

# A dataset of completeness of radiosonde humidity observations based on the IGRA

António P. Ferreira (ap.ferreira@uvigo.es), Raquel Nieto (rnieto@uvigo.es) and Luis Gimeno (l.gimeno@uvigo.es)

Environmental Physics Laboratory | Faculty of Science - University of Vigo  
Campus As Lagoas, 32004 Ourense, Spain

=====

**Readme file** | António P. Ferreira | Last updated 2 August 2018

=====

-----

## TABLE OF CONTENTS:

-----

### [A. OVERVIEW](#)

### [B. SOURCE DATA](#)

### [C. SHORT DESCRIPTION OF DATA FILES](#)

#### [D1. FORMAT DESCRIPTION OF FILE: humetadata-yearly.txt](#)

#### [D2. FORMAT DESCRIPTION OF FILE: humetadata-ascent.zip](#)

#### [D3. FORMAT DESCRIPTION OF FILE: stations-hum-record.txt](#)

### [E. CONTACT FOR QUESTIONS](#)

-----

### [A. OVERVIEW](#)

This dataset describes the completeness of historical radiosonde humidity observations based on the sounding data compiled in the Integrated Global Radiosonde Archive (IGRA) Version 2, until the end of 2016. The stations in the IGRA 2 reporting a minimal amount of radiosonde observations (RAOB) – namely, RAOB in at least 5% of the annual observations taken for at least one year, including pilot balloon observations – were evaluated according to specified completeness parameters for every year in their period of record, regarding humidity measurements of either relative humidity or dewpoint-depression.

Humidity completeness for a radiosonde station and a year is defined by: the annual count of humidity soundings; the fraction of days in a year having humidity observations; the average vertical resolution of humidity data; the average atmospheric pressure at the highest level with humidity data; and the largest interval of days without humidity data. The Surface-to-500-hPa observations eligible for calculating precipitable water vapor – having sufficient vertical sampling between the surface and the 500-hPa level – are also evaluated. Completeness for an individual sounding is defined by the vertical resolution of humidity data, the height and pressure at the highest level with humidity data, and whether the sounding qualifies as a Surface-to 500-hPa humidity sounding.

The main sub data sets consist of 1) statistical metadata describing humidity completeness for each station and year within the station's period of record and 2) metadata describing humidity completeness in individual observations for the same stations, sounding by sounding. An additional data set 3) provides the observing periods for humidity and the corresponding amount of observations for each of the selected stations: 1723 stations out of the 2761 of IGRA 2.

-----

## B. SOURCE DATA

This dataset is derived from the following dataset:

Durre, I., Vose, R.S., Yin, X., Applequist, S., and Arnfield, J. (2016): Integrated Global Radiosonde Archive (IGRA) Version 2. [Sounding data for the full period of record]. NOAA National Centers for Environmental Information. DOI:10.7289/v5x63k0q

---

## C. SHORT DESCRIPTION OF DATA FILES

	File name _____	Content _____
(1)	humetadata-yearly.txt	Meta-data of humidity completeness on a yearly basis for each station
(2)	humetadata-ascent.zip	Meta-data of humidity completeness in individual observations from each station
(3)	stations-hum-record.txt	Information on the upper-air stations represented in files (1) and (2): IGRA ID codes, station names, countries, geographic coordinates, full periods or record until 2016, record periods for humidity and corresponding amounts of humidity observations

---

### D1. FORMAT DESCRIPTION OF FILE "humetadata-yearly.txt"

-----  
Note:  
-----

The file begins with a header record with the names of variables, followed by the data records. Stations appear in alphabetical order according to IGRA 2 station identifier codes; the station locations are given in the table described in section D3. For each station, the data describing the completeness of humidity observations are shown chronologically in one record per year until 2016. Years without source data are not shown. The file has 60175 data records for 1723 stations.

