Major Comments:

1. I am wondering whether the interpretation of Fig. 5 is fully consistent with the statement in the manuscript. Specifically, Fig. 5a appears to suggest that the differences between the TCKF1D-Var and ERA5 mainly occur above 600 m, whereas only minor differences are evident within the boundary layer. Similarly, Fig. 5e shows nearly no differences between the TCKF1D-Var and ERA5. These results do not seem to be consistent with the statement that "for temperature mean bias, the differences between the TCKF1D-Var and the ERA5 are mainly confined to the boundary layer." Clarification or further explanation would be helpful.

Re:

Thank you very much for this insightful comment. We agree that the original wording in the manuscript was not sufficiently clear and may have caused confusion when interpreted together with Fig. 5 (now Fig. 6). Our original intention was to emphasize that, in plain language, for temperature mean bias, the differences between TCKF1D-Var and ERA5 are pretty much limited within (mostly confined to) the boundary layer. To address this issue and improve clarity, we have revised the corresponding sentence in the manuscript. It now reads: "For temperature mean bias, the differences between the TCKF1D-Var and ERA5 are predominantly limited within the boundary layer, while detectable improvements are found above 3000 m above ground level." (**Line 243 – 245**)

2. In the response, the figure reference may need to be corrected to "Figure 5f" instead of "Figure 6f". Furthermore, the term "comparable levels" is somewhat unclear, and it might be helpful to replace it with a more explicit description (e.g., "slightly higher").

Re:

We thank the reviewer for the careful check and helpful suggestion. Concerning the figure reference, we would like to explain that the figure numbering was updated during the revision process in response to the comments from Reviewer #1.

Specifically, an additional figure entitled "Figure 4. Schematic of the coupling between the thermodynamic constraint and the WSM3 single-moment microphysics scheme" was inserted after the original Figure 3, which led to the renumbering of subsequent figures. As a result, the correct reference in the revised manuscript is Figure 6f.

We also appreciate the suggestion regarding the wording. The term "comparable levels" was indeed not sufficiently clear, and we have revised the sentence accordingly to provide a more explicit description. The revised sentence now reads:
"However, at night (Figure 6f), the random errors of the TCKF1D-Var temperature profiles increase substantially above 8500 m, changing from being comparable to ERA5 during daytime to slightly higher than those of ERA5." (**Line 250 – 251**)

3. Thank you for your detailed reply. However, perhaps due to my previous imprecise wording, I would like to clarify that the relatively large errors below 500 m are not confined to the 1D-Var results. Elevated errors are also evident in ERA5 and TCKF1D-Var, as shown in Figs. 4b–4d, 5a–5d, 5f, and 5h. In some cases (e.g., Fig. 5e), the error increases upward from near zero, whereas in the other cases listed above, the error decreases from relatively large values near the surface. This contrast represents another interesting feature that merits further explanation. In addition, my primary concern previously was that the low errors of TCKF1D-Var relative to observations may partly arise from an inherently unfair comparison. Given that TCKF1D-Var does not explicitly incorporate background (B) or observation (R) error covariance matrices and relies strongly on the observations themselves, it is expected that low errors would be obtained if the GMWR observations are sufficiently accurate (e.g., comparable to radiosonde measurements). However, I now realize that this characteristic may instead highlight an important advantage of TCKF1D-Var, namely its potential as an effective alternative in regions where GMWR observations are available but radiosonde data are sparse or absent.

Re:

We deeply appreciate the further clarification and for highlighting this interesting feature. We agree that the relatively large errors below 500 m are not confined to the 1D-Var results, but are also evident in ERA5 and TCKF1D-Var, as shown in the figures you referenced. This behavior reflects a common characteristic of near-surface temperature retrievals rather than a limitation of a specific method.

From the perspective of the GMWR measurement principle, elevated errors in the lowest atmospheric layers are expected. GMWR observations represent vertically integrated brightness temperatures, and the weighting functions of the oxygen and water vapor absorption channels exhibit strong overlap near the surface. As a result, the vertical resolution and information content below several hundred meters are inherently ill-defined. In addition, the diurnal cycle of surface emission, reflection effects, and rapidly varying thermodynamic conditions within the surface layer under different weather conditions further contribute to increased uncertainties. These factors affect all retrieval-based products, not only 1D-Var but also the TCKF1D-Var.

Regarding the contrasting vertical error structures, the decreasing error with height can be attributed to the diminishing surface influence, in cases such as Fig. 5e, where errors increase upward from near-zero values, this behavior likely reflects reduced surface emission and reflection effects at nighttime (12:00 UTC, approximately 20:00 local time). The coexistence of these two patterns therefore results from the combined effects of surface influence, atmospheric stability, and the height-dependent information content of MWR observations.

