This is the data format description of the files of the R, C and S versions of the Comprehensive Historical Upper Air Dataset (CHUAN).

PART A: Files with data on pressure levels

Column	Data
1	YEAR
2	MONTH
3	DAY, set to -9999 for monthly mean versions, left out for monthly means of the TD52/53 and German datasets.
4	HOURUTC: hour of measurement in UTC, set to -9999 for daily and monthly mean versions, left out for monthly means of the TD52/53 and German datasets.
5-317	Contain the GPH [m a.s.l.], T (temperature) [°C], WD (wind direction) [°], WS (wind speed) [m/s], RH (relative humidity) [%], DIFFTAU (dewpoint difference) [°C] and SH (specific humidity) [g/kg] DATA on all fixed pressure levels, each followed by a flag. The pressure levels are: 1000, 950, 925, 900, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 175, 150, 125, 100 and 50 hPa. Additional pressure levels at 70, 30, 20 and 10 hPa have been introduced for the radiosonde data in order to preserve all the data available.

Flags: are set to -9999 by default and have subsequently been set to the respective values in the original datasets.

The humidity measures have initially been introduced because the original radiosonde data, contrarily to the radiosonde dataset included and described in Grant et al. (2008), contains this information.

PART B: Files with data on geometric altitude levels

Column	Data
1	YEAR
2	MONTH
3	DAY, set to -9999 for monthly mean versions, left out for monthly means of the AE Monthly, AE Daily, African, TD52/53 and German datasets.
4	HOURUTC: hour of measurement in UTC, set to -9999 for daily and monthly mean versions, left out for monthly means of the AE Monthly, AE Daily, African, TD52/53 and German datasets.
5	HOURUTCFLAG: has been introduced because of time zone inconsistencies found in the AE Monthly dataset in the case of the time zone indicator flag being set to 1 (meaning GCT) in those original files. In this case, this flag is set

to 1111.
6-1340 Contain the p [hPa], T (temperature) [°C], WD (wind direction) [°], WS (wind speed) [m/s], U (u wind) [m/s],

V (v wind) [m/s], RH (relative humidity) [%] and SH (specific humidity) [g/kg] DATA on all
fixed geometrical altitude levels, each followed by a flag.
The geometrical altitude levels are: 0 m a.g.l. and 150, 200, 250, 300, 500, 750, 1000, 1250, 1500, 1750,
2000, 2250, ..., 1616250, 16500, 17000, 18000, ..., 29000 and 30000 m a.s.l..

Flags: are set to -9999 by default and have subsequently been set to the respective values in the original datasets.

Other flags are set to 1111 whenever linear interpolation has been applied to the original data.

The flag value 2222 for the data flags means that the information is considered doubtful, the flags 4444 or 1 that the information is implausible (e.g. data reported on levels below station elevation).

The flag value 3333 is applied to interpolated data considered doubtful, 5555 to interpolated data that is implausible.

The flag value 8888 replaces flags of value 7777 if values are considered doubtful. 9999 replaces the same flag value if data are considered implausible.

The flag value 0 is equivalent to the value -9999.

This has been applied to the early MWR kite and radiosonde data from the years 1922-1938 (MWR_1922). For the kite data no information on the number of observations for each level is available. However, for 6/1929-12/1931, mean altitudes reached are given on a monthly basis for each station. These are relatively stable (show no clear annual cycle) over time and have therefore been used to flag all data above the mean level over the period given. For the radiosondes, the number of observations is documented for each level for the period 8/1938-11/1942. For each month and station, we have determined the level reached by more than 13 sondes. Indeed the time series of these values display an annual cycle, but they also show a trend to lower values towards the end of the period, similar to or even larger than the amplitude of the annual cycle. Therefore, we have once again used the mean for the whole period to flag values above this level.

The humidity measures have initially been introduced because the original radiosonde data, contrarily to the radiosonde dataset included and described in Grant et al. (2008),

contains this information.

REFERENCES

Grant, A. N., S. Brönnimann, T. Ewen and A. Nagurny (2008): A new look at radiosonde data prior to 1958, accepted at J. Clim.