

## ***Interactive comment on “A global historical Radiosondes and Tracked Balloons Archive on standard pressure levels back to the 1920s” by L. Ramella Pralungo et al.***

**Anonymous Referee #3**

Received and published: 16 February 2014

General comments:

This work presents an historical archive of temperature and wind observations collected by radiosondes and pilot balloons in which observations have been interpolated the standard pressure levels and to the standard synoptic times of 0000 and 1200 UTC. In this form, the dataset simplifies the task of conducting timeseries analysis on standard-level temperatures and wind components, thus complementing the archives of raw radiosonde observations from which it is derived. Both the dataset and manuscript are well-structured, the data processing procedures are state-of-the-art, and the data and metadata are accessible as specified in the text. However, the

C337

language of the manuscript require some polishing to bring it up to par with standard English as well as to ensure clarity and correct typographical errors. I recommend that the authors seek the assistance of a native English speaker in this regard. I have provided examples in the "Technical Corrections" section to highlight some of the language issues. In addition, while the manuscript generally provides an appropriate level of detail, some aspects of the assignment of WMO identifiers, temporal interpolation procedure, and merging process (Sections 2, 5, and 6) require further explanation.

Specific comments:

Page 839, "and data have been provided only on significant levels, not pressure levels": In many cases, significant levels are reported and appropriate pressure values. Therefore, this statement is only correct if "not pressure levels" is changed to "not standard pressure levels", Although some further refinement of the statement may be required to clarify what is meant by "significant levels". From a stylistic standpoint, "not" could also be replaced with "rather than on", eliminating the need for the preceding comma.

Page 843, "station relocations have often split records ... second phase": It is difficult to understand what is meant by the first sentence in this paragraph. Also, there has not been a reference to the first phase in the text. In General, the discussion in this section is rather difficult to follow, and the process would be difficult to reproduce based on the information provided. I would suggest inserting a paragraph early in the section that summarizes the process described before this section goes into the details and clarifying the language throughout this section.

Page 844: Are these Interpolation steps performed regardless of how close together or far apart consecutive heights are in the original data? Has direct cross validation been performed in soundings that contain both altitude and pressure-level data?

Pages 845-846: In the context of the temporal interpolation to synoptic hours, I have three questions/requests for clarification in the text: 1. I am wondering how the authors distinguished between observations that were truly taken at asynoptic hours (e.g., 2100

C338

UTC) and those that were actually taken around the synoptic hours, yet there observation hour was reported as the launchtime which is often one or two hours prior to the synoptic hour. In the latter case, Which is particularly frequent after 1957, no adjustment of the observations would be needed to bring the observation in line with the true synoptic hour. 2. In addition, was the reported observation hour or the launch time used to determine the actual hour of observation, or how was the availability or absence of both of these pieces of time information handled? 3. Finally, it is mentioned in the paper that care is taken to prevent the inclusion of duplicate times following the temporal interpolation. How was this done? For example, in the not infrequent situation in which observations are available every six hours, Were the soundings closest to the true synoptic hours used and the others discarded?

Pages 845-846: How are discrepancies in station location among source data archives handled in the merging process? E.g., if there are two station records representing data for the same location in two different archives, and these records are identified with different WMO numbers (e.g., because the WMO ID changed over time) and have a coordinate discrepancy greater than  $0.5^\circ$  because of a coordinate error in one of the archives, would they appear as separate stations in the merged dataset?

Page 847: Regarding the order of priority in which the different data archives are merged, please clarify how the reader should interpret the order shown. Is archive (1) the one that ends up contributing the most data values to the merged dataset? More generally, please clarify how the merging process is performed. Is archive (1) loaded in first, and is archive (2) then used to fill in any holes in archive (1), and so on? Are data merged for each pressure level and variable separately? I.e., in the final data set, can a profile at a particular time and location contain values from multiple archives?

Page 848: What kinds of Spike and consistency tests were performed?

Page 849: Approximately how many values are removed as a result of the spike tests,

C339

inconsistency tests, and 20CR-departure tests?

Page 850: If observations at Lindenberg extend back to 1900, as is stated multiple times throughout the paper, why does the merged data set go back only to 1905?

Page 855, "only available for ERA-Interim and ERA-40 archives, where biascorrections procedure has been performed by ECMWF": does that mean that the data used from those two archives have actually already been homogenized? If so, perhaps it is worth clarifying that earlier in the paper when the different source archives are introduced.

It appears that in the station list provided along with the dataset, no name is listed with most of the stations. Although many disparities exist among station names in different archives, and station names have changed over time, many users of climate data lookup stations of interest by name, and excluding the station name entirely makes it difficult to cross-compare metadata in different station archives because the assignment of WMO station IDs is not always consistent through time or across data sets.

Technical corrections:

Page 838, "employing geopotential information National Oceanic and Atmosphere Administration": Some words appear to be missing here, and "Atmosphere" should be "Atmospheric".

Page 840, "to characterize the climate of troposphere and stratosphere in 1940–42 related to particularly strong El Niño event": Again, some articles and perhaps other words appear to be missing.

Page 840, "did a first look on": Rephrase.

Page 840, "since the ERA-40 (Uppala et al., 2005) and NCEP/DOE (Kistler et al., 2001) reanalyses that went back to": Remove "that".

Page 847, "to merge these archives in a global one": The use of the word global is confusing here. Rephrase.

C340

Page 848, "spike and consistency statistic tests have been performed in order to discard values have been performed erroneously": Something is incorrect in this sentence.

Page 848, very bottom: "Sibera" should be "Siberia".

Page 850, Footnote 2: "decate" should be "decade".

Page 857: "avaible" should be "available".

Page 849, "In the years 1945/1950 in Europe, Russia (even if in Moscow temperature observations has been maintained since 1938) and Japana rudimentary upper air observation network is present, but also in Australia, New Zealand, the Hawaii, Polonesia and Africa few, but important, stations are working": Awkward wording. Rephrase.

Page 851: "International Geographical Year" should be "International Geophysical Year".

Page 855: "identificator" should be "identifier".

---

Interactive comment on Earth Syst. Sci. Data Discuss., 6, 837, 2013.