

Review of 'Comparison of the GRUAN data product for Meisei RS-11G and Vaisala RS92-SGP radiosondes at Tateno (36.06 N, 140.13 E), Japan by Eriko Kobayashi et al.

The authors describe the comparison of a new GRUAN data product derived from the Meisei-RS11 sonde with the longstanding Vaisala RS92-SGP product. The piece is overall well structured and written and as such publishable following relatively minor corrections.

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.

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.

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.

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.
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.

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.
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.
5. P. 10 line 10 and line 34 delete 'degrees'
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.