that is gradually being implemented at an increasing number of observing sites. Thanks to the addition of these data sets, the records at stations that have switched to only submitting soundings in BUFR can again be updated in IGRA, and the continuity of updates at other stations that undergo the transition to BUFR in the future is ensured.

The IGRA archive consists of five components: quality-assured individual soundings, monthly means, sounding-derived humidity and stability parameters, station history information, and the homogeneity-adjusted time series of temperature anomalies that constitute the Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC).

All of these components are updated on a regular basis and are made available in ASCII format. The sounding data and derived parameters are used for a wide variety of applications and analyses that require vertical profiles of atmospheric temperature, humidity, or wind in the troposphere and lower stratosphere. They aid in the verification of observations from other instruments and platforms, assist in the calibration of satellite observations, and serve as input to, or as a validation tool for, retrospective reanalyses, air pollution models, and other atmospheric models for both research and operational purposes. Although IGRA have not been adjusted for biases introduced by changes in instrumentation, observing practices, and station location over time, they nevertheless also provide important information about atmospheric conditions during weather events since the early 20th century.

3. Data Field Description

The five components of IGRA contain a large number of data and metadata variables. The IGRA sounding data include the primary variables of temperature, humidity expressed as either RH or dewpoint depression or both, and wind direction and speed. The vertical location of a data level within the atmosphere is always identified by pressure, geopotential height, or both. In some cases, elapsed time since launch is also available.

For specifics on file formats, measurement units, and precision of variables within each IGRA component, the user is referred to the readme files and format documentation files that accompany the data. The IGRA sounding data, derived parameters, monthly means, and station history information can be accessed at https://ncei.noaa.gov/data/integrated-global-radiosonde-archive/ and ftp://ftp.ncei.noaa.gov/pub/data/igra. All RATPAC-related files can be found at https://www.ncei.noaa.gov/data/ratpac/ and ftp://ftp.ncei.noaa.gov/pub/data/ratpac.
In general, observing sites are distributed unevenly across global land areas as well as some ships. In the sounding data, observations are available at standard and variable pressure levels, fixed and variable-height wind levels, and the surface and tropopause. Data from the higher-resolution BUFR profiles have been down-sampled to match this historically standard vertical resolution. Variables include pressure, temperature, geopotential height, relative humidity, dew point depression, wind direction and speed, and elapsed time since launch.

Soundings and sounding-derived parameters are typically available once or twice daily, although higher-frequency observations can also be found in the dataset. The spatial coverage, temporal gaps in the data, and the vertical resolution and extent of the profiles vary over time, as documented in Durre et al. (2006, 2018).

4. Data Sources and Processing *(several paragraphs)*

IGRA contains data from 42 historical and real-time data sources of varying, and often overlapping, spatial coverage and temporal extent. The creation of the integrated IGRA sounding data from these individual sources involves a complex set of steps that, broadly speaking, include converting each data source into a standard format, defining the set of geographical locations that are represented within the collection of data (i.e., IGRA stations), merging the data available for each of these IGRA stations into a single time series of soundings, applying a comprehensive suite of quality control procedures that remove gross errors from the data, and compositing the records from some collocated, or nearly collocated, sites. Additional details, as well as the creation of the accompanying station history information, can be found in the accompanying Supplement as well as in Durre et al. (2018). The sounding-derived parameters and monthly means are calculated from the integrated sounding data (Durre et al. 2006, 2018; Durre and Yin 2008, 2011), and the RATPAC time series are, from 1997 onward, derived from the IGRA monthly means (Free et al. 2005; Durre et al. 2016).

Although the IGRA processing system has been improved through extensive parallelization of processes in order to accommodate the much larger volume of the BUFR data, the science behind, and output from, all of these steps remain the same in IGRA 2.2 as in IGRA 2.0. The primary change in IGRA v2.2 compared to v2.0 is that the NCEI/NCEP GTS data source, which provides most of the real-time updates to IGRA, is now augmented with low-resolution versions of soundings available in two new BUFR data sources in order to be able to continue to update data as stations migrate to transmitting in BUFR. The two new data streams originate from the National Weather Service's Telecommunications Gateway (NWS TG, beginning in 2017) and the European Centre for Medium-Range Weather Forecasts (ECMWF, beginning in 2014) and are archived at NCEI. Because the vertical resolution of profiles transmitted in BUFR is often much higher than that of typical IGRA soundings, the BUFR profiles are first downsampled to match the spacing of standard and significant levels that is common in IGRA. Next, for each

IGRA station, the downsampled soundings are used to fill in data on days on which no observations are available in the NCEI/NCEP GTS data source.The augmented data source then enters the IGRA processing system like all other sources. Additional details on the processing of the BUFR data streams can be found in Sections 1.4 and 3.2 of the Product Description Supplement.