-----  
Data Record Format:  
-----

Variable	Columns	Type
STN_ID	1-11	Character
YEAR	13-16	Integer
-----		
SNDS	18-21	Integer
TEMP	23-26	Integer
HUMa	28-31	Integer
RESa	33-36	Integer
GAPa	38-41	Integer
FDYa	43-46	Integer
TOPP	48-51	Integer
HUMb	53-56	Integer
RESb	58-61	Integer
GAPb	63-66	Integer

These variables have the following definitions:

**STN\_ID** is the station's identification code used in IGRA 2.

**YEAR** is the year in the period of period to which the following metadata corresponds.

**SNDS** is the annual number of soundings, regardless of having RAOB or PILOT (wind-only) data.

**TEMP** is the annual number of soundings with RAOB data – i.e., having temperature data at any pressure levels.

**HUMa** is the annual number of RAOB soundings having any amount of humidity data besides temperature data.

**RESa** is the annual mean vertical resolution of humidity data [in decametres];  
-999 = missing (no humidity data in the entire year).

Note: the vertical resolution of a single sounding is defined as the geometric mean distance between all consecutive levels with humidity data, obtained from collocated pressure and temperature.

**GAPa** is the largest interval of time without humidity data within the year [in days];  
if  $HUMa = 0$ , then  $GAPa =$  number of days in the year.

**FDYa** is the fraction of days in a year having humidity data [in %].

**TOPP** is the annual geometric mean pressure at the top of humidity soundings, rounded to the nearest kPa [in hPa];  
-999 = missing (no humidity data).

**HUMb** is the annual number of Sfc-to-500-hPa humidity soundings, i.e., of RAOB soundings such that:

(i) Humidity data are not missing at the surface level and all standard pressure levels (other than 925 hPa) between the surface and the 500 hPa, or additional levels nearby, within a distance of less than 5% the height of a missing standard level (measured from the surface);

(ii) The distance between any two consecutive levels with humidity data between the surface level and the lowest level at a height larger than 1 km above the surface must be less than 1 km, unless the station's elevation is larger than 500 m.

**RESb** is the annual mean vertical resolution of Sfc-to-500-hPa humidity data [in decametres].  
-999 = missing (no Sfc-to-500-hPa humidity data).

Note: the vertical resolution of an individual Sfc-to-500-hPa humidity sounding is calculated as noted above for RESa but limited to pressure and temperature data from the surface up to 500 hPa. If missing, the upper limit can be replaced non-standard level near 500 hPa, according with the above definition of Sfc-to-500hPa humidity soundings.

**GAPb** is the largest interval of time without Sfc-to-500-hPa humidity data within the year [in days];  
if  $HUMb = 0$ , then  $GAPb =$  number of days in the year.

**FDYb** is the fraction of days in a year having Sfc-to500-hPa humidity data [in %].

---

## [D2. FORMAT DESCRIPTION OF FILE "humetadata-ascent.zip"](#)

-----  
**Notes:**  
-----

1. The ZIP file contains 1723 plain-text files, each one corresponding to a station among the selected IGRAs listed in the file described in Section D3. E.g.: "ACM00078861-humetadata.txt" refers to the station with IGRAs identification code ACM00078861.

2. Each individual file consists of a header record (with the names of variables) followed by the data records describing the completeness of humidity observations in every vertical sounding of the corresponding station: one record per sounding covering the station period of record, as given in IGRAs until December 31, 2016.

-----  
**Data Record Format:**  
-----

Variable	Columns	Type
LAUNCH_DATE	1-10	Character
HOUR	12-15	Character
GND_LAT	17-23	Real
GND_LONG	25-32	Real
-----		
RAOB	36-36	Integer
RESa	39-42	Integer
RESb	45-48	Integer
TOPP	51-54	Integer
TOPZ	57-60	Integer

These variables have the following definitions:

**LAUNCH\_DATE** is the date of the balloon launch in the format YYYY-MM-DD.

**HOUR** is the nominal hour (UT) of the sounding in the format hhZ, with valid values of hh being 00 through 23; "99Z" = missing.

**GND\_LAT** is the balloon latitude at the ground (launch latitude) [in decimal degrees]:  
→ for fixed stations, the latitude of most recent location is assumed in IGRAs (see D3);  
→ for mobile stations, it is the latitude at the time of the balloon release.

**GND\_LONG** is the balloon longitude at the ground (launch longitude) [in decimal degrees]:  
→ for fixed stations, the longitude of the most recent location is assumed in IGRAs (see D3);  
→ for mobile stations, it is the longitude at the time of the balloon release.

