We would like to thank the anonymous reviewers for their very detailed comments on our work. The two reviewers have highlighted important imprecisions in the paper. We hope that our answers and the modifications added to the manuscript will help future readers better understand our work and avoid any misinterpretation. The updates appear in track-change mode in the article file.

This document presents our answers to RC2. For added clarity, we colour coded our answers as follows:

Black: original comments from Anonymous Reviewer #2
Blue: our direct answer
Green: corrections and/or additions as they appear in the updated manuscript. The new line number also appear.

1. General Comments

This paper presents a novel method for assessing humidity fields from numerical weather prediction models with estimates from the SAPHIR instrument. The probabilistic methodology used to estimate relative humidity from SAPHIR is exploited to provide a new approach for model assessment. The methodology also allows for a confidence interval to be placed on comparisons where classical 'bulk' comparisons.

This study demonstrates an innovation that yields more nuanced results for satellite and model inter-comparisons. This is important for relative humidity, where uncertainties in satellite measurements can be as high as 10% RH for some instruments (especially heritage infrared sounders). Overall, I find that this study is of scientific value and recommend it for publication, after all the issues that I have highlighted are addressed.

2. Specific Comments

Line 21: The final sentence in your abstract is illustrating a key point of your study but it is missing the "why" of its importance. Adding another sentence or editing this final one will make it more impactful.

Thank you for the suggestion. Following this remark, we worked on the abstract and developed as follows:

Lines 21 to 24: "Specifically, it reveals cases where the ARPEGE simulated values falling within the deterministic confidence range actually correspond to extreme departures in the reference distribution, highlighting the shortcomings of the too-common Gaussian assumption on the reference error, on which most current deterministic comparison methods are based."

Line 42: Why use a reference for precipitation when talking about humidity? There are plenty of water vapour retrieval algorithm papers that perform an inversion between an atmospheric stat vector and observed radiances/brightness temperatures. Please update.

We understand and agree to that comment. The following references related to water vapour were added instead:

References:
Rosenkranz P. 2001. Retrieval of temperature andmoisture profiles from AMSU-A and AMSU-B measurements. IEEE Trans. Geosci. Remote Sens. 39: 2429–2435.

Lerner, J. A., Weisz, E., and Kirchengast, G., Temperature and humidity retrieval from simulated Infrared Atmospheric Sounding Interferometer (IASI) measurements, *J. Geophys. Res.*, 107(D14), doi:10.1029/2001JD900254, 2002.
Karbou F, Aires F, Prigent C, Eymard L. 2005. Potential of Advanced Microwaves Sounding Unit-A (AMSU-A) and AMSU-B measurements for atmospheric temperature and humidity profiling over land. J. Geophys. Res. 110: D07109, DOI: 10.1029/2004JD005318.
Divakarla, M., et al. (2014), The CrIMSS EDR Algorithm: Characterization, Optimization, and Validation, *J. Geophys. Res. Atmos.*, 119, 4953– 4977, doi:10.1002/2013JD020438.

Lines 45-47: Averaging is not the only method used to get data on the same resolution. The discussion here does not include the use of averaging kernels, which are used to smooth model or in situ profiles relative to the vertical resolution of the satellite measurement. See "Rodgers, C.D. and Connor, B.J., 2003. Intercomparison of remote sounding instruments. Journal of Geophysical Research: Atmospheres, 108(D3)."

We reworked the sentence to mention the use of averaging kernels. It now reads as follows:

Lines 47 to 49: "In any case, the comparisons usually involve spatial and/or temporal averaging, sometimes involving error bars or the use of averaging kernels to smooth models or in situ profiles relative to the vertical resolution of the satellite measurement (Rodgers and Connor, 2003)."

Lines 65-67: It reads a bit strange when you talk about RH and then reference a precipitation paper for further discussion. If this is the only suitable reference there needs to be slightly more elaboration as to why. For instance, is the discussion point in the paper about representativeness but in the context of precipitation?

Although we understand the concern, this a point for RH that is also valid for any geophysical variables. We developed the sentence that now reads:

Lines 69 to 71: "These issues are not confined to the study of RH but are, to an extent, common to those of all geophysical variables (see for instance Kirstetter et al., 2020 for a discussion on precipitation)."

Line 93: A figure here might illustrate this point better for the channels on SAPHIR. Not all readers may be familiar with MW remote sensing, especially the 183 GHz region where the +/- values relate to where on the wings of the 183 GHz feature SAPHIR is sampling. Alternatively, the sentence could be updated to reflect this point and why it is done.

Thank you for this suggestion. We think that adding a figure may not be necessary for the presentation of the instrument. We added precision to the following sentence, which now reads:

Line 96 to 98: "SAPHIR spectrally samples the 183 GHz line with 6 channels ranging from 183.31 +/- 0.2 GHz (close to the center of the line for upper tropospheric sounding) to 183.31 +/-11GHz (wings of the line for a deeper sounding)"

Lines 96-108: Is the SAPHIR measurement noise (measurement uncertainty) used at all in the RH retrieval?

Yes, the in-flight radiometric noises were used in the training phase of the model, as discussed in Sivira et al, 2015 and Brogniez et al, 2016.