5. Quality Control *(one or two paragraphs with one supporting table or figure and references)*

The IGRA 2.2 QA system is identical to the one in the IGRA 2.0 version which is largely based on the QA procedures in the IGRA 1.0 system (Durre et al. 2006, 2008, 2018). It consists of a deliberate sequence of specialized algorithms, each of which 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.

The procedures in the IGRA QA system can be grouped into eight categories: fundamental "sanity" checks, checks on the plausibility and temporal consistency of surface elevation, internal consistency checks, checks for the repetition of values, checks for gross position errors in ship tracks, climatologically-based checks, checks on the vertical and temporal consistency of temperature, and data completeness checks. See the IGRA Product Description Supplement and Durre et al. (2018) for specifics on the individual QA procedures.

6. Validation *(one or two paragraphs with one supporting table or figure and references)*

The IGRA 2.2 processing system was run non-operationally in parallel to the then-operational IGRA 2.0 system for approximately two years. During this time, extensive comparisons have been made between the operational v2.0 output and the non-operational v2.2 output, and adjustments to the system were made in the rare cases in which this was deemed necessary. Comparisons focused on the respective run times, output data volume in terms of directory sizes and file counts, and spot checks of soundings created by the two systems. In addition, a spot check of several soundings over the US revealed no differences between the non-operational v2.2 versions running in parallel, one on the production server and one on the development server.

During the development of the process for decoding and incorporating the soundings received in BUFR, several forms of validation were used to verify that new processing steps yielded the desired outcomes. The two most important of these are presented briefly here.

1. Verification of downsampling algorithm: Comparisons of selected TAC-based soundings with BUFR-based soundings that have been downsampled from the native one-second vertical resolution to mandatory and significant pressure levels. Figure 1 shows an example of such a comparison which demonstrates that the downsampled BUFR-based sounding matches the corresponding TAC sounding that was also received for this sample station.

Figure 2 shows a histogram, for each standard pressure level, the percent of available soundings in the year 2018 that included that particular level. The figure illustrates that downsampled BUFR-based soundings contain approximately the same number of standard levels as TAC soundings, suggesting that users of mandatory-level data should not experience a significant change in completeness due to the introduction of the additional data streams. In fact, the BUFR soundings are slightly more likely to extend above 100 hPa. For example, the percentage of soundings with a 100-hPa level is 75% among TAC reports compared to 80% among BUFR reports. The likely reason for the slightly better completeness of the BUFR reports in the stratosphere is that, unlike BUFR messages, TAC reports are split into separate reports for levels below and above 100 hPA, sometimes leading to the above-100-hPa portion to be lost during transmission or ingest.

2. Station count and distribution: a comparison of the numbers of TAC-only, BUFR-only, and TAC-and-BUFR transmitting stations to expectations based on reports from the WMO. Figure 3 shows time series of the annual number of stations with data in IGRA v2.0 and v2.2, while Figure 4 displays a map of the 119 stations that, as of late 2022, had fully migrated to BUFR and were therefore no longer being updated in IGRA v2.0. There is a general decrease in the number of upper-air stations that is reflected in both IGRA v2.0 and v2.2 due to some stations closing for lack of funding and other reasons. However, the decline in stations is less pronounced in v2.2 than in v2.0 due to the fact that the stations that have fully migrated to BUFR can again be updated in v2.2. The geographical distribution of these stations is consistent with the countries that are known to have switched to transmitting only in BUFR, e.g., China and Spain.