**RAOB** indicates the type of radiosonde observations reported in the sounding data. It has four possible values:

- 0 = no RAOB (wind-only data);
- 1 = temperature (at any levels, regardless of simultaneous humidity or wind data);
- 2 = temperature and humidity (at any levels, regardless of simultaneous wind data);\*
- 3 = temperature and humidity enough complete to have a 'Sfc-to-500-hPa humidity sounding'.\*\*

(\*) Valid humidity data requires collocated temperature data, as part of IGRA's quality checks

(\*\*) For the definition of 'Sfc-to-500-hPa humidity sounding' see section D1 (parameter HUMb)

**RESa** is the vertical mean resolution of humidity data [in decametres].

RESa is defined as the geometric mean distance between consecutive levels having humidity data, obtained from collocated pressure and temperature data.

-999 = missing (number of humidity data levels < 2).

**RESb** is the vertical resolution of Sfc-to-500-hPa humidity data [in decametres].

RESb is calculated as RESa but restricting to data between the surface and 500 hPa, including these levels. Whenever humidity is missing at 500 hPa, a non-standard nearby level is equivalent provided it has humidity data and it is at about the same height from the surface, within 5% difference.

-999 = missing (no Sfc-to-500-hPa humidity data).

**TOPP** is the atmospheric pressure at the highest level with humidity data [in hPa];

-999 = missing (no humidity data).

**TOPZ** is the vertical extent (altitude above the mean sea level) of humidity data [in decametres];

-999 = missing (no humidity data / missing data at the surface level / missing station elevation).

### [D3. FORMAT DESCRIPTION OF FILE "stations-hum-record.txt"](#)

Notes:

1. Below the header record (with the names of variables) data records are organized in alphabetical order, according to the IGRA 2 stations' identification codes. Fields are delimited by pipes.

2. Whenever a humidity time-series from a station is interrupted by one or more calendar years, the humidity-related variables for that station are shown in multiple records along with the same content for the other variables (2797 data records in total for 1723 stations). The last year shown for active stations is 2016.

Data Record Format:

Variable	Columns	Type
STN_ID	1- 11	Character
LAT	13- 19	Real

LONG	21- 28	Real
ELEV	30- 33	Integer
NAME	35- 64	Character
COUNTRY	66- 96	Character
- - - - -	- - - - -	- - - - -
STN_POR	98-106	Character
HUM_POR	108-116	Character
NHOBS	118-122	Integer

-----

These variables have the following definitions:

**STN\_ID** is the station's identification code used in IGRA 2.

**LAT** is the latitude of the station (most recent location) [in decimal degrees\*];  
-98.888 = mobile station.

**LONG** is the longitude of the station (most recent location) [in decimal degrees\*];  
-998.888 = mobile station.

**ELEV** is the elevation of the station (most recent location) [in meters\*];  
-998 = mobile station;  
-999 = missing.

**NAME** is the name of the station.

**COUNTRY** is the name of the country. "Ocean" denotes floating stations.

**STN\_POR** is the station's period of record for the years before 2017, in the format YYYY/YYYY.

STN\_POR ranges from the first to the last year of record in sounding data, regardless of possible interruptions (missing years) and of the kind of weather-balloon used over time (including pilot balloons).

**HUM\_POR** is a period of record for humidity (relative humidity or dewpoint depression) in the format YYYY-YYYY;  
" - " = missing (no humidity data within STN\_POR).

HUM\_POR denotes consecutive years with any amount of humidity data in the year.

**NHOBS** is the number of humidity observations (i.e., sounding reports with humidity data) within HUM\_POR.

(\*) horizontal coordinates are rounded to the nearest thousandth of a degree (equivalent to 1.8" precision:  $\pm 56$  m distance in the N-S direction, and in the E-W direction at the Equator). Elevation is rounded to the nearest meter.

---

## [E. CONTACT FOR QUESTIONS](#)

Questions and feedback should be emailed to the authors of the dataset. Luis Gimeno (l.gimeno@uvigo.es) and Raquel Nieto (rnieto@uvigo.es) are principal investigators at the Environmental Physics Laboratory (EPhysLab) at the Faculty of Science of the University of Vigo. António Paulo Ferreira (ap.ferreira@uvigo.es) is a postdoctoral research fellow of the EPhysLab group.