## Response to the comments (RC2)

Eriko Kobayashi[1], Shunsuke Hoshino, Masami Iwabuchi, Takuji Sugidachi, Kensaku Shimizu and Masatomo Fujiwara
[1]Aerological Observatory, 1-2 Nagamine, Tsukuba-shi, Ibaraki, 305-0052, Japan
eriko-kobayashi@met.kishou.go.jp

Thank you very much for your thoughtful comments and suggestions.
Our responses to your comments are as follows:

Major points

My first major point would be that the fact that the two products are far more frequently found to be inconsistent with one another than would be expected by chance if both products had a comprehensive metrological traceability. We would expect only 5% of measurements to fall outside k=2 yet far more cases occur (Section 5) particularly for daytime temperatures and tropospheric humidity. While I would not expect the authors to definitively ascertain the causes of this it would be useful to: i) more explicitly flag this; ii) in the discussion more directly allude to various possible reasons; and iii) point to ways that this could in future be addressed. For the latter point the authors may wish to point to the value of repeating the analysis once a third generation RS92 product is available and whether known issues to be addressed in that product version may help or hinder the explanation of discovered differences.

-Thank you very much for giving us very valuable comments. We have verified the distribution of temperature differences for daytime observation (see Figure R1) on the basis of your comments and the short comments (SC1, SC2). (Figure R1 has been added to the Supplement.) As for the temperature differences in the troposphere (bottom panels of Figure R1), although we had excluded outliers from the samples, some samples have large negative differences, which are either due to some issues during the flights or to possible calibration issues. On the other hand, the histograms of the temperature differences in the stratosphere (top panels in Figure R1) are considered to be normally distributed, with a non-zero mean, and we think that extreme cases such as those in the troposphere are not included. It is speculated that the main cause of the temperature differences in the stratosphere is due to systematic effects which still have not been fully corrected by both the GDP processes. As for RH differences in the troposphere (Figure R2), we have identified that the humidity sensor of RS-11G has a dry bias in the lower troposphere and a wet bias in the upper troposphere based on comparison observations with CFH. Thus, we consider that the RS-11G GDP needs to be upgraded. It is our future task to investigate the sources of the errors by e.g., making comparison observation with high-performance instruments. Also, we agree that we need to repeat the analysis with the third generation RS92 GRUAN data product to explain more details. We have added these notes to Sections 5.3 and 5.4 as follows:

-Section 5.3 : 'Possible reasons for the fact that the percentages of "inconsistent" and "significantly different" categories are larger at pressures < 150hPa at daytime are as follows. We investigated the histogram of temperature difference and found that it is normally distributed and that the number of samples is large enough. Therefore, the temperature difference in the stratosphere at daytime is thought to be caused by unexpected systematic effects. Also, some samples showed large temperature differences (about -0.5 K) even in the troposphere, which is considered to be due either to some issues during the flights or to possible calibration problems. Further works, including comparisons with high-performance temperature instruments and additional ground checks, are required. Also, the RS92 GDP version 3 is supposed to be available in the near future (Ruud Dirksen, private communication, 2018), and it would be useful to redo the analysis with the new RS92 GDP.'

-Section 5.4 : 'There are some samples with large RH differences (more than 10 %RH), which is considered to be either due to evaporative cooling effects or related to the sensor hysteresis characteristics as mentioned in section 5.2. In addition, the authors have identified that the humidity sensor of RS-11G shows drier values in the lower troposphere and wetter values in the upper troposphere when compared to chilled-mirror hygrometer measurements as mentioned in Section 6. It is our future work to improve the RS-11G RH GDP when more intercomparison data with chilled-mirror hygrometers become available.'