Figure 1: Sample Vertical Profile from TAC and BUFR Reports

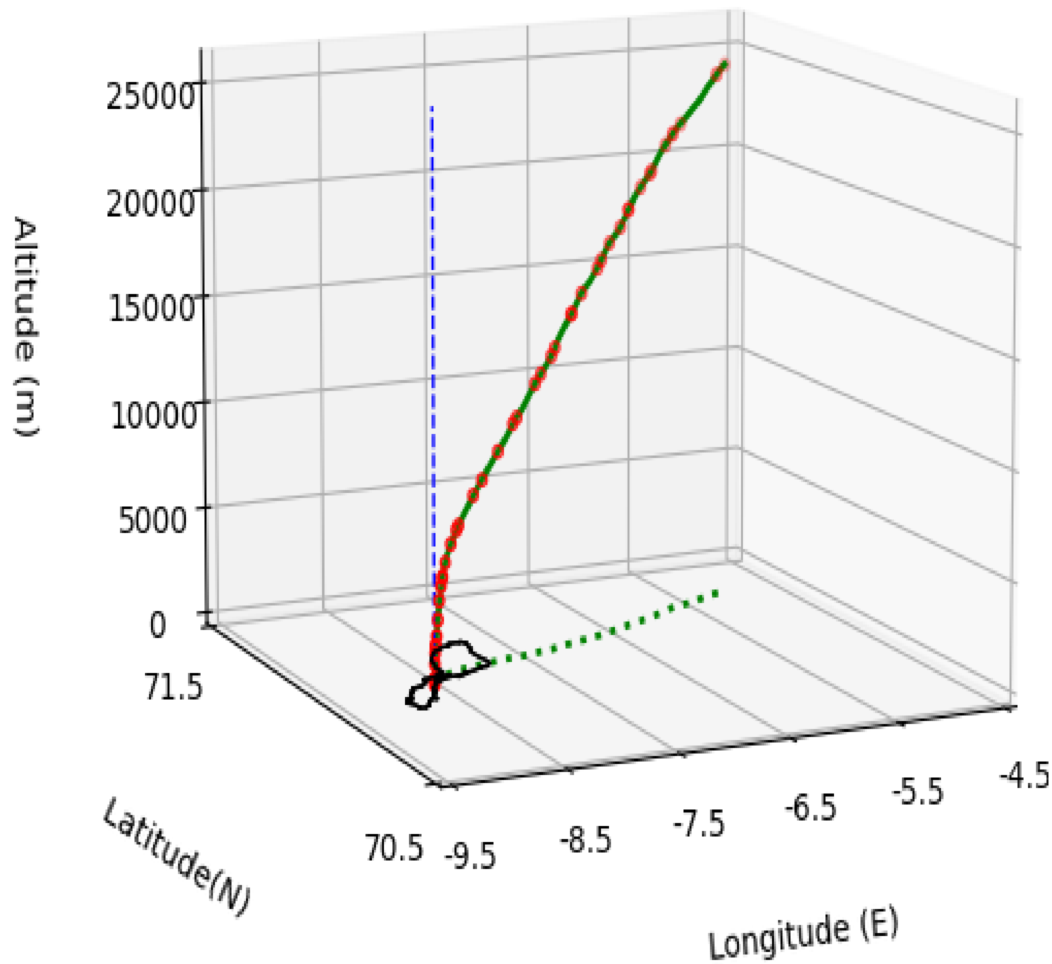


Figure 2: Percent of Available Soundings at Each Standard Pressure Level in TAC and Down-Sampled