We agree that, because TCKF1D-Var relies primarily on information from ground-based microwave radiometer observations and does not explicitly prescribe background (B) or observation (R) error covariance matrices, relatively low errors with respect to independent observations may be obtained when the GMWR measurements are of high quality. At the same time, this aspect may raise concerns regarding the interpretation of the comparison results under certain conditions. Moreover, we acknowledge the point that this characteristic also suggests a potential role for TCKF1D-Var in situations where radiosonde observations are limited or unavailable. In this sense, TCKF1D-Var may serve as a complementary approach rather than a replacement, particularly in observation-sparse regions.

4. What I pointed out is that in the manuscript the mean bias of water vapor at~1700 m appears larger than the RMSE, which is mathematically implausible given the standard definitions: for errors =model - observation, the root mean square error (RMSE) and mean bias (MB) satisfy RMSE = $\sqrt[2]{\overline{e^2}}$, MB = $\overline{e}$, and therefore RMSE ≥ | MB (because $\overline{e^2} - \overline{e}^2$ = Var(e) ≥ 0). If the plotted MB exceeds the RMSE, that suggests an inconsistency. Please check and clarify the corresponding analysis.

Re:

Thank you very much for pointing out this important issue and for carefully examining the statistical consistency of the results. We agree that, by definition, the root mean square error (RMSE) should always be greater than or equal to the absolute value of the mean bias (|MB|), and we appreciate your attention to this detail.

After carefully re-examining the results, we confirm that the situation you noted around approximately 1700 m does indeed occur. However, this feature is present in the ERA5 reanalysis and the conventional 1D-Var retrievals, rather than in the TCKF1D-Var profiles. The TCKF1D-Var results consistently satisfy the expected statistical relationship between RMSE and MB.

Further analysis of Figure 7 indicates that the cases in which this behavior appears are primarily associated with clear and cloudy conditions for both ERA5 and the 1D-Var retrievals. For ERA5, this behavior can be attributed to limitations related to its relatively coarse horizontal resolution, which affects the representation of non-precipitating clouds and associated water vapor structures in the lower to middle troposphere (e.g., Prange et al., 2023; Virman et al., 2021; McDonald et al., 2025). These limitations can introduce systematic biases that dominate the error statistics at certain altitudes.

For the 1D-Var method, the background error covariance matrix is constructed based on the statistical differences between ERA5 reanalysis and radiosonde observations. Consequently, the aforementioned deficiencies of ERA5 are implicitly propagated into the 1D-Var retrievals through the background constraint, leading to similar bias characteristics around 1700 m.

In contrast, no such inconsistency is observed in the TCKF1D-Var profiles. This indicates that the use of a thermodynamic constraint based on virtual potential temperature conservation, together with the coupling to a cloud microphysical parameterization scheme, provides a robust framework for water vapor retrievals under both clear and cloudy conditions.

Reference:

Prange, M., Buehler, S. A., and Brath, M.: How adequately are elevated moist layers represented in reanalysis and satellite observations? Atmospheric Chemistry and Physics, 23, 725–741, https://doi.org/10.5194/acp-23-725-2023, 2023.

Virman, M., Bister, M., Räisänen, J., Sinclair, V. A., & Järvinen, H.: Radiosonde comparison of ERA5 and ERA-Interim reanalysis datasets over tropical oceans. Tellus A: Dynamic Meteorology and Oceanography, 73(1), 1929754, https://doi.org/10.1080/16000870.2021.1929752, 2021.

McDonald, A. J., Kuma, P., Panell, M., Petterson, O. K. L., Plank, G. E., Rozliaiani, M. A. H., & Whitehead, L. E.: Evaluating cloud properties at Scott Base: Comparing ceilometer observations with ERA5, JRA55, and MERRA2 reanalyses using an instrument simulator. Journal of Geophysical Research: Atmospheres, 130, e2024JD041754, https://doi.org/10.1029/2024JD041754, 2025.

Minor Comments:

1. The accuracy information for CPR_CLD_2A I recommend is the retrieved hydrometeor profile errors compared to the ground observations such as radiosonde.

Re:

Thank you for this valuable suggestion. We fully agree that independent ground-based observations would be highly desirable for evaluating the accuracy of the retrieved hydrometeor profiles in CPR_CLD_2A.

We initially attempted to validate the retrievals using routine ground-based observations. However, conventional radiosonde measurements do not provide observations of cloud liquid water content or cloud ice water content, which prevents a direct quantitative comparison with the retrieved hydrometeor mass concentration profiles. Although Ka-band millimeter-wave cloud radars are deployed at some sites, these instruments in our study region currently provide primarily basic reflectivity measurements and do not routinely produce quantitative cloud liquid or ice water content products. This limitation makes a robust and consistent validation against ground-based observations challenging.

Therefore, we acknowledge that the specific validation requested cannot be fully addressed within the scope of the present study. We have clarified this limitation in the revised manuscript. In future work, we plan to mitigate this shortcoming by extending the temporal coverage of the dataset and incorporating observations from a larger number of sites, as well as by exploring additional synergistic ground-based measurements where available.