We would like to thank referee#2 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.

The manuscript "Comparison of M10 and M20 Meteomodem radiosondes relative humidity measurements with ECMWF ERA5 above France: focus on the upper troposphere" presents the results of a comparison of the relative humidity values in the upper troposphere (200-300hPa), a part of the atmosphere relevant for contrail formation, from ERA5 reanalysis data versus operational in situ observations with the Meteomodem M10 and M20 radiosondes.

In addition the authors investigate the differences between manufacturer-processed data and data processing developed for GRUAN.

The authors find a clear dry bias for ERA5 data at high RH values, which is attributed to the fast cloud formation scheme in the numerical weather model that basically suppresses supersaturation.

For lower RH values, far away from saturation, the difference between ERA5 and radiosonde RH data is greatly reduced.

Furthermore, the authors report a relatively small dry bias for the manufacturer-processing, and also address the potential effects of change of the sounding model (M10 to M20) to the data.

A major deficiency of the manuscript is that there appears to be no clear common thread or storyline. It feels like the collage of loosely-connected topics and results that are lumped into one document.

As a consequence, the various topics are not covered with the necessary depth and detail, which stands in the way of reaching appropriate scientific value of the work presented in the manuscript.

We thank the reviewer for this clear summary of our work and for the constructive assessment of the manuscript.

We agree with the reviewer that the initial version of the manuscript did not sufficiently highlight a clear and coherent scientific storyline. In particular, the different analyses (ERA5 vs radiosondes comparison, processing differences, and instrument changes) were not explicitly linked to a single central question, which may have given the impression of a collection of loosely connected results.

The manuscript has been substantially revised to clarify the main objective. Some comparison between ERA5 and aircraft and lidar measurements already exist and some comparison between different radiosondes have also been performed but comparison between the quite recent METEOMODEM radiosonde and ERA5 have never been performed. This was the aim of this paper. To address this question, the manuscript follow the following progression:

(i) A comparison between ERA5 and METEOMODEM radiosonde RH is done on the whole troposphere with the two processing methods available for radiosondes (manufacturer and GRUAN-like)

(ii) A focus on the upper troposphere (200-300 hPa) to understand the highest difference observed between measurements and model is done.

(iii) The comparison between M10 and M20 sondes is used to assess the robustness of the results with respect to instrumental changes.

We hope that this revised structure provides a clearer scientific rationale and allows a more consistent interpretation of the results.

The introduction and discussion sections have been revised accordingly to better emphasize the scientific question and to provide a more coherent link between the different parts of the analysis.

High-level comments

The main motivation for this study provided in the introduction is the need for accurate humidity observations in the upper troposphere connected to the alleged overestimation of cloud formation by ERA5, as is indicated by the description of ERA5's cloud formation scheme by Tompkins et al. 2007.

To get an idea of the importance of this issue it would be good if the authors can provide additional background information in the form of references to relevant studies. Are there observations that support this alleged overestimation of the cloud cover by ERA5, do for example satellite-based cloud observations show discrepancies with ERA5?

Suitable references in this regard may be Wolf et al. (https://doi.org/10.5194/acp-25-157-2025) or Hildebrandt et al. (https://doi.org/10.5194/egusphere-2025-3048)

We thank the reviewer for this relevant comment and for suggesting these references. We agree that the motivation needs to be better supported. The suggested studies (Wolf et al., 2025 and Hildebrandt et al., 2025) have been added to the introduction to strengthen the discussion.

The description of ERA5 is a bit coarse. The authors could provide some information on the numerical model that ERA5 employs (the IFS). For example does the model include atmosphere-ocean coupling, what are the constraints, which schemes are used, what are performance-wise the strengths and weaknesses of the model?

The description of ERA5 (Section 2) has been revised to include information on IFS cycle 41r2.

At some locations in the manuscript the description of the Meteodem radiosonde systems is reads too much like a marketing brochure, please keep it neutral and factual.

We agree with this remark. The description of the Meteomodem radiosondes (Section 2.2) has been revised to ensure a more neutral and factual tone. In particular, promotional or subjective formulations (e.g., “innovative”, “efficient”, “environmentally friendly”) have been removed or rephrased, and the section now focuses on technical characteristics relevant to the measurements.

