Review of

Direct assimilation of ground-based microwave radiometer observations with machine learning bias correction based on developments of RTTOV-gb v1.0 and WRFDA v4.5

Overall, this paper is clearly written, including detailed descriptions of the experimental design. The analysis and presentation of the results are also generally clear. However, I have some concerns regarding the bias correction method and the way the experiments are compared. I would suggest that the authors address these points before the paper can be considered for publication.

Major comments:

1. The paper includes a focus on bias correction, but its impact on the analysis and forecast fields is not discussed in depth. While the improvements in GMWR diagnostics in observation space (e.g., Figs. 6–8) are somewhat expected, the more crucial aspect is how the bias correction affects the model space. It would strengthen the paper to include a comparison between assimilation experiments with and without bias correction for GMWR.

2. The bias correction approach based on offline O–B statistics essentially assumes that all biases originate from the observations. However, this assumption may not always hold, and such correction could potentially mask model bias (e.g., Auligné et al., 2007; Eyre, 2016), especially when more complex predictors or bias-prediction schemes are used to make the correction more expressive. Therefore, such offline bias correction that simply brings the O–B mean close to zero does not necessarily indicate a successful correction. It may be a sign of success, but could also be a result of compensating for model bias, rather than removing observation bias. Although in practice it remains difficult to fully separate the sources of O–B bias, I believe it is important for the paper to acknowledge this fundamental limitation of the current offline bias correction method based on O-B statistics.

Auligné, T., McNally, A. P., & Dee, D. P. (2007). Adaptive bias correction for satellite data in a numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *133*(624), 631-642.

Eyre, J. R. (2016). Observation bias correction schemes in data assimilation systems: A theoretical study of some of their properties. *Quarterly Journal of the Royal Meteorological Society*, *142*(699), 2284-2291.

3. Regarding the wind analysis and forecast, the results appear somewhat inconsistent depending on the diagnostic used. For example, Fig. 9 shows a degradation, while Figs. 10–11 indicate marginal improvements. However, the paper does not seem to acknowledge or discuss these discrepancies across different diagnostics. A brief discussion of these differences would help clarify the interpretation of the results.

Minor Comments:

1. Figure 3(b)(d): It would be useful to show the same scatter plot after bias correction to examine whether the two distinct clusters merge.

2. L233-235: Existing approaches can also address nonlinear relationship between the physical variables, e.g., skin temperature (TS), and the bias since the selection of predictors can be completely general. E.g., consider predictors $p_1 = TS,\ p_2 = (TS)^2,\ p_3 = (TS)^3,\ p_4 = (TS)^4 \ldots$

3. L326-327: Since Figure 8 does not show the CTRL results, it is difficult to determine whether the assimilation of GMWR really improves the fit (even though such improvement is expected)

4. L480-481: This may not be true, as discussed in the major comment (2).

5. L498-499: A larger STD in the K-band compared to the V-band does not necessarily imply that the model's humidity accuracy is worse than its temperature accuracy. First, it is inherently difficult to directly compare the accuracy of humidity and temperature fields. Second, the brightness temperature STD also depends on its sensitivity to temperature and humidity. For example (using hypothetical numbers), a 1K change in K-band may correspond to 1% change in humidity, but 1K change in V-band may correspond to a much larger 10% change in temperature.

6. Overall, the paper includes a large amount of numerical detail (e.g., bias reductions by a few degrees or a certain percentage). If some of these values are already shown in the figures, I believe it is not necessary to restate all of them in the text. Instead, the paper could focus on highlighting the meaning and implications of these numbers. This would help make the manuscript more concise and easier to follow.