

Response to the comments (SC1)

Eriko Kobayashi¹, Shunsuke Hoshino, Masami Iwabuchi, Takuji Sugidachi, Kensaku Shimizu and Masatomo Fujiwara

¹Aerological Observatory, 1-2 Nagamine, 5 Tsukuba-shi, Ibaraki, 305-0052, Japan

eriko-kobayashi@met.kishou.go.jp

5

Thank you very much for your valuable comments and suggestions.

Our responses to your comments are as follows:

10 1) The paper considers the mean in each of the 13 pressure layers of Kobayashi et al. (2012) and compares the ensemble average of the two instruments for each layer. The statistical approach is quite simple and clear. Nonetheless, state of the art literature is not considered.

In my opinion, it should be important to deepen the literature review and cite recent papers on alternative methods for radiosonde comparisons such as the approach based on functional data analysis of Fassò et al. (2014).

15

-Thank you for your very useful comment. The analysis method of Fassò et al. (2014) would really be helpful for radiosonde data comparisons as an alternative approach. However, we are afraid that it will take long time for us to work on functional data analysis, and thus we would like to try it in a future work. We have added the citation of Fassò et al. (2014) as one of our future tasks.

20

2) Figure 22 shows that the two instruments have differences with a probability much larger than stated, under zero mean Gaussian assumptions, in Tab. 4, leading to inconsistent measurements. This is especially true for daytime as noted by the authors in the abstract, in Section 5.3 and in the Summary.

Now, this may be due partly to random and correlated effects and partly to a non-Gaussian distribution of the differences.

25 Using notation of Eq. (9), we may have that

$$d > 3u_c$$

for various (and combined) reasons:

a) The ensemble average difference, estimated in Eq.(6) and mentioned in point 1 of this discussion, is not zero. For example, the observed difference of 0.4 K is enough to justify Tab. 4?

30 b) d is not Gaussian;

c) the number of samples is small.

I do not consider point c) in detail here because M is not very small and, moreover, the number of seconds per layer is relatively large.

Regarding point a), we have information given by Figure 8 where M (independent) launches are averaged and each layer gives an average of possibly correlated differences (d). So the following question arises:

- 5 Is the ensemble average difference (6) large enough to justify the inconsistency or other co-location issues are present, or measurements are not Gaussian, so the use of Immler's constants ($k=1,2,3$) is not appropriate?

I think that addressing this point using e.g. the histogram of differences (d) in each layer and/or other techniques to analyse the distribution of the differences (d) may help.

- 10 - We followed your suggestion to create the histogram of temperature differences to confirm the distribution. Please see the following Figure R1, which shows the histogram of temperature differences for 70-50hPa and 500-300hPa pressure layers. (Figure R1 has been added to the Supplement.) Firstly, as for the possibility that sample data has outliers, we detected outliers with a criterion of $d > 3\sigma$ (d and σ are estimated in Eq. (3) and Eq. (7) for all data during a comparison period) and the detected data were excluded from the statistical analysis (i.e., in Fig.8 in the original manuscript).
- 15 Therefore, extreme outliers are not included in the results of the statistical analysis (i.e., in Fig.8 through Fig.20 in revised manuscript). However, the distribution at pressures between 500hPa and 300hPa (the bottom panels of Figure R1) shows that there are some samples with large negative differences. This is probably because there were some issues in some of the flights in the troposphere, or there may have been some cases with calibration issues. At the moment, we do not know the exact reasons.

- 20 The top two panels of Figure R1 are for the histogram of temperature differences for 70-50hPa pressure layer, which show that the mean of the temperature differences at daytime was about -0.4 K. The temperature differences are considered to be normally distributed. Therefore, the large temperature difference for the daytime observations may have occurred due to systematic effects. We have added this note to Section 5.3.

