

## Enhanced Data Coverage in Version 2 of the Integrated Global Radiosonde Archive

Imke Durre  
NOAA National Climatic Data Center

Xungang Yin  
STG, Inc.

Russell S. Vose  
NOAA National Climatic Data Center

The Integrated Global Radiosonde Archive (IGRA) is a collection of historical and real-time radiosonde and pilot balloon observations from around the world. It is maintained by the United States National Oceanic and Atmospheric Administration's (NOAA's) National Climatic Data Center (NCDC). Version 1 of IGRA was released in 2005 and consists of historical radiosonde and pilot balloon observations from 1538 stations worldwide. Variables include pressure, temperature, geopotential height, dew point depression, wind direction, and wind speed at standard, surface, tropopause, and significant levels. A compendium of data from 11 different sources, IGRA is the largest dataset of quality-assured radiosonde observations freely available to all users and is updated on a daily basis.

However, IGRA 1 has several limitations of relevance to reanalysis efforts. It is currently limited to fixed stations on land, gives preference to thermodynamic data over pilot balloon observations, and is limited in its spatial coverage before 1963. Consequently, NCDC is making an effort to produce a new version of IGRA, IGRA 2, with the primary goal of improving the temporal and spatial coverage of the database. Improved coverage will be achieved in two ways: (1) by including sources of data that were not included in IGRA 1 and (2) by modifying the IGRA processing system to include as many pilot balloon and ship observations as possible along with the land-based radiosonde soundings.

The most resource-intensive step in such an undertaking is the identification, acquisition, and preprocessing of additional data sources. A search of the literature on past reanalysis projects and the archives of both NCDC and the National Center for Atmospheric Research (NCAR) yielded more than 70 potential data sets. In addition, NCDC acquired recently digitized data from the Climate Database Modernization Program (CDMP), bringing the total to more than 90 potential sources. Of these, 42 were reformatted into a common data format for further processing. The acquisition of another approximately 30 was postponed until future versions of IGRA because they were either not readily available or extremely small. The remainder were rejected because of exact duplication with more easily accessible data sets, inadequate documentation, or significant problems with the data files themselves.

The 42 reformatted data sources include 16 from CDMP, 5 from NCDC's archive, 3 from NCAR, 2 from Meteo-France, and 1 each from the U.S. Air Force 14th Weather Squadron, the British Antarctic Survey, the Institute for Atmospheric and Climate Science at ETH Zurich in Switzerland, the Japan Agency for Marine-Earth Science and Technology (JAMSTEC), and the National Geophysical Data Center. Of particular interest to the reanalysis community are NCAR's datasets of reports sent over the Global Telecommunications System (GTS) since 1958; pilot balloon observations for the United States between 1919 and 1960; CDMP's and Meteo-France's data for Africa, a region where historical observations have been particularly sparse; and the additional pre-1960s records supplied by various other sources. Considering all of these sources together, the earliest observations are from 1902, and the record extends up to the day before present.

All of the reformatted data sources are being further pre-processed, checked for duplicates, merged into one data set, and quality assured. While much of this processing can be

accomplished closely following the procedures developed for IGRA 1, two modifications to the merging process are being implemented to accommodate the additional data sets, a number of which consist of solely pilot balloon observations and/or provide records for locations and periods not covered by any other source. First, the algorithm for determining whether station records in one source match with any of the stations contained in other sources is being expanded, so that it does not solely rely on comparisons of mandatory pressure level data during overlapping periods of record. Second, the procedure for merging multiple sources of data into one station record, which currently gives preference to thermodynamic observations, is being modified such that it maximizes the number of both thermodynamic and wind levels in the merged records.

**Corresponding Author:**

**Name:** Imke Durre  
**Organization:** NOAA National Climatic Data Center  
**Address:** 151 Patton Avenue,  
Asheville, NC 28801  
United States