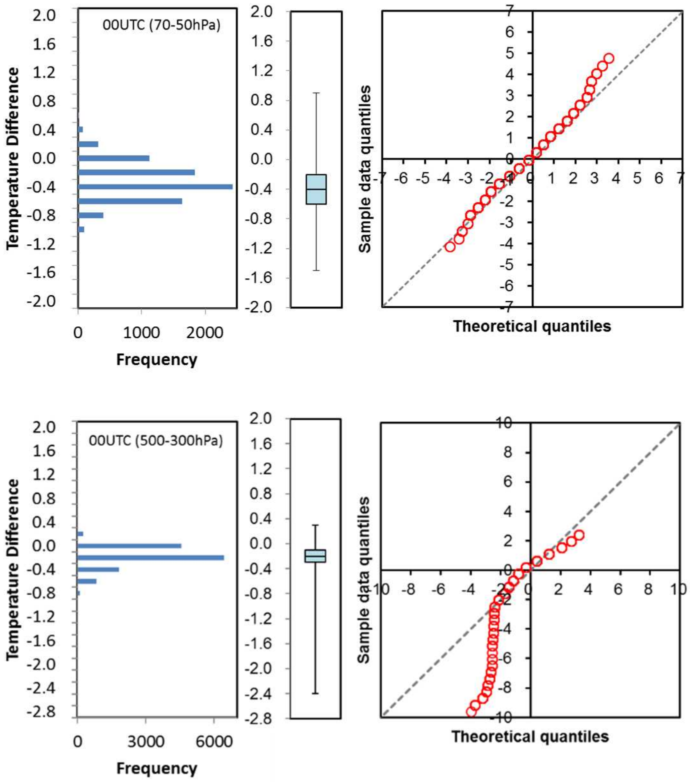


**Figure R1: Distribution of temperature difference at pressure between 70 hPa and 50 hPa, and 500 hPa and 300 hPa for daytime. Left:histogram, center:box plot, and Right:Quantile-Quantile plot.**

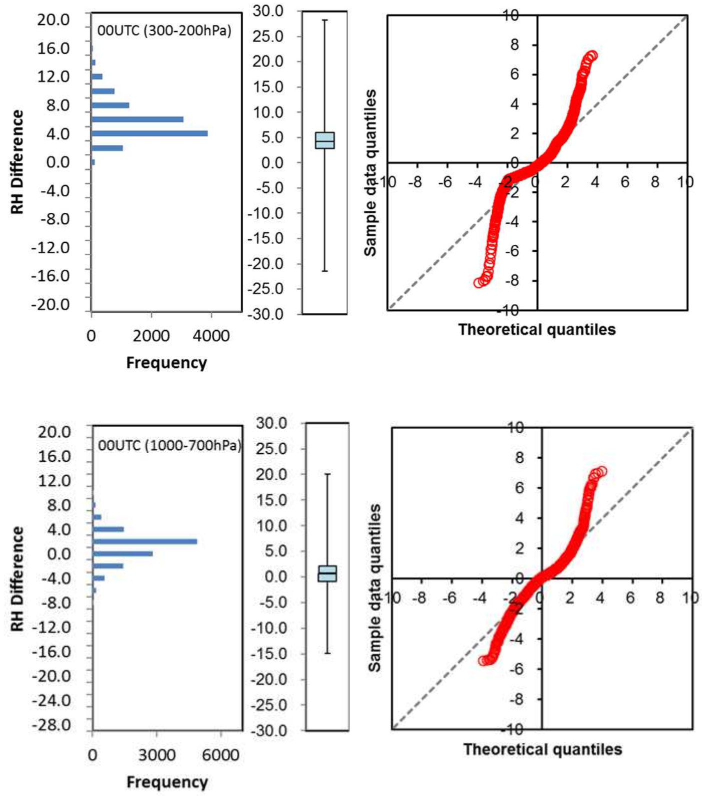


**Figure R2: Distribution of RH difference at pressure between 300 hPa and 200 hPa, and 1000 hPa and 700 hPa for daytime. Left:histogram, center:box plot, and Right:Quantile-Quantile plot.**

My second major point would be that substantially more detail on the product derivation for the RS-11G sonde would be useful in Section 2. Is there a diagram similar to the PTU diagrams produced within GAIA-CLIM by NPL? (e.g. http://www.gaia-clim.eu/document/product-traceability-and-uncertainty-gruan-rs92-radiosonde-temperature-product). Can the traceability analysis be shown in a similar manner by expanding / modifying Figure 1? Can further details on the important sources of uncertainty and how they were derived be shown? The GAIA-CLIM approach provides a basis to do this and the tables and analysis they employ in the PTU document linked above could be used and placed in a technical appendix in the final draft, likely at little effort to support Figures 1 through 4 and provide the reader with a better expectation of the RS-11G product. Tom Gardiner of NPL may be able to help the completion of this.

- We have added a diagram of traceability of the RS-11G sensors as in Figure R3 (Figure 1 in the revised manuscript). Further details of the traceability of the RS-11G sensors can be found in Section 5 of Kizu et al. (2018).

