

A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905 – Part 2: Homogeneity adjustments for PILOT and radiosonde wind data

The reviewed paper describes the big effort which was made to homogenize the available upper air wind data from different sources starting in 1905 as far as possible. Main emphasis was put on the early years, although there the hardest conditions occur due to low data coverage.

Unfortunately neither the paper nor the data meet the scientific style conventions. In the current state it is not recommended to publish the article.

The diction makes the paper hard to read – the authors should improve the English considerably.

General remarks

- PILOT: The format for messages from fixed land stations which contain only wind data. They are called PILOT messages (see <http://www.ofcm.gov/fmh3/pdf/13-app-e.pdf>); pilot balloons or pibal is used more often.
- Names and symbols should be unified through the hole paper (as well in text, formulas and figures), e.g. NOAA-20CR, NOAA 20CR, NOAA20CR, 20CR, bg or ff, WS or Φ , WD or Δ , τ or 'an', '20CR'
- Meteorological conventions should be considered, e.g. that Φ is used as symbol for geopotential (like in part1)
- Formula should be written like in part1: the dependence of a variable on time is easier to read in Eq 2, part 1 than in Eq. 1 of part 2
- The paper does not explain the content of the data files.

Detailed remarks:

Page 337 Line 20: Thorne et al. (2011) does not deal with temperature biases of radiosondes or upper air temperature trends.

Page 339 Line 11: Brönnimann et al. (2012) analyses temperatures of the arctic not for mid-latitudes like cited here...

Page 340 Line 2: "data starting in 1905 at Lindenberg": Fig. 1 does not show a dot for Lindenberg for 1930-1940 and the data in the according netCDF-file 010393_t_U.nc (if the crude time axis was interpreted well by the reviewer) start in 1950.

Page 340 Line 8: "Since then the network has not changed much." – This is not true as the former gaps (South America, South Africa and Pacific) are filled.

Page 341 Line 10: Wrong sign, correct is: the value is corrected to value minus 360° in case of being larger than 180° .

P 343 Line 7: N=1470 means that the chosen interval is 4 years, this contradict to page 344 line 16 where 'the intervals a, b are generally chosen longer (1-8 years)' for break adjustment. Later (page 346 Line 8) it is written that 'interval for the break size estimation has been set to 8 years'.

Please, give the information in which cases N varies (as written p 342 line 17). The authors did not mention that k has to be larger than 730 and smaller than sample size – 730.

Page 343 Line 23: $\langle Q_k \rangle$ should have the index t_m as you have different means for 00:00 and 12:00 GMT time series.

Page 344 Line 4: It becomes not clear how the break probability was calculated, please include the conversion function.

Page 344 Line 18: Should it be the standard deviation of τ_ϕ ? (instead of the mean)

Page 344 Line 18: Do the authors mean 'the vertical mean of break sizes'?

Page 345 Line 22: Citation is missing.

Page 345 Line 23: The quotient was calculated, not the difference.

Page 345 Line 28: 'for the scatter plot' – which scatter plot is meant?

Page 347 Line 3: Figure A1 by Compo et al. (2011) shows low bias above 500 hPa and only for latitudes -30° to 40°N . Stickler et al (2010) only mention NOAA-20CR, but do not show or comment any data. Fig. 3 from part 1 does not show wind speeds but U and V for 150 hPa.

Page 347 Line 18: Time period is not given for Figure 4 data and Figure 4d is not mentioned. It is not clear, why data from 4a vanish during the step to 4b (as well in case 4c data below 60% vanish on the way to 4d).

Page 349 Line 1: The high variability of the time series in Fig. 2 makes it hard to see any shift in the sense of the authors.

Page 349 Line 3: 'only three pressure levels that have an almost complete time series back to 1935' – in Figure 2 only the 700-hPa-line goes back to that time.

Page 349 Line 19: 'No visible inhomogeneities remain.' As the reviewer understood all colored areas under the SNHT curve indicate inhomogeneities and such areas are still visible in Fig. 8.

Page 349 Line 21: The authors should provide a number of found inhomogeneities in the 1930s for wind speed over US. At page 350 Line 6 they wrote 'Generally it has been found that wind speed breaks over the US are relatively rare and relatively weak'. Please, check whether the adjustments have an influence on the data Brönnimann used.

