Reviewer #2:

This paper uses routine aircraft-based temperature and humidity measurements from the AMDAR and WVSS-II programs, respectively, to evaluate the operational thermodynamic retrievals from the IASI instruments aboard Metop-A and Metop-B. They find that the retrievals from the two satellites have almost identical biases relative to the aircraft obs. Furthermore, the evaluation of the satellite retrievals by day/night, viewing angle off nadir, and season show very small relative differences.

I find this paper very interesting, and very relevant for this journal. It is well written, the figures are clear and informative, and the conclusions generally well supported by the writing. I believe that this manuscript only needs a small amount of work to make it suitable for publication in EGUsphere. I have one "major" (really not too major) and several minor comments.

We thank the reviewer for the positive comments and their support of this manuscript for eventual publication.

Major comment:

Does the IASI team independently retrieve RH and SH (i.e., retrieve the two variables separately)? It seems that they must, because otherwise the mean biases don't make sense to me. I have two examples that I'll use to illustrate my confusion:

The IASI Level 2 algorithm retrieves SH. The values for RH used in this study are postprocessed from the retrieved T and SH by first converting the SH to vapor pressure and then obtaining the saturation vapor pressure via the Bolton formula:

$$e_s(T) = 611.2\ \exp\left(\frac{17.67\ T_c}{T_c + 243.5}\right)$$

where e_s is the vapor pressure in pascals and T_c is the temperature in Celsius (not Kelvin). This conversion was carried out by the team and is not a part of the routine IASI Level 2 output. We have added some text on this point to the manuscript to aid readers.

Fig 6: at 205 K, the bias in SH is about 0 (or slightly negative) and the bias in T is positive. For a given SH, a positive error in T would result in a decrease in RH, and similarly, decreasing SH should result in a decrease in RH. But for the temperature bin, the RH bias is +3 RH%. I know we are looking at means here, but the tails of the distribution (for that T) are even worse. This is inconsistent.

The three panels in Fig. 6 are independent. The top panel is showing bias in temperature as a function of temperature. The middle panel displays bias in SH as a function of SH, not temperature, and the RH panel shows bias as a function of RH, not as a function of the other variables. This was done in order to separate the analysis of the variables from the confounding impacts of altitude and pressure which are implicitly present in Figs. 2–5. Furthermore, the temperature statistics represent a much larger and more diverse dataset than the SH and RH observations do, as indicated in the maps in Fig 1. Therefore, quantitative comparisons between the various panels are not possible.

In Fig 2 at 1000 mb, the bias in T is zero and the bias in q is -0.8 – which seems to lead to a bias in RH of about 2 %RH. I did some back-of-the-envelope calculations, and for a range of T (from 10 to 35 C), and starting from two points (roughly RH=80% and RH=40%) and assuming that the aircraft data were about 0.8 g/kg drier than IASI (to give the bias shown in Fig 2), the resulting difference in RH ranges from -10 %rh to -2% RH. Since there is a year of observations from 2017 in this figure, I think that the mean RH bias should be much later than the 2% shown in Fig 2 if the q and RH retrievals from IASI are consistent with each other (I think the mean RH bias should be closer to -5 %rh). (see the attached figure on the next page)

In the example at 1000 hPa, an unbiased temperature and a negatively-biased SH should lead to a negatively-biased RH, which is what is observed. However, the relationship between temperature and relative humidity is nonlinear and asymmetrical. While the random temperature uncertainties may average to zero bias, those uncertainties may have a non-zero impact on the RH bias.

For example, let's assume conditions of 20 C, 11.7 g/kg, and 1000 hPa. This works out to a RH of 80.0% according to the Bolton formula described above; this looks to correspond well with the figure you included with your review. We did a quick monte carlo analysis in which we held the SH constant but perturbed the T by random values chosen from a normal distribution with a mean of 0 and a standard deviation of 2.4 (the uncertainty calculated at 1000 hPa in Fig. 2). After a million calculations, the mean RH was 81.0%. Despite unbiased temperatures going into the calculation and no change to the specific humidity, the net result was an RH that was biased moist by an entire percentage point. Such values partially offset the dry bias from the analysis you included with your review. The difference in scope between the airborne temperature and moisture datasets is an additional hindrance to these quantitative comparisons.

Anyway, I think the authors need to be more explicit about the consistency between the IASI retrieval of q vs RH (or if one is derived from the other), and if they are independently retrieved, to spend more time discussing the implications of this.

We do appreciate this concern, and we have added discussion on this issue to the text. We thank you for encouraging us to be more explicit and precise in our writing.

Minor comments:

We thank you for each of these suggestions, which significantly improve the clarity and readability of our work.

- Line 41: twice-daily revisits by a single satellite – this should be clarified

Clarified.

- Line 93: "first constituted with of the order of 10^8 of real IASI" – this is very awkward, and should be rewritten

Changed to "…a training base is constructed with over 100 million real IASI and AMSU/MHS observations collocated with model reanalysis data…"

- Line 149: a laser diode hygrometer

We have changed this to slightly different language based on the suggestion of Reviewer #2.

- Line 177: there is also significant displacement (and perhaps even more) for descents. The way this was written suggests that only ascents (not descents) were analyzed here, which I don't think was true. Please update

You are correct. The sentence has been updated.

- Line 273: does the shape of the vertical profile matter to the IASI retrieval? For example, in the daytime, the water vapor specific humidity is often pretty constant in the convective boundary layer, but that isn't true at night. This should at least be added to the text as a possible explanation for the small day/night differences in the bias

We have included the potential for shape to influence the retrieval accuracy.

- Caption of Fig 1 and Fig 5: you state "2019" when I am pretty sure you mean 2017

You are correct, and the captions have been updated.

- Caption of Fig 2: you state “AMDAR-minus-IASI” when all of the other results are the reverse. I believe this is a typo, based upon the results shown later

This, too, was a mistake on our part, and we thank you for catching that.