Line 115: "RH fields range between -5 and +5 % (resp. 5 and 25%)" what do the values in brackets relate to?

The values in brackets relate to the RMSE, we however reorganized the sentence to avoid any misunderstanding:

Line 126 to 128: "In the Tropics (30°N-30°S) and at this forecast range, the ARPEGE biases on RH fields range between -5 and +5 % and rmse between 5 and 25% with respect to both radiosondes and the ECMWF analysis (Chambon et al., 2014)."

Line 120: Does the vertical averaging account for SAPHIR weighting functions? – in a similar way to which upper tropospheric humidity is calculated?

No. As detailed in Sivira et al (2015), the weighting functions of SAPHIR have only been used during the design phase of the RH retrieval scheme, to determine the 6 atmospheric layers from the set of 6 measurements (6 BTs for every footprint). A UTH product is also available from the 3 upper channels of SAPHIR (see Brogniez et al, 2015, JAMC, DOI: 10.1175/JAMC-D-14-0096.1), but we are analyzing here the RH profiles defined on fixed atmospheric layers. Therefore, over the former lines 120-121 we match two profiles of RH: one with only 6 wide layers (RH from SAPHIR) and the other one with 18 thin layers (RH from ARPEGE). The vertical averaging in thus only performed to make ARPEGE match SAPHIR.

Line 133: do you mean uncertainty in a metrological sense? If not, you might want to change the word used. This is linked to the comment about lines 96-108.

Yes indeed, we replaced "uncertainty" with "shape" in the sentence. It now reads:

Lines 149 to 150: "The averaged PDF encompasses all the available information of the reference RH such as the mean (first moment), spread, shape, and extremes of the distributions."

Line 164: what is the uncertainty here? Source, magnitude? Or is it an error?
This paragraph has been reworked and it now provides more details about the uncertainties in the reference dataset. We added a sentence explaining the simplification and our choice to keep a common uncertainty value representing all layers:

Lines 179 to 181: "The 15% RH uncertainty value is the smallest that allows to encompass the uncertainties of all pressure layers (see paragraph 2.1 or Brogniez et al, 2016 for a more complete analysis of uncertainty)."

Line 192: I don't think you mentioned what you're a priori error assumption is before this point, what is it? Do you get an a-posteriori error? Do you calculate the error reduction?
We do not use an a priori error nor an a posteriori error. The term "a priori" is used to state that in the deterministic comparison there is an assumption made on the uncertainty and based on the characteristics of the retrieval scheme. In other words, assuming "a priori" that the error is Gaussian yields to bias the analysis.

Figure 5: Did 12:00 UTC look different? Is there any correlation to convection?
Each case shows different patterns but yes, there seems to be a correlation between convection and the patterns in the model's biases.

<u>3. Technical Comments</u>

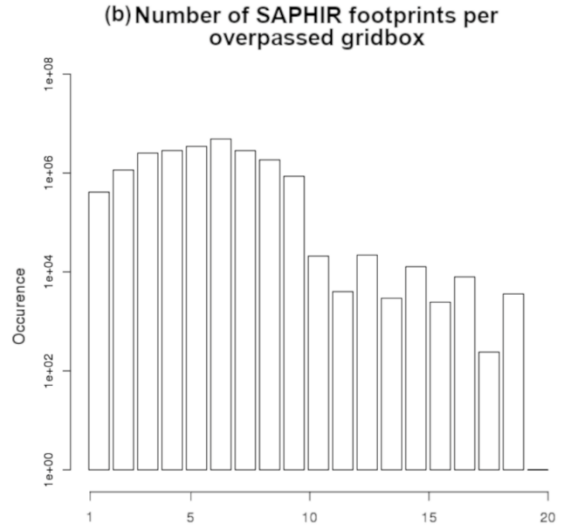
Line 17: ".The study first ..." – change to ". This study first ..."
Changed in the reviewed manuscript.

Line 18: "It warrants the need ..." - this sounds like you are eluding to a future direction in a conclusion. Would something more like "We demonstrate the need ..."
Changed in the reviewed manuscript.

Line 33: change "relies" to "rely"
Subject "The accuracy" is singular. "relies" stays as-is.

Line 72: "such a probabilistic approach." – missing 'a'
Changed in the reviewed manuscript.

Figure 1b: X axis label missing, also cannot see bars for values > 10, log scale might help here
The figures axis missing has been rectified and we used log scale in the updated figure. Figure 1b has been changed as follows:
Figure 1b:



Line 137: "complementarities" – change to similarities
Changed in the reviewed manuscript.

Lines 232-236: need a space between %RH, i.e. % RH. There is no need for a space between the value and the percent, e.g. 12% RH.
Thank you for the precision, this and all the following occurrences were modified.

Line 265: need a space between %RH, i.e. % RH

Line 280: need a space between %RH, i.e. % RH

Line 295: need a space between %RH, i.e. % RH

Figure 6: need a space between %RH, i.e. % RH

Figure 7: need a space between %RH, i.e. % RH

Lines 343-359: need a space between %RH, i.e. % RH

Line 412: need a space between %RH, i.e. % RH

Line 427: need a space between %RH, i.e. % RH

Lines 439-440: need a space between %RH, i.e. % RH