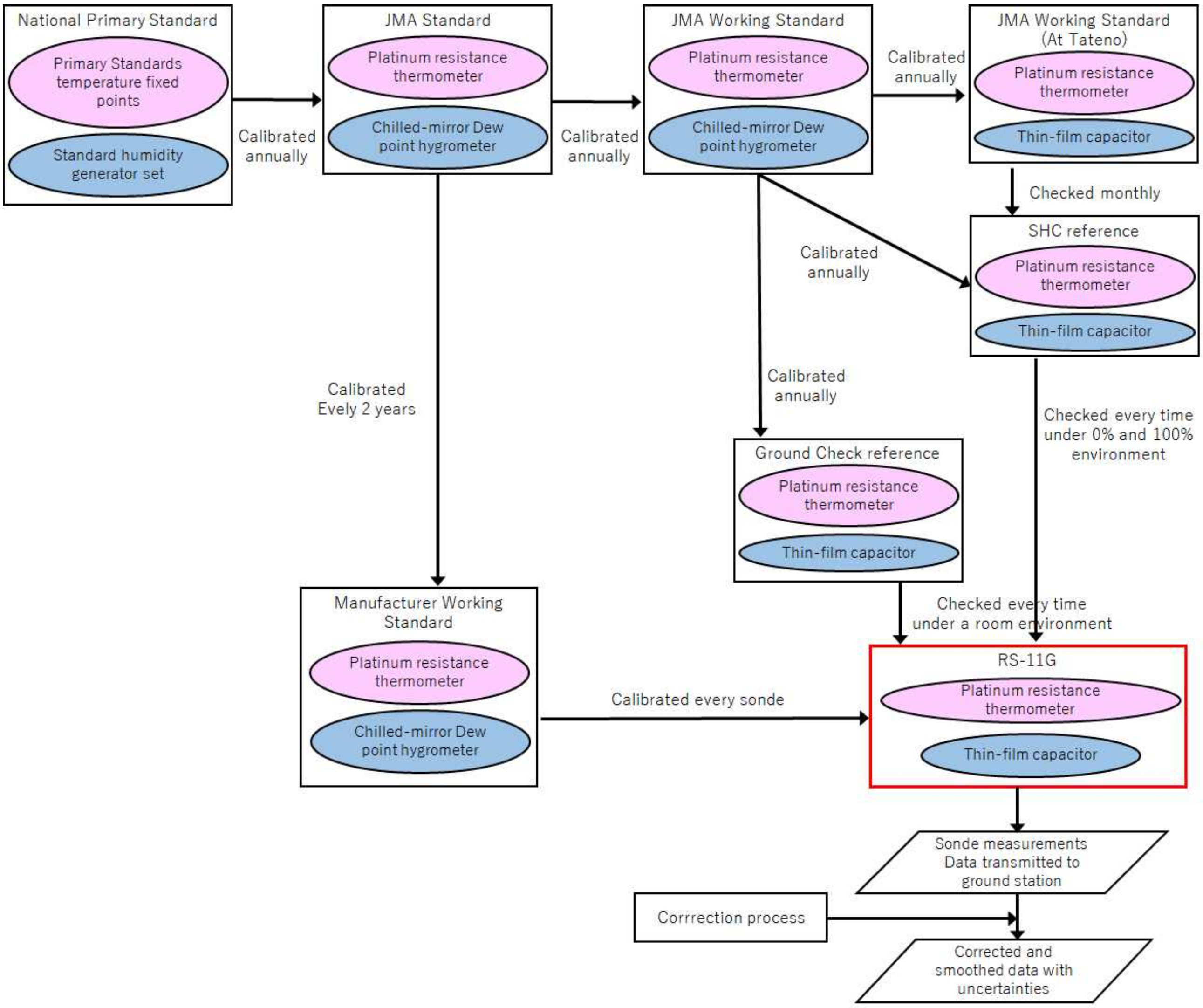


**Figure R3: Traceability of the temperature and RH sensors on RS-11G**

**Pink and blue ellipses indicate temperature and RH sensors, respectively. Parallelograms indicate data. The details of the correction procedures are shown in Figs. 2 through 5 in the revised manuscript.**

Third, given the differences found between different rod configurations (Section 3) would there be value in undertaking a split analysis where the results for each rig are considered in turn? If the resulting populations are too small to justify this then I would expand Section 3 to directly state this. Otherwise it would be a reasonable question that many readers may have.

