

Interactive comment on “Radiosounding HARMonization (RHARM): a new homogenized dataset of radiosounding temperature, humidity and wind profiles with uncertainty” by Fabio Madonna et al.

Anonymous Referee #1

Received and published: 23 September 2020

The authors tried to obtain a homogeneous data set of temperature, humidity and wind profile through developing a Radiosounding HARMonization (RHARM) approach. This approach first post-processes the data since 2004, and then detect and adjust systematic errors in the historical observations with the help of the documented metadata. This involves one important issue, that is, the stations used in the study have complete metadata records. How complete are the metadata records for the stations used? Incomplete metadata records will lead to many unadjusted jumps in the final data set, which can't be used for climate research and other applications, similar to the raw

C1

dataset. Secondly, the data set derived from this study is mostly only since 2014 and lacks lots of stations in USA, Russia and China (Fig. 1) so as to prevent its applications in the future. Third, the verification of the derived data set is far from enough. This manuscript only shows the mean differences and PDF comparisons (whose differences are not obvious), which is not easy to see improvements from your approach. The authors will seek other ways to show the improvements of the derived data set. Fig. 14 shows time series that are not significantly different from the IGRA raw dataset and your homogenized dataset. Does this imply that the RHARM approach does not make significant adjustments for the raw data set? This reminds me that whether this small difference is due to incomplete metadata records, so that some jumps have not been adjusted? Lastly, the narrative structure of the piece and figure plots do not engender easy comprehension of claims, methods or main takeaways, which prevents a robust assessment. These may help improve the manuscript and requires extensive revisions.

Specific comments: 1. L21-25, do you mean the variables except relative humidity were harmonized on 16 standard pressure levels? Please clarify it. 2. L27-34, it is not clear that how to post-process, adjust and homogenize in each step? 3. L34, what is systematic effects? What's different from 'systematic biases or errors'? 4. L51, 'these climatic time series' is changed to 'these biased time series'? 5. L81, please cite other literatures at the end of the sentence and move (Hersbach et al., 2020) to the back of the ERA5? 6. L103, how many stations? instead of 'several sites' 7. L122, 'since 2004' is changed to 'from 2004'? 8. L132 and 179, pls revise 'a subset of 650 radiosounding stations'? 9. This study missed lots of stations in USA, Russia and China in Fig. 1? 10. Fig. 1 shows sonde types probably from different decades at different stations, which may mislead the readers. If there are several sonde types for one station, which type did you use? 11. L251, 'reasonably complete?' please see #8 12. L307, how did you handle surface data?

2020.

C3