The common parameter for the comparison is relative humidity (RH). For this purpose the air temperature and specific humidity produced by ERA5 must be converted to RH. A similar exercise is performed for the radiosonde data, although radiosondes do measure RH directly. Is this because radiosonde BUFR data is used in this study? Wouldn't it be more practical to use the radiosonde's RH data directly? This would eliminate the need for a lengthy and complicated discussion of the conversion of dew point data to RH. Since one of the authors is directly affiliated to the manufacturer, this should be feasible. Furthermore, GRUAN data products include RH values.

In this study, we use the files with the HR_complet.csv extension, which are available here: https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=97&id_rubrique=33. These files report the dew point temperature which need to be converted into RH. However it wouldn't be more practical to use the radiosonde's RH data directly because this RH data are estimated over liquid water calculated using the Sonntag formulation. In all cases, conversions are needed to ensure consistency between the formulations used for the different datasets. The formulations used in this study issue from the IFS documentation allow to calculate RH over ice for temperatures below 250.16 K, over liquid water above 273.16 K, and over a mixed-phase in between. This allows identifying the occurrence of supersaturation with respect to ice in the upper troposphere relevant for contrails studies.

Clarifications have been done lines 201 to 206 and 273-274 in the revised version of the manuscript.

A description of the GRUAN data processing for the M10 should be provided, as well as a discussion of the differences between GRUAN data processing and manufacturer data processing. Latter is essential for the reader to understand and interpret the observed differences in the measurement data.

The description of the GRUAN data processing was already provided in the manuscript but is now more detailed to clarify the differences between GRUAN data processing and the manufacturer data processing. We have also added a brief statement highlighting that these corrections account for known limitations of capacitive humidity sensors.

The description of the plots in the figures often contain a listing of obvious and not necessarily relevant facts. Please focus on what the data presented in the plots tell us. For example, mean values of differences/biases are more interesting than the maximum values of outliers.

We agree that focusing on mean differences provides the most relevant statistical assessment of the bias. However, it may also obscure the fact that, for individual collocated profiles comparisons, the differences may be larger or even have the opposite sign. We therefore believe that it is important to keep this in mind, whilst ensuring that the discussion remains focused and does not become unnecessarily descriptive or confusing.

Throughout the manuscript the measurement uncertainties are referred to as % (which I presume to be relative errors) or %RH, which is quite confusing to the reader. Please be clear, and consistent, in this regard.

The differences expressed as percentages in the manuscript are absolute differences, not relative differences. This is now clarified throughout the manuscript.

Section 3.3, that discusses the impact of changes to the data processing is not very convincing. Apparently changes to the processing software of the manufacturer data product are made regularly, and the data sets used for figure 4 and table 1 contain a mix of processing versions, where the time windows of the individual versions are not clear. I would prefer to see a comparison with two different processing versions applied to the complete period 2020-2024.
The numbers given in table 1 are hard to interpret. Are min,max and Q25-Q75 all necessary to make the point?

After discussion with Meteomodem, we confirm that only one manufacturer's processing upgrade was implemented affecting the RH data during the study period, on 30 January 2024. No processing update was applied to the temperature data. Since less than one year of observations is available following this update, a robust assessment of its impact over the 2020-2024 period is not possible. We therefore agree that the former section 3.3 is no longer relevant and have removed it in the revised version of the manuscript. The key information previously given in Table 1 has been incorporated into section 3.1.

Similar applies to section 3.4, discussing the impact of switching from M10 to M20. The plots in figure 5 are quite cluttered, and it is difficult for the reader to follow the conclusions the authors draw from these plots. To what extend can real instrumental effects be separated from meteorological effects?

Figure 5 has been simplified in the revised version of the manuscript.

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, 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.

The difference between the daytime and nighttime results for Mime (upper two plots) are quite striking (the green and red traces seem to shift from aligning with the blue trace to aligning with the black trace). A qualitative discussion of this effect would be appropriate.

A discussion about the difference between the daytime and nighttime results for the manufacturer processing method has been added lines 410 to 418 in the revised version of the manuscript.

The authors should refrain from using subjective adjectives when describing results. E.g. write good agreement instead of a very good agreement, or even better: data A and B agree within x%.

The subjective adjectives have been removed in the revised version of the manuscript.

Is the detailed discussion of RH over water or ice relevant? Radiosondes report RH over liquid water as stipulated by WMO.

Yes, radiosondes report RH over liquid water but as mentioned before, it is not the case in this paper as we recalculate the RH according to the IFS documentation's formulations.

Detailed comments

l20: Now the sentence implies that the data processing was developed by Meteo France. The correct situation is that the soundings are performed by Meteo France, but the processing is done with the software provided by the manufacturer. To prevent confusion refer to this stream as manufacturer data processing.

