We would like to thank referee#3 for his/her review. Below we provide our responses, with the referee's comments in black and our replies in blue. The lines mentioned are those of the revised manuscript with Simple Markup for better readability.

Review of "Comparison of M10 and M20 Meteomodem radiosondes relative humidity measurements with ECMWF ERA5 above France: focus on the upper troposphere." By Sidy Diarra et al.

The paper is an interesting piece that fits within the overall journal remit and is likely ultimately publishable. Overall, greater eOorts on producing a tight scientifically rigorous text are required before this can be accepted. There are also some key issues that are missed in the current text. None of this is impossible to address so I would encourage the authors to revise and resubmit additionally addressing the comments from the two other reviewers.

Major comments

1. Humidity in the UTLS region is important in its own right because the low water vapour concentrations mean the UTLS is not radiatively saturated in the WV band and hence changes in WV irrespective of condensation to form clouds is relevant to understanding climate sensitivity and climate feedbacks. As such the scoping in the abstract (and the paper) feels restrictive compared to the real scientific impact of high quality observations in this region of the atmosphere.

We thank the reviewer for this comment. We agree that humidity in the UTLS region plays a key role not only in cloud formation processes, but also directly in the Earth's radiative balance due to the low water vapour concentrations and the resulting sensitivity of the radiative budget to small changes in humidity.

Following the reviewer's suggestion, we have revised the abstract (lines 16 to 18) and the introduction (lines 58 to 59) to better emphasize the importance of accurate humidity measurements in the UTLS for understanding radiative processes, climate sensitivity, and feedback mechanisms, in addition to cloud formation.

2. I find it odd that the introduction would choose not to mention the frostpoint hygrometer measurement technique even if only to note that whilst these provide much better characterised measurements, costs and technical complexity prohibit widespread adoption of this technique across the global network of observing sites with upper-air measurement capabilities.

The frost-point hygrometer measurement technique is now mentioned in the introduction (lines 77 to 86) with other techniques providing balloonborn water vapour measurements to be more exhaustive.

3. The ERA5 description in Section 2 feels a little cursory. In particular it feels like the issues with the reanalysis in the UTLS region including the need to reissue a segment as ERA5.1 to address these issues should be discussed even if they were principally to address issues in the temperature reanalysis fields in the UTLS region.

We agree that the description of ERA5 should be expanded to include known limitations, particularly the cold bias identified in the UTLS region. This cold bias, affecting the tropical lower stratosphere in the early 2000s, led to the release of the ERA5.1 reanalysis by the ECMWF, which is mentioned by the reviewer and corrects temperature-related issues primarily associated with the assimilation of satellite radiances. However, this issue does not affect the period considered in our study (2020–2024), and therefore ERA5.1 is not required for our analysis. Nevertheless, we acknowledge that such limitations highlight broader challenges in representing UTLS thermodynamic conditions, which may also have implications for humidity. The manuscript has been revised accordingly to clarify these points and the description of ERA5 has been expanded.

4. In section 2 I would include a photograph of the M10 and M20 sondes and in particular their sensors to enable a reader to understand the degree of distinction in housing, mounting and instrumentation between these two models.

As asked, photographs of the Meteomodem's sondes and sensors have been added on a new figure (Figure 1).

5. In section 3.1 is the assumption that the sonde is directly above the site? Given that there is GPS tracking applied it would surely be better to interpolate to the actual location which may under strong polar jet conditions vary from launch location by >100Km and be in a separate airmass to the air directly above the site particularly in the winter half year. Regardless, some more explicit consideration and quantification of irreducible co-location errors in the comparison feels warranted to avoid an unduly high frequency of false positive detections of a discrepancy – see Immler et al., 2008 GRUAN uncertainty paper.

As mentioned lines 358 to 360, "For each radiosounding profile, a colocated ERA5 RH profile has been calculated by 3D interpolation in space of the ERA5 reanalysis at 12:00 PM for daytime profile or 12:00 AM for nighttime profile taking into account the sonde displacement." We thus take into account that the sonde does not stay above the site but drifts with the wind during the ascent. Its exact position is used to calculate a collocated ERA5 profile, thereby reducing the uncertainties associated to co-location errors. Since this approach concerns all the comparisons discussed in the paper, we added this information to the data description lines 198 to 200.

Regarding the cited reference, it seems that the relevant paper is Immler et al., 2010 rather than Immler et al., 2008. The Immler et al., 2010 reference has therefore been added to section 2.3 to highlight the added value of GRUAN data products in the estimation of RH uncertainties.