BUFR Reports During 2018

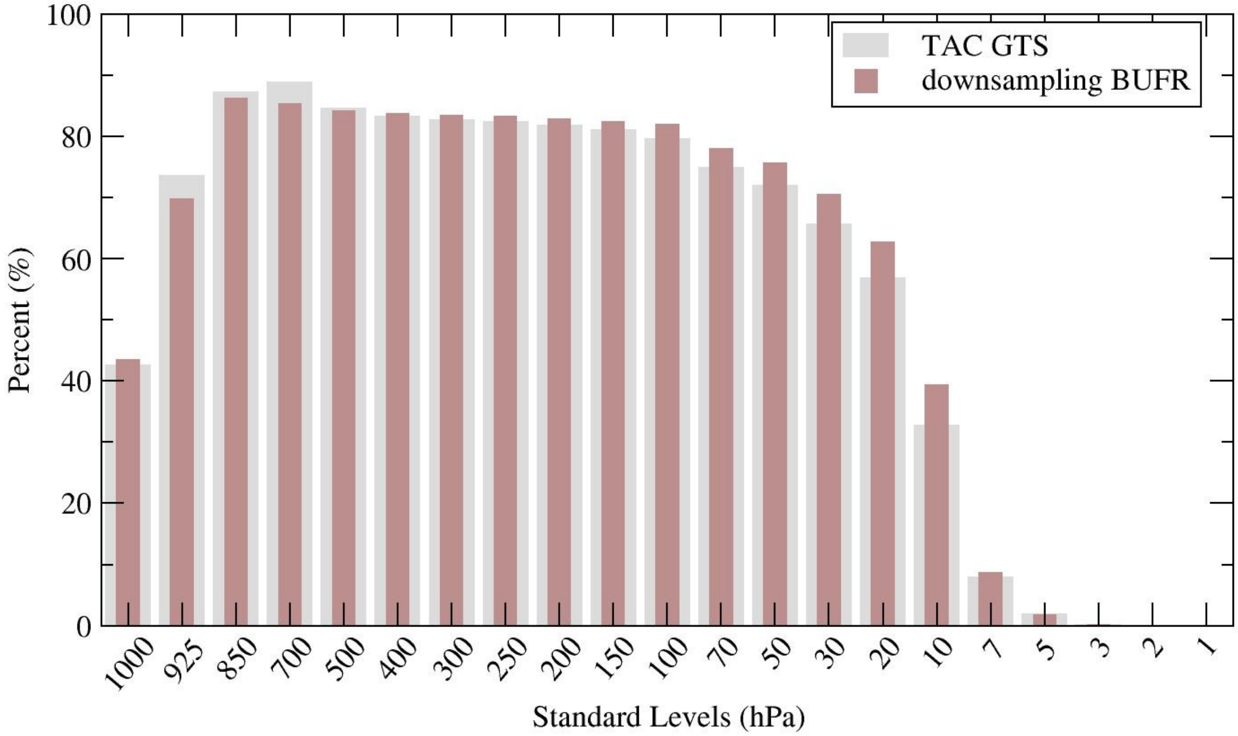


Figure 3: Annual Number of stations in IGRA v2.2 and v2.0 Since 2014

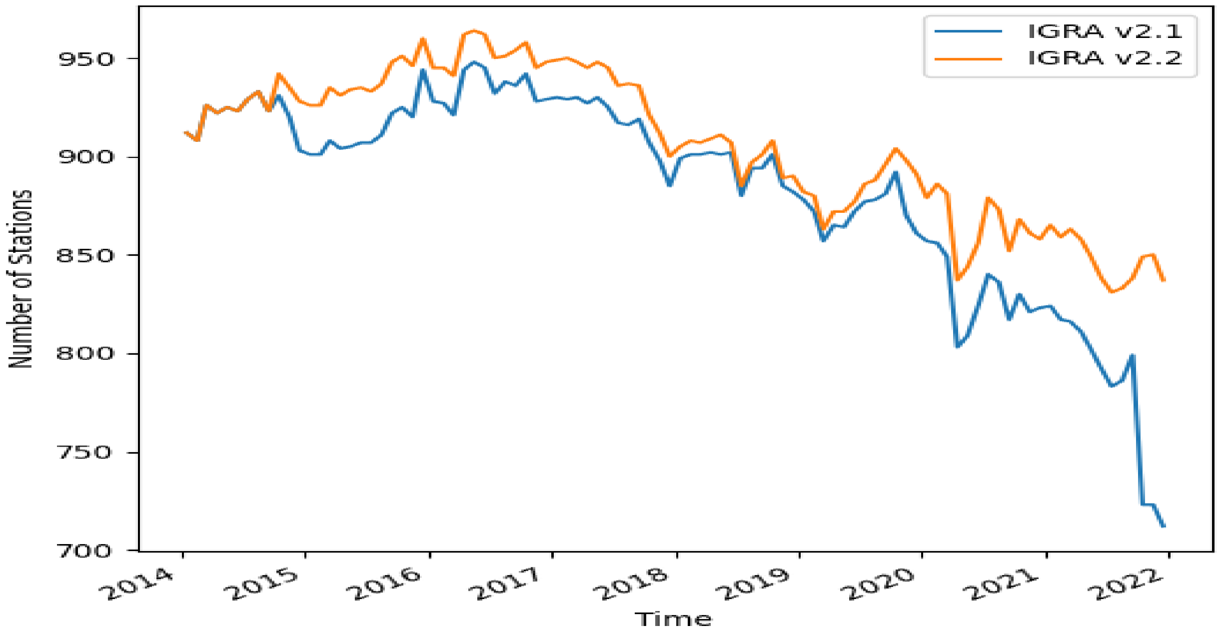
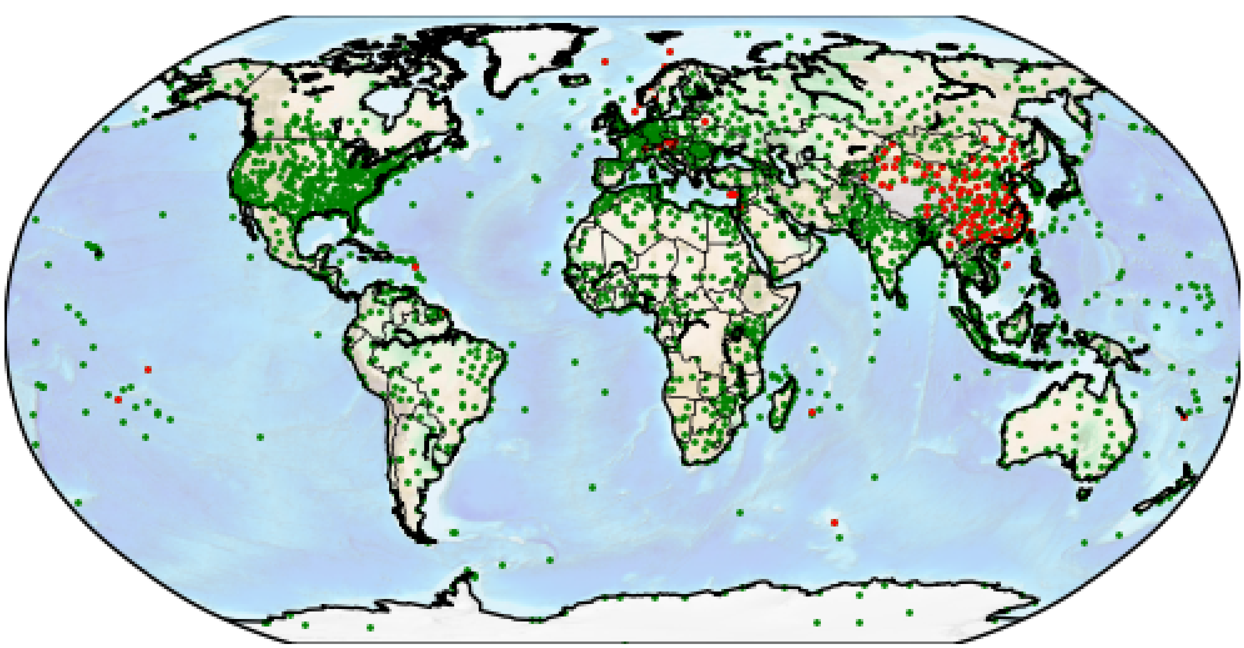


