Review of "Comparison of GRUAN Data Products for Meisei iMS-100 and Vaisala RS92 Radiosondes at Tateno, Japan" by Shunsuke Hoshino et al.

The study undertakes a comparison of two radiosondes at the GRUAN site Tateno according to GRUAN principles of change management. Both sonde types have quantified uncertainties using the metrological principles of measurement traceability. The study is valuable both to GRUAN but also to the broader radiosonde user community in that it highlights issues around both sonde types. The journal is an appropriate target audience for this work and the study is clearly germane. The work in my view is publishable following corrections.

Major comments

1. The methods described in sections 2 through 4 are possibly in several cases a bit too brief and too heavily reliant upon the reader going back to and reading a number of previously published papers. Perhaps some key additional aspects need to be documented to ensure broad-scale methodological reproducibility. In particular section 4.5 is too brief given its overall importance in a GRUAN product comparison and should be expanded.
2. For figure 3 and subsequent similar figures it would be useful to describe in the figure 3 caption what the different shapes of the boxes denote? What is a circle, a parallelogram, a diamond and an oblong?
3. It would probably be worth spending some time discussing the very marked seasonality of the rejection rates shown in Figure 10 which is cited but not really discussed. Rejection rates are low in winter but very high in spring and summer. Why is this?
4. I always hate making this comment because it is immensely impressive for non-native speakers to produce papers in English. However, the paper would be much more readable if you could get a native speaker to edit for clarity. There are numerous places where minor edits would improve the readability and make the messaging stronger.

Minor comments

1. A reader would reasonably ask in the abstract why you mention 99 dual soundings but proceed to analyse only 57 of these. Can a few words be added to clarify why this 57 subset of the sample were analysed? Something like "Following data quality checks 57 flights were considered of sufficient data quality to produce GDP profiles and this subset is analysed"
2. Line 8 I would delete "with RS92-GDP" as this is already clear from earlier in the sentence
3. Line 19-20 I would write. "While the RS92-SGP radiosonde has a GDP it was required to seek alternative radiosonde models to use for operational reasons as the payloads often fall within the greater Tokyo metropolitan region and for health and safety reasons use of lighter instrumentation is necessary."

4. In lines 21-22 I would suggest making explicitly clear that the RS-11G has already been developed and certified as a GDP
5. Line 56 and not ant (typo)
6. Should line 105 not be 10’ x 10’ ? A geoid model at 10 degrees by 10 degrees feels implausibly coarse and the change to 5’ by 5’ later in paragraph is then huge.
7. Perhaps in 2.2.4 make clear why no pressure sensor is fitted. Presumably this is to save weight and because the errors in the hydrostatic equation based approximation are considered sufficiently small to justify the omission of such a sensor?
8. Lines 205-206 are not possible to follow logically. How did a single payload last 7 months? What is a logistic regression? Work required here to clarify this sentence please.
9. Lines 218-219 are again unclear. I think you mean something like: The criteria applied here apply solely to the present study and are not applied to the GDP. However, you also probably need to explain why this is the case. Surely the same processing should be applied to the GDP screening or is there some reason why this would not be possible?
10. The paragraph starting line 302 is presumably applicable to all meteorological elements measured and not just temperature. As such its placement here rather than elsewhere – perhaps most logically the discussion section of the paper – feels strange to me. There is already similar text on lines 398-405 so maybe you can just delete this?
11. Line 388-389 it is unclear for what parameter this finding applies. I assume pressure but it needs to be stated explicitly.