-Figure R4 in this document shows temperature differences between RS-11G and RS92 for four different rig configurations. The sample size of each configuration is 11 for bamboo rod with fixed radiosondes, 21 for paper cardboard rod, 8 for bamboo rod with radiosondes hanged freely, and 17 for plastic cardboard rod. Although each sample size is not large enough, we think that it is useful to show the results for each configuration. The standard deviation for the paper cardboard rod tended to be larger than that for other rods. Also, the standard deviation with the plastic cardboard rod at pressure < 30 hPa is smaller than others in daytime data. In addition, when the radiosondes are fixed to the rod directly, sensor orientation can influence the temperature measurements (Rohden et al., 2016). We are planning to write another paper on the influences of different rig configurations on the temperature measurements. We have added Figure R4 (as Figure 7 in the revised manuscript) and revised the text in Section 3 as follows:
'However, the paper cardboard rod was thicker than the bamboo rod and kept much air inside, this might have caused unexpected heat flow and influenced the temperature measurements. The temperature differences investigated for each of the four different rig configurations are shown in Fig. 7. The temperature differences are averaged for each pressure layer based on the method described in Section 4.2. Note that the five outliers are not excluded in Fig. 7. The temperature difference and standard deviation for the paper cardboard rod tend to be somewhat larger than those for the bamboo and plastic rods in the lower troposphere and the lower stratosphere. In the main analysis (Figs. 8-20 in the revised manuscript) three among the soundings with the paper cardboard rod were excluded because of very large temperature differences. When these three outliers are excluded, the mean difference for those with the paper cardboard rod is found to be essentially within the standard deviation for all the four configurations combined.'