Done

l21: the term GRUAN standard does not exist for data products. In case of the M10 radiosonde a GRUAN data product is under development. Better use the term GRUAN processing.

Done

l23: 800 hPa -> 800 hPa, and

Done.

l23: is it 2% or 2%RH? This applies to various other locations in the manuscript as well

The differences expressed as percentages in the manuscript are absolute differences, not relative differences. This is now clarified throughout the manuscript.

l25: maximum -> largest

Done

l26: can you offer an explanation why the difference is larger for nighttime measurements?

The larger difference observed for the nighttime measurements can be due to the temperature-gradient correction applied before the 30 January 2024 which was different for daytime and nighttime data in the operational MF processing method. This is now mentioned in the abstract and detailed in section 3.1.

l30: mention that 200-300 hPa corresponds to the altitude relevant for contrails

Done

l31: indicate -> measure [or record]

Changed for record.

l49: For completeness, you can mention that water vapor is important for the transport of latent heat.

Done

l50: Thusfar you have described the role of WV in the atmosphere's energy budget, but did not discuss sources, whether natural or anthropogenic. Just state that aviation contributes to global warming by contrail formation in the upper troposphere.

Done

l53: is calculating the right term here? Suggestion: deriving, obtaining, or establishing.

Changed for establishing.

l54: However ... temperature. This stand-alone sentence doesn't add much.

This sentence has been replaced by this one: “Accurate RH derivation depends on accurate measurements of both the temperature and the water vapour concentration.” allowing a better transition with the following sentence.

l56: complex -> challenging

Done

l63: reproducing -> capturing

Done

l64: insert "in situ" after other hand.

Done

l65: full -> global

Done

l68-69: Awkward sentence. I assume you want to provide a reference to the dataset. Please rephrase.

The sentence has been rephrased.

l76-81: this sentence covers both homogenisation, trend detection and actual trends. Use separate sentences.

The sentence has been separated in two sentences.

l81: "In the past" radiosonde intercomparisons is not a thing of the past, this method is still very much employed today. Insert the word performance before evaluations.

"In the past" has been removed as we agree that radiosonde intercomparisons are still made. It is even necessary in the steps to become a new GRUAN data product. The word performance has been added.

l85: this statement may be a bit too bold. I would prefer to say that it helped to improve data quality radiosoundings.

The sentence has been modified according to the reviewer suggestion.

l88: insert comma after world.

Done.

l90: Goal of GRUAN is to provide reference quality measurements of ECVs, such as water vapor.

The sentence has been modified and is now "GRUAN (GCOS Reference Upper-Air Network, (Seidel et al., 2009)) was created to provide reference quality measurements of essential climate variables, such as water vapor."

l94: capacitative sensors are used for RH measurements.

Yes, there is also thermistors on the radiosondes. The sentence has been modified as follow (lines 104 to 106): "The humidity sensors of these radiosondes are capacitive sensors consisting of a dielectric polymer placed between two electrodes, whose dielectric constant varies proportionally with RH."

l101: Nothing spurious about time lag or radiative heating. Just state that important error sources for RH measurements are time lag at low temperatures and radiative dry bias.

The sentence has been changed for "Important error sources for RH measurements are time lag at low temperatures and radiative effects." (lines 111-112)

l104: what are these "GRUAN standards"?

The word standards has been replaced by requirements.

l106: what is the cause for this remaining bias between M10 and RS92? Calibration issues?

Extracted from Dupont et al. (2020): "The correction is much more efficient for nighttime profiles (OHP campaign) in relation with the temperature gradient that is very low for nighttime periods. For the daytime profile obtained during the SIRTA field experiment, the most important difference between M10 and RS92 (RH_{M10} . RH_{RS92}) appears between 5 and 10 km, where a lot of cirrus clouds occur during the different field experiments. In some events, the

temperature-gradient correction can have a too strong impact on the relative humidity profiles compared to the RS92 radiosonde."

l108: are there other relevant studies on evaluating the performance of the M10?

A campaign called MALICCA was done in April 2013 at La Réunion. Only one comparison on 8 April 2013 between M10, RS92 RH and H2O Lidar and has been published in Keckhut et al., 2015. When similar corrections are applied for the two sondes (radiative and response time), they are in agreement. Several measurement's campaigns have been done after at the Maïdo Observatory in La Réunion from 2014 to 2017 (MORGANE, HAIC …) with 12 launches including M10 and CFH sondes and sometimes also RS92. The results have been presented in GRUAN meetings but not yet published.

