The authors have made many changes in line with the previous two reviews which improve the clarity and scientific accuracy of the paper. In particular, the addition of the instrument specifications, several clarifications about the algorithm and the comparison with the radiometer's standard retrieval. However, there remain several issues that I would recommend be corrected before publication.

It remains unclear to me why there are so many oscillations in the figure of the temperature retrieval bias, compared to the plots of mean temperature retrieval and mean temperature observations. Also, it is not clear exactly what is presented in the results with the humidity/temperture bias correction applied.

The discussion and conclusion section now seem to repeat many of the same points. Could this be condensed into a single section to avoid repetition? Also the conclusion has not been changed since the previous review round, and still makes claims such as the 'high precision' retrieval. Please bring this into line with the rest of the paper.

**Line 212**: Some things here are badly formatted. Equation 2 seems badly rendered and does not match the other equations. T(Z) is not well aligned.

**Line 213**: here you mention tau, but this does not feature in the equation. You integrate physical temperatures and absorption coefficients over a height, but the optical depth/opacity does not figure in this.

**Line 325**: "In this study, the adjacent-layer limits are set to $\delta 1 = 8$ K for temperature and $\delta 2 = 60\%$ for relative humidity."

When this is applied to all heights regardless of the grid resolution, I would suspect that it does not have much effect. Your highest resolution (smallest spacing) on the vertical grid is 25m, and as you only prevent jumps of 8K between adjacent levels, this only prevents jumps of more than +/- 320K/km (compared to typical tropospheric lapse rates of -6 to -7 K/km). I suspect that this limit is not having so much effect in the retrieval... I don't think that this can reasonably called a lapse rate constraint either.

**Line 415:** "The radiosonde profiles shown here represent ensemble-averaged means over 38 matched cases"

What is the meaning of an ensemble-average mean in the case of observations? What is the difference from a regular mean?

**Figure 7**: As previously mentioned, the zoom plots on here do not give any insight into the plots. For Temperature: either remove, or make the zoom in the first 2km of the plots where there are details that may not be fully appreciated by the reader. For the humidity plot: we see even less and it seems that there are fewer points in the zoom. I would advise that you remove the zoom from this plot. For plot c, the bias seems to cross the 0K lines at several points below 1km, but on the two mean plots the retrieved only exceeds the observed between ~1.3 and ~1.6km. However the bias is calculated, it should equal the difference between these two curves. This leads the reader to conclude that a) the bias was calculated from different data or b)additional processing was applied to one plot that wasn't on the other. Could the authors please explain this?

**Table 2**: Could you comment on why the RMSE goes up for attitude angles above 2.5°? Does this represent much of the dataset?

**Figure 9**: you state here ‘based on 38 valid matchups’, so is this figure made with all cases? If so then I am not sure how to interpret the systematic bias correction described in lines 357-360, where you state “The 38 collocated samples were randomly divided into a training set (80%) and an independent testing set (20%). The training set was used to derive the systematic bias correction model, while the testing set was used exclusively for independent validation of the correction performance.”

Shortly afterwards you state (line 366): “After the systematic error model is established using this method, it is applied to correct all retrieval profiles.”

Is the plot in figure 9 made by applying the bias correction, that was made with 30 out of 38 cases, to all 38 cases? If so I struggle to see the value of this. It would be better to compare the stats from the 8 testing cases before vs after bias correction. Even though it is not a big sample size, it is more meaningful.

**Figure 10**: It would be more informative to see the effects of constraints if the plot showed a retrieval with and without constraints (as well as observations). Now that it has been explained that the ‘lapse rate constraint’ simply did not allow 8K jumps, the plot as it is is not so informative.