We thank the two anonymous referees for their careful review of our manuscript and for their suggestions. Please find our replies to your comments written in blue.

Response to Anonymous Referee #1

This paper decribes the effort to develope a database based on historical radiosonde intercomparison activities. Knowing the instrument and correction information is the key to establishing a reliable upper air time series for long term climate change study. However, information for historical radiondes is often fragment and incomplete. I understand that collecting the data and compiling a database is a tedious task. This is an excellent work. It can be very helpful for the global radiosonde archive, thus for the global change study. I'm looking forwarding to seeing the database to be published in its final version.

The paper is well organized; data backgroud and database structure have been clearly illustrated. There are a lot more information in the historical upper air observations but I feel the important and necessary information is presented by the authors. I only have some minor points here:

1) L39-40: data base--> database
2) L42: use of it-->use of them
3) L57-58: the transmission of the data and their processing--> data transmission and processing
4) L66-67: Were the soundings on International Days and Week launched at the same UTC hours?
The soundings were launched simultaneously through a telegraphic signal (Brückner, 1899). The International Aeronautical Commission agreed then in 1909 to launch balloons at 7.a.m. Greenwich time on the appointed days (Dines, 1912).

Brückner, E. Bericht über den VII.internationalen Geographen-Kongress. Jahresbericht der Geographischen Gesellschaft von Bern, 17 (1898-1899). Geographische Gesellschaft Bern, 187-194, 1899.

Dines, W. H. The Vertical Temperature Distribution in the Atmosphere over England, and Some Remarks on the General and Local Circulation. Philosophical Transactions of the Royal Society of London. Series A, 211 253-278, 1912.

We will add this information to the manuscript.

5) L74-75: Payerne (Switzerland) --> Payerne, Switzerland
6) L80: Shlyakhov. -->remove the dot at the end
7) L92-93: "GCOS (Global Climate Observing System), the GRUAN (GCOS Reference upper-air network) " maybe changed to: "Global Climate Observing System (GCOS), the GCOS Reference Upper-Air Network (GRUAN)"
8) L93-94: and with this the quality is --> with which the radiosonde quality was
9) L102: amount of studies on radiosondes has been --> number of studies on radiosonde have been
10) L117: "which is useful to determine, were corrections have been applied." , this sentence is not clear.
11) L130: the 1930 and 1990 --> change to "the 1930s and 1990s" or "1930 and 1990"
12) L148: Figure 1: Left: Three types of Väisälä radiosondes --> Figure 1: Three types of Väisälä radiosondes. Left:
13) L226: have been-->were
14) L227: deviate--> differ

15) L230: "resp.", do you want to say "or"?
16) L266-267: "the individual campaign" --> individual campaigns
17) L267: unique identifier (UID)
18) L283: relate-->related
19) L285: have-->has
20) L291: end of line ":"--> "."
21) L335: intercomparison-->intercomparisons
22) L340: report-->were reported
We did not change this, because the verb refers to the present.

Thank you for the careful proof-reading of the manuscript. We will correct all the above mentioned spelling mistakes and add the additional information.

Response to Anonymous Referee #2

1. L93 on GRUAN can include "Bodeker et al. (2016). (see the list of GRUAN publications on https://www.gruan.org/documentation/articles).

2. L95-L100: Dirksen et al. (2014 for RS92) includes all those methods and should be mentioned.

Thank you for the suggestions! We will add these two references.

3. L11 & L23: "climate" should be added here too.

We relate here to e.g. the weather "analysis" from the assimilation processes of instrumental data. Thus, it does not seem necessary to refer to climate specifically in this context.

4. L29: Those two references should be mentioned.
Zhou, C., J. Wang, A. Dai and P. Thorne, 2021: A New Approach to Homogenize Global Twice-daily Radiosonde Temperature Data from 1958 to 2018. J. Climate, pp.1-64.
Dai, A., J. Wang, P. W. Thorne, D. E. Parker, L. Haimberger, and X. L. Wang, 2011: A new approach to homogenize daily radiosonde humidity data J. Climate, 24, 965-991.

Thank you for the suggestion! We will add these references.

5. In term of how reliable the error assessment is, it would be good to explain some of outliers in Figure 6, such as two in Fig. 6d.

The strong outliers in Figure 6d stem from the sondes Thommen JR-3 (comp009th3) and Airsondes (comp009air). Airsonde was only experimentally used in this intercomparison as it underwent further changes after the campaign, which we also note in the discussion and the summary tables of the intercomparison campaigns. It is unclear to us what caused the strong deviations for the Thommen JR-3 radiosondes at the 400hPa level. We could hypothesize that it might be caused by an error during the data processing of the campaign or due to a measurement error, rather than being a real deviation, because it only occurs on one level. We know from metadata that for the Thommen JR-3 radiosonde at this level only a total of six comparison values were available, but we were not able to find the valuable

raw data of the radiosonde ascents, thus we can not reproduce the error assessments made by after the campaign, which are presented in Figure 6c and d.

In order to illustrate the reliability of the error assessments, we will add a short paragraph to the discussion section, mentioning these outliers and also stating that such quality problems have to be considered when using the dataset.

6. This is a really nice archive of results from early radiosonde intercomparison campaigns. However, I think that the main application of the study is to help adjust errors in early radiosonde data, so the data can be part of homogenized radiosonde data for climate studies, esp trend analysis. A lot of prior work including those from the second author have devoted to homogenize global radiosonde temperature data. It would be good to provide some examples on whether the adjustments made from prior homogenization efforts are consistent with the errors archived by this work. If not, how does it affect the conclusions made by those studies? This would greatly highlight the usefulness of this archive.

The main motivation to create the herein presented database is the assimilation of early upper-air data in reanalyses. Therefore, estimates of the errors as described here are necessary, which can either be derived directly from the raw data, or from the assessments provided here. As you state, the data can also inform in homogenization procedures. It is however only a relative information, i.e. mainly errors of one sonde relative to other sondes. Some work in this direction has been performed (see e.g. http://othes.univie.ac.at/65166/1/70520.pdf, in German), but results are mixed and preliminary. Main challenge is to find comparison pairs and to make sure that really the same sensors were used during the campaigns and operationally. We consider such work beyond the scope of the present paper, but we will refer to the work cited above.

7. The examples given in the manuscript are all temperature data. Is it also true in your database? If so, it should be clarified.

The database contains the variables pressure, temperature, geopotential height, relative humidity, wind speed and wind direction. We realize that this is never explicitly stated in the manuscript itself (only in the tables of the database) and thus, we will add this information in section 3.1 of the manuscript.