Reference: Keckhut et al., 2015 : Introduction to the MALICCA campaign dedicated to the validation of Upper Air Meteorological parameters, J. Appl. Remote Sens, 9(1), 094099.

In May/June 2025 was organized the TRACIS campaign at SIRTA to evaluate the capability for measuring water vapor in the upper troposphere. This campaign relies on the release of around twenty multi-payload radiosonde probes including Meteomodem sondes. The results of this campaign in not yet published.

l112: formed by -> constructed from GDPs of

Done.

l113: insert comma after radiosondes.

Done.

l123: what are these "GRUAN standards"?

The introduction has been modified to fulfill referee's comments and this sentence has been removed.

l124: add: and to evaluate the manufacturer processing vs the GRUAN processing.

The introduction has been modified to fulfill referee's comments and this sentence has been removed.

l133: including -> such as

Done.

l134: such as -> including

Done.

l150: p is not defined

P is the atmospheric pressure in Pascal, it is now mentioned.

l163: text suggestion: twice per day, at/around noon and midnight.

Suggestion accepted.

l167-169: there is an inconsistency in the numbers, unless the sampling frequency is 0.5Hz

The sampling frequency is indeed 1 Hz, it is the average vertical sampling that is 5 m. it is now corrected.

l176: remove innovative (marketing jargon)

Done.

l185: calibrate -> correct (?)

Yes, modified. In fact, the relative humidity retrieved with the humidity sensor is those at the temperature of the humidity sensor. To get the relative humidity of the air, the temperature of the air and the temperature of the humidity sensor have to be known to correct the relative humidity and get the relative humidity of the air.

l185: who is the manufacturer of the RH sensor, Meteomodem?

No, UPSI. It is now written on line 252.

l187-190: marketing speak

We have removed the first sentence, but the other two allow the reader, on the one hand, to identify which Meteomodem sonde is being used based on the date and, on the other hand, to introduce the next section of our article.

l195: tunable heating system. Remove "which can be switched off"

Done.

l199: Which sensor is meant here (temperature, pressure, RH)?

The humidity sensor. It is now clarified in the paper.

l203ff: I assume this elaborate recalculation is necessary because you use BUFR data. If so, please mention this.

No, we use the files with the HR_complet.csv extension provided here: https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=97&id_rubrique=33. It is now added in the revised version of the manuscript.

l217-219: in principle radiosondes measure/report RH over liquid water.

Yes but relative humidity was recalculated for all datasets to ensure consistency between ERA5 and radiosonde observations as mentioned before in the fourth high-level comment.

l230: more appropriate references on GRUAN are Seidel2009 (10.1175/2008bams2540.1) or Bodeker2016 (10.1175/bams-d-14-00072.1).

Vömel et al. (2009) has been replaced by Seidel et al. (2009).

l231: the GRUAN data product for the M10 is still under development. The data processing does indeed exist.

This sentence has been reworded as section 2.3 has been expanded.

l240: for above reason these data are not processed by GRUAN. Correct is to say that GRUAN processing method has been applied.

Yes, all the section 2.3 has been reworded.

l255: I am not sure whether considering the mixed-phase is relevant here. WMO stipulates radiosondes to report RH over liquid water.

It is always to ensure consistency between ERA5 and radiosonde observations.

l261: I think you mean manufacturer-processing here.

Yes, each time “Météo France processing” was written, we changed for “manufacturer processing” in the revised version of the manuscript.

l290: explain why should the location make a difference

The location can make a difference because the climates are very different, continental for our study, tropical for the study of Virman et al., 2021. The water vapour mixing ratio is larger close to the surface and in the low troposphere for tropical climate compared to continental climate. In addition, the study of Virman et al., 2021 was during the convective period. The convection is more intense in the tropics than in the midlatitudes and the ability of ERA5 to simulate a realistic vertical distribution of water vapour mixing ratios depends strongly on its convective scheme. This last point has been added in the revised version of the manuscript.

l393: processings -> processing methods

Section removed in the revised version of the manuscript.

l404: describe in the caption what is represented by the boxes and the lines.

Section removed in the revised version of the manuscript.

l406: remove very

Section removed in the revised version of the manuscript.

l408: precision -> uncertainty

Section removed in the revised version of the manuscript.

l529-535: this is too speculative. I see no basis for this statement.

This paragraph has been removed in the revised version of the manuscript.