- To investigate possible systematic effects in more detail, we need to make much more comparison observations with high-performance instruments in actual flights. Also, more detailed ground checks may be necessary. These are our future tasks.
- 25

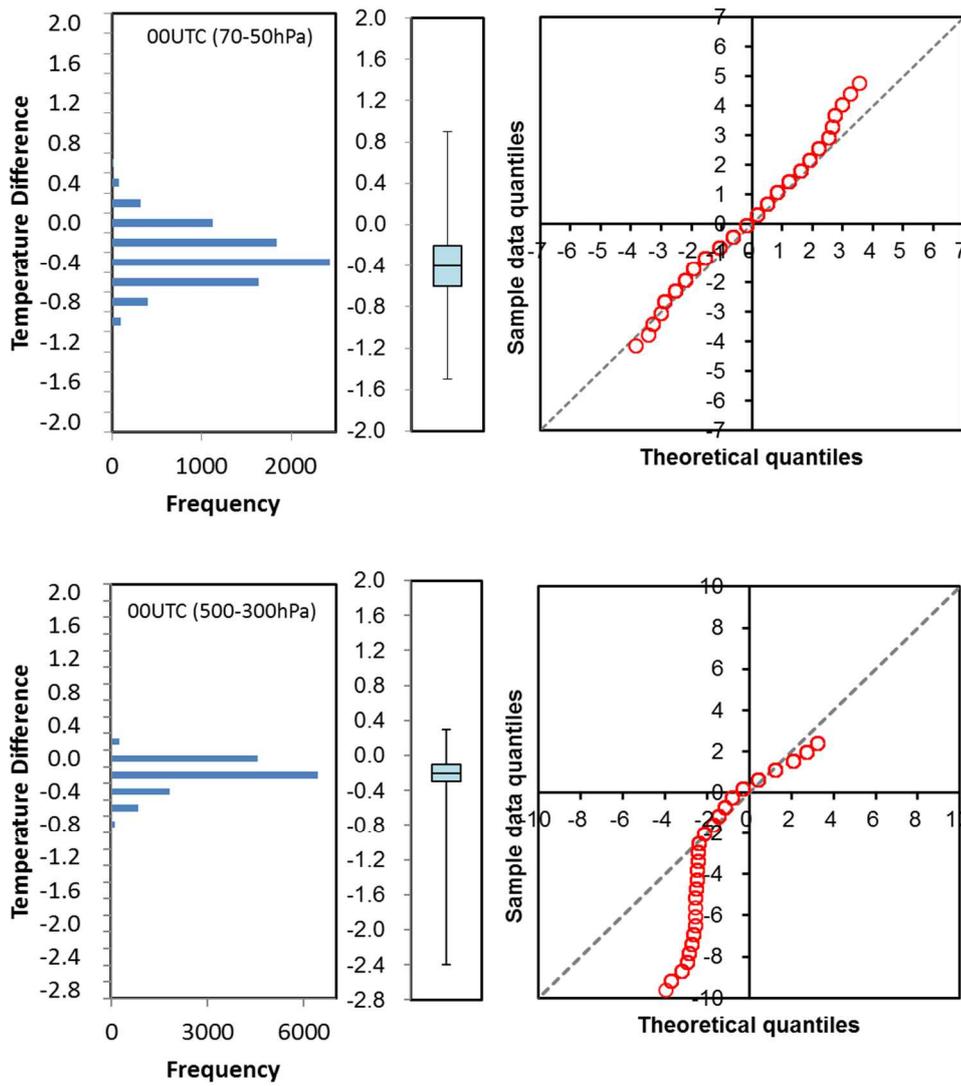
It is also possible that different rig configurations used for the comparison flights influenced the measurements, but this is not considered in the uncertainty budget in this study.

- 30 Minor points & typos:

Fig. 22 legend misspelling

Eq. (8) is a square missing?

-We have corrected these.



5 Figure R1: Distribution of the temperature differences at pressures between 70 hPa and 50 hPa, and 500 hPa and 300 hPa for daytime.

Left: histogram, center: box plot, and Right: Quantile-Quantile plot.