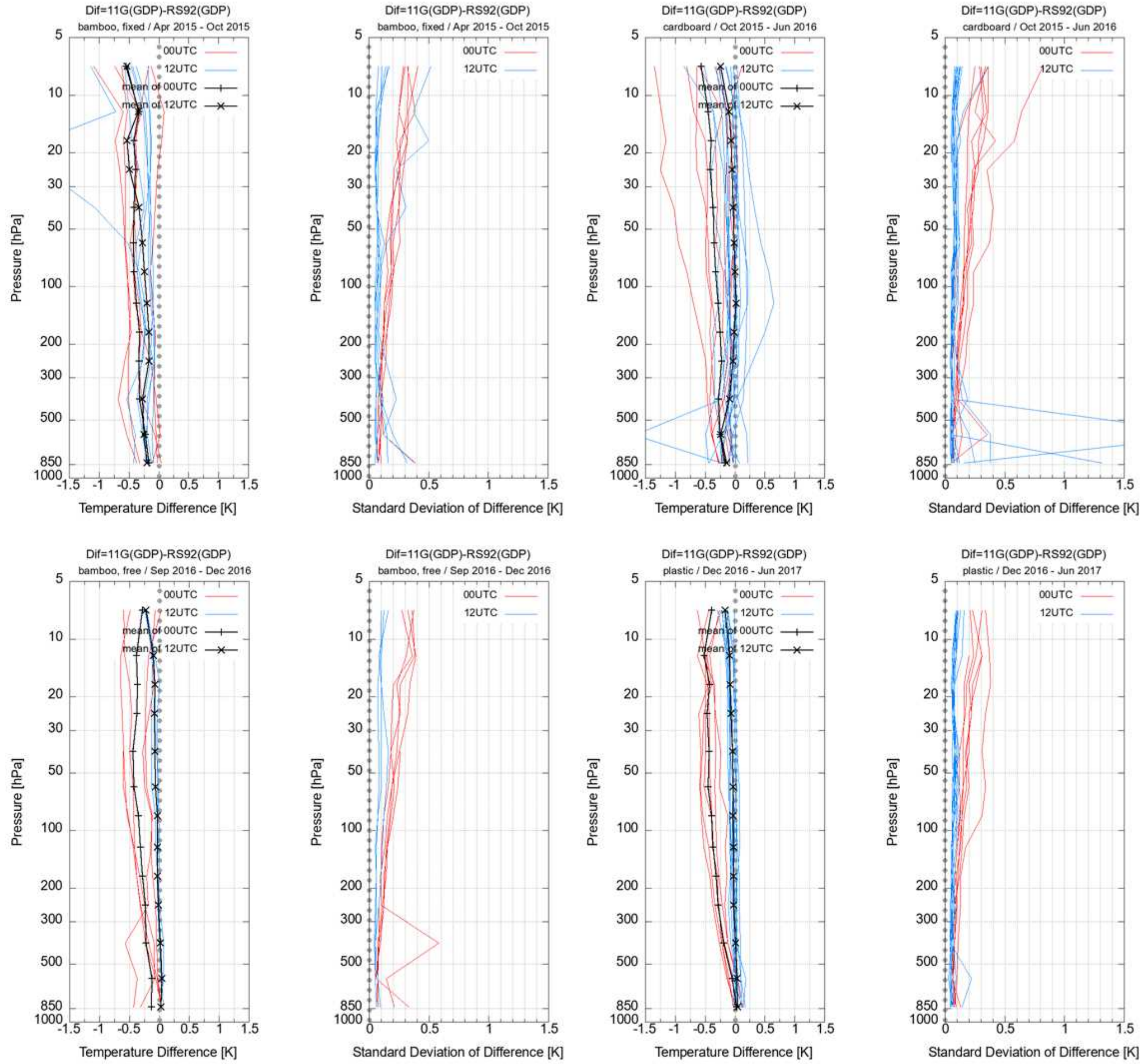


**Figure R4: Temperature differences and standard deviation for four different rig configurations**

**The temperature data were allocated to four categories, i.e., bamboo rod with fixed radiosondes (upper left two panels), paper cardboard rod (upper right two panels), bamboo rod with radiosondes hanged freely (lower left two panels), and plastic cardboard rod (lower right two panels). Red and blue lines show the results in daytime and nighttime observation, respectively. Black lines show means of temperature differences for daytime and nighttime data.**

Specific comments:

1. In the abstract it would be useful to be clear that the departures experienced are typical values. Currently it could be read as being more exact than the results show.

-We have revised the text about the temperature and RH differences as follows:

The temperature measurements of RS-11G were, on average 0.4 K lower than those of RS92-SGP in the stratosphere for daytime observations. The relative humidity measurements of RS-11G were, on average 2%RH lower than those of RS92-SGP under 90–100%RH conditions, while RS-11G gave on average 5%RH higher values than RS92-SGP under ≤ 50%RH conditions.

2. Figures 1 through 4 the rationale for the different shaped boxes should be clearly described in the figure 1 caption and this referred to in the caption to subsequent figures.

-We have added the explanation for the different shaped boxes to Fig.1 through 4 (Fig. 2 through 5 in the revised manuscript).

3. To Section 2.2 it would be worth appending text that alludes to the fact that the Lead Centre plan to reprocess the RS-92 product to v3 based upon new insights.

Some allusion to what the impacts of this are likely to be would, I think, be useful.

-We have added the following sentence to Section 2.2:

While the authors used version 2 of the RS92 GDP, version 3 is supposed to be available in the near future (Ruud Dirksen, private communication, 2018), and it would be useful to redo the analysis with it.

4. Page 6 paragraph starting line 5 why were there so many flights for which a GDP was not derived? Was there a pattern? Or an underlying reason why? This would be important to describe for full traceability of the analysis.

- The quality control procedures for the RS92 GDP are as follows (Dirksen et al., 2014): The first step verifies the results of the ground check procedure; after the GRUAN corrections have been applied to raw RS92 measurements, the second step checks that profile data are within valid ranges to ensure that the estimated uncertainties of GDPs are within the manufacturer-provided uncertainties. One of the excluded soundings of RS92 had more than 1.5%RH difference between two RH sensors at the ground check. Most of the excluded RS92 GDPs are considered to have failed the second step of the quality control procedures. Two-third of the excluded soundings were daytime observations. It is our future task to clarify the exact causes. We have added these explanations to Section 3.

5. P. 10 line 10 and line 34 delete ‘degrees’

-We have revise the text as suggested.

6. In table 6 does eq. 8 refer to equation 8 in your paper (seems unlikely) or rather in Dirksen et al? If the latter please be explicit.

-It is Equation 8 in Section 5.3 or our paper. We have added the text, ‘Uncertainties for frequency splitting, contamination correction, and moving averaging are associated with the use of filtering or of moving averaging, which are determined by using the standard deviation of the correction amounts.’ to the caption of Table 6 (Table 7 in revised paper).

Also, we have revised the description of ‘Averaging (filtering)’ in Table 5 as follows: ’Derived by Eq.(8) in Section 5.3, determined by using the standard deviation of the correction amount’ (Table 6 in revised paper.)

Thank you very much again for your valuable comments and suggestions.

References

Kizu N., Sugidachi T., Kobayashi E., Hoshino S., Shimizu K., Maeda R. and Fujiwara M.: Technical characteristics and GRUAN data processing for the Meisei RS-11G and iMS-100 radiosondes (GRUAN-TD-5), GRUAN Lead Centre, 2018.

Rohden, C., Sommer M. and Dirksen R.: Rigging Recommendations For Dual Radiosonde Soundings (GRUAN-TD-7). GRUAN Lead Centre, 2016.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4463-4490, 2014.