6. Overall the results feel like they could be considerably better organised and tightened up. There is a little too much describing what the figures show and perhaps a little too light on aiding the reader to interpret the figures in the way you intend with supporting inferences made. Suggest thinking about structure of the results section and really ensuring that the key inferences are much more clearly articulated with less text just describing the figures.

The structure of the results section has been revised by removing section 3.3 as it did not provide added-value beyond sections 3.1. and 3.2., particularly given that only one manufacturer's software upgrade was performed on 30 January 2024. The findings in section 3.1 were examined in detail, particularly with regard to the interpretation of the differences between day and night.

7. At several points in the results an inference is made that the radiosondes are correctly reporting the true value. This can't be true as they are only ever measuring a proxy for the true measurand. They also may not be free of biases. When radiosondes and reanalyses disagree it needs more than the radiosondes are measurements to definitively reach an inference as to the biases in one or other series. E.g. lines 375-378 assume that the radiosondes are correct but they may be biased for various reasons such as contamination on the sensor after emerging from the cloud top or systematic sensor performance limitations.

We agree that we cannot claim that radiosondes report the true value. However, these last ten years a lot of intercomparison campaigns have been performed to evaluate and correct know errors on radiosondes (Dupont et al., 2020; Dirksen et al., 2024 for examples). Several comparisons where performed using other technologies as frost point hygrometers like the MORGANE campaign at La Réunion in 2015. The comparison give confidence on the fact that the radiosondes have no systematic bias or at least no bias exceeding the given respective uncertainties without being completely sure. Previous lines 375-378 have been completed to give more information and more caution is taken.

8. In 3.3 why are daytime and nighttime ascents mixed? This seems unduly limiting if part of the discrepancy arises from diOerences in radiative corrections which would be much larger by daytime. I would suggest performing this separately for daytime and nighttime ascents. If this is due to sample size limitations this should be made explicit here.

Initially, we don't make the distinction between daytime and nighttime ascents has it was already done in section 3.1 and also in section 3.4. This point is no more relevant because, as explained in the answer to question 6, section 3.3 has been removed.

9. In 3.4 it sounds like a rudimentary version of double diOerencing using ERA5 as the comparator. Formalising it as such and rewriting in that context would help. Otherwise how do you account for the fact that the two sites are sampling slightly diOerent climates?

We recognize that it is difficult to distinguish instrumental effects from meteorological variability when comparing different periods. In this section, the Trappes site is used to illustrate interannual variability recorded by the sondes, and ERA5 serves as a reference for estimating climatic variability between the two sites. Analyses were carried out to ensure that the sondes and ERA5 reflect the same climatic variability between the two sites; the results are summarized in lines 561 to 569.

10. The conclusion feels like a halfway house between a substantive summary (a true conclusion) and a discussion of broader context and next steps (discussion). It would be better to have a discussion section (with more details than present) followed by a conclusion section of at most 2 paragraphs highlighting the key findings.

The title of this section has been changed to “Conclusions and perspectives”. The discussion has been expanded upon in section 3, and we have revised the conclusion to highlight the main findings.

Minor comments

1. The sentence lines 31-34 does not make logical sense as written. Please rewrite for clarity

Done.

2. Line 41 I would use reported rather than observed. The humidity is inferred from a proxy for the true measurand. Implying that the radiosondes have directly observed humidity feels an over-reach.

Done.

3. 2.3 is mistitled as its about GRUAN processing rather than the GRUAN network per se

Title changed for “The M10 GRUAN data product”.

4. You could better clarify that the list in lines 290-293 pertain to the Virman et al paper as it took me 3 reads to try to understand which study (theirs or yours) was being discussed here

Done.

5. Why on line 351-352 is this tied to a single ascent? This makes little immediate sense and is insufficiently justified.

Although the balloon's ascent rate for each radiosounding is approximately 5 m/s, it fluctuates during the ascent as well as the wind encountered and give at the end a different vertical sampling for each radiosounding. In order to calculate the percentiles of the RH differences between M10 radiosondes and ERA5, we need a common vertical sampling. In all cases, interpolation is necessary. As the vertical sampling is similar for each radiosounding, we use the sampling corresponding to the first date selected in our study to interpolate all other profiles and ERA5 reanalyses. This point is now explained in the revised version of the manuscript.