Page 349 Line 25: 'The breaks in the 1970s and 1980s are no longer visible in the adjusted innovation time series.' The language is not precise enough: nothing was done for wind speed at station Bismark, so SNHT is unchanged.

Page 350 Line 12: 1997 no break is to be seen – what is meant?

Page 351 Line 19: 'generally directed southward' is not true as the adjustment has positive and negative sign.

Page 353 Line 6: The geographical name for the region is Antarctica.

Page 353 Line 14: Give the time range considered (1900-2010?).

Page 354 Lines 25ff: Please explain, why known breaks (Gruber and Haimberger (2008) are not considered.

Page 355 Line 2: Dee and Uppala (2009) did not made wind bias adjustments, only surface winds were mentioned.

Pages 361 to 383: The figures, especially the captions, suffering from inaccurateness.

Fig. 5: Lat=45N does mean what?

Fig. 6: The WS-Threshold is about 20, whereas it is higher than 25 in Fig. 8; the color of the triangles is not explained.

Technical remarks:

Language corrections are mentioned only in exceptional case, the article has to be revised (ideally with the help of a native speaker).

P 337 Lines4-6: “they still are an essential component of the observing network” is the same as “Even then they remain an essential component of the observing network.”

P 338 Line 4: “called GASP” and “(GASP)” is one to much.

Page 342 Line 12: better: Gaussian distributed

Page 343 Line 23: Eq number and ‘with’ is missing.

Page 346 Line 1: in the range 1.0–1.2 ~~provided~~ assumed there

Page 346 Line 2: For easier reading, please start the sentence with ‘Beyond 1960 the mean...’

Page 346 Line 10: It is not clear what is meant by ‘between the current break and the beginning of the time series’. (The last detected break?)

P346 Line 14: to track ~~of~~ the ascending

Page 347 Line 24: monthly means are shown in Fig. 5

Page 349 Line 19: in Fig. 6

Page 350 Line 11: No data for 700 and 300 hPa are shown.

Page 352 Line 20: Written is ‘at least 10 years’ ; caption of Fig 17 ‘at least 15 years’

Page 353 Line 13: Figure 22b

Fig. 1: The caption content “Between the 1960s and the 2000s the difference is relatively small.” does not belong to a caption.

Fig. 2: ‘Note shifts in 1938 and 1948’ does not belong to the caption if they are not marked.

Fig. 8: Changed right y-axis is not mentioned.

Fig. 9: WMO station number should be given in caption of Fig.3

Fig. 23: Replace BeakNum by total number of breaks.

Technical remarks to one file out of the data set (010393_t_U.nc):

The format of the data (netCDF) is highly appreciated. But the main advantage of netCDF (self-explaining files) is not given, because the file structure is orderless, variables are unexplained, certain attributes are not used. Variables not belonging to the file (e.g. Varno_R_Humidity etc.) should be excluded from the file. Ancillary Data and Flags (<http://cfconventions.org/1.6.html#ancillary-data>) should be used for the variables ,status' and 'anflag'. The use of standard_names for all variables is highly recommended. unit="" is not valid. 'event' is not explained. 'sonde_type' is set to missing value though the whole tested file. Version number of RAOBCORE is missing as well as the origin of the data.

The described dataset includes the self-demand to accord the climate and forecast conventions 1.4 – this is not the case, at least in the tested case of file 010393_t_U.nc (U-wind at Lindenberg, Germany)

The attribute of the variable 'pressure-layers' should be pascal, the valid range is certainly not 0 to 23 as given in the file. The time axis is not given properly: there are date (which should have the dimension unlimited, the numbers in the variable does not fit to the unit, _FillValue = -999 is not allowed in this case: data without time are useless), obs_time and original_time (with a wrong valid range 0 to 24 and additionally values up to 30) and index_days. Index_days starts with 18263 – if this is the content belonging to the unit 'days since 1900-01-01 00:00:00' than it would mean that the data set starts in 1950 and not in 1905 as written at page 340.

The global attributes are incomplete; one should follow the recommendations of cf: <http://cfconventions.org/1.6.html#identification-of-conventions>

The content of the file was not checked because of the inadequate file structure.