Figure 4: Map of Stations Fully Migrated to BUFR and Recovered in IGRA v2.2



7. Additional Information

Dataset Citation

Cite this dataset as: Durre, Imke; Xungang, Yin; Vose, Russell S.; Applequist, Scott; Arnfield, Jeff. (2016) Integrated Global Radiosonde Archive (IGRA), Version 2. [indicate subset and exact version used]. NOAA National Centers for Environmental Information. DOI:10.7289/V5X63K0Q [access date].

Supplementary Information

Additional technical information is available in a Product Description Supplement which will be available at https://www.ncei.noaa.gov/pub/data/igra/igra2-product-description-supplement-rev00.pdf upon release of IGRA v2.2. Operators of the IGRA process at NCEI can find guidance on processing steps in the Operations Document and Operations Document Supplement within NCEI's internal Gitlab repository.

References

Durre, I., and X. Yin, 2008: Enhanced radiosonde data for studies of vertical structure. Bulletin of the American Meteorological Society, 89, 1257-1262.

Durre, I., and X. Yin, 2011: Enhancements of the dataset of sounding parameters derived from the Integrated Global Radiosonde Archive. 23rd Conference on Climate Variability and Change, Seattle, WA, 25 January 2011. [Available online at https://ams.confex.com/ams/91Annual/webprogram/Paper179437.html.]

Durre, I., R. S. Vose, and D. B. Wuertz, 2006: Overview of the Integrated Global Radiosonde Archive. Journal of Climate, 19, 53-68.

Durre, I., R. S. Vose, and D. B. Wuertz, 2008: Robust automated quality assurance of radiosonde temperatures. Journal of Applied Meteorology and Climatology, 47, 2081-2095.

Free, M., D. J. Seidel, J. K. Angell, I. Durre, and T. C. Peterson, 2005: Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): A new dataset of large-area anomaly time series. Journal of Geophysical Research, doi: 10.1029/2005JD006169.

8. Dataset and Document Revision History

Rev 0 – 24 Jan 2023 - This is a new document for IGRA v2.2.