I acknowledge the authors' prompt responses to the comments. However, there are still some outstanding questions that merit further discussion.

Major Comments:

1. Line 211: In the figure, it is not clear that "the differences among the three products are mainly confined to the boundary layer." The authors may clarify whether this statement is supported by the figure or revise it accordingly. Lines 217–219: In Fig. 5f, the random errors of the TCKF1D-Var temperature profiles remain significantly smaller than those of 1D-Varabove 2000 m, which seems inconsistent with the statement "increase substantially" in the text.

   Reply:

   We sincerely apologize for the confusion caused by our original wording. We agree that the previous statement may have led to a misinterpretation and did not accurately reflect what is shown in the figure. The sentence has been revised to: "For temperature mean bias, the differences between the TCKF1D-Var and the ERA5 are mainly confined to the boundary layer." (Line 242–243)

   Re:

   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.

   Reply:

   We thank the reviewer for pointing out the inconsistency between the text and the results in Fig. 5f. We apologize for the misinterpretation created by the earlier description. The corresponding statement has been corrected to: "However, at night(Figure 6f), the random errors of the TCKF1D-Var temperature profiles increase substantially above 8500 m, shifting from being equivalent to ERA5 during daytime to comparable levels." (Line 248–250)

   Re:

   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").

2. In the overall comparison among TCKF1D-Var, 1D-Var, and ERA5, it appears that temperature differences between TCKF1D-Var and ERA5 are generally small, whereas 1D-Var exhibits large errors in the upper atmosphere, and below 500 m errors remain high. These issues seem unresolved and warrant further discussion. For the first two points, could they be attributed to the dependence on the R and B matrices? If so, it may be slightly unfair to generalize, as in other cases (with different R and B) 1D-Var might perform better. Moreover, since the new cost function depends explicitly on GMWR observations, it is unsurprising that TCKF1D-Var outperforms 1D-Var when the

distance between radiosonde and GMWR is minimal. Clarifying these points would be helpful.

Reply:

We deeply appreciate the reviewer for this thoughtful comment and fully agree that the performance of 1D-Var is closely linked to the specification of the background (B)and observation (R) error covariance matrices. We acknowledge that the temperature differences shown in the manuscript—particularly the larger upper-level deviations in 1D-Var and the persistent errors below 500 m—may partially reflect the characteristics of the chosen B and R matrices. Our intention is not to suggest that1D-Var is inherently inferior; indeed, from the standpoint of computational efficiency and operational robustness, 1D-Var offers clear advantages over the proposedTCKF1D-Var method. The central objective of this study is instead to introduce an alternative retrieval approach that satisfies dynamical constraints and incorporates microphysical parameterizations, providing the community with a complementary option beyond 1D-Var. The comparisons included in this work aim to demonstrate that the performance of TCKF1D-Var can reach or exceed that of 1D-Var under the tested conditions.

We also recognize the reviewer's concern regarding regional representativeness. The evaluation sites in northern China do have localized meteorological characteristics, and we do not exclude the possibility that 1D-Var may outperform TCKF1D-Var in other regions — particularly those with weaker water vapor variability or less baroclinicity. While alternative atmospheric profile references exist, radiosonde observations remain the widely accepted standard for directly probing upper-air thermodynamic structure. Thus, using radiosondes as verification data is essential. To mitigate the limitation of sparse co-located radiosonde stations, we additionally compared both retrieval methods against ERA5 reanalysis. The combined results indicate that TCKF1D-Var can effectively exploit GMWR observations to improve the thermodynamic structure relative to ERA5, especially in the lower and middle troposphere.

A synthesized clarification addressing these points has been added to the "Summary and Concluding Remarks" section of the revised manuscript (Line 431-443) to ensure that the discussion is transparent and balanced, reading as: "We also acknowledge that the performance of the classical 1D-Var approach is inherently shaped by the prescribed background (B) and observation (R) error covariance matrices, and the differences highlighted in this study should not be interpreted as a universal limitation. Rather than positioning TCKF1D-Var as a replacement for 1D-Var, our intention is to provide a complementary retrieval framework that incorporates moist-thermodynamic constraints and a microphysical closure, features that are not explicitly represented in the classical formulation. The evaluation sites in North China exhibit regional characteristics, and it is fully plausible that in regimes with weaker humidity gradients or reduced baroclinicity, 1D-Var may perform similarly or even more favorably. Radiosonde observations remain an essential benchmark for upper-air thermodynamic verification, and to address the limited availability of co-located soundings, additional comparisons with ERA5 were included. Overall, the combined evaluation suggests that TCKF1D-Var can extract additional thermodynamic information from GMWR measurements and thus serves as a useful complement to existing 1D-Var techniques under the conditions examined. These considerations have been incorporated to ensure that the inter-method comparison is presented within a balanced and context-appropriate framework."

Re:

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.

3. Only seven sites are equipped with radiosonde observations. Therefore, in composite analysis, large differences may arise between mean bias and RMSE for the same variable. For instance, in Figs. 4c and 4d, the mean bias of water vapor at ~1700 m appears larger than the RMSE, which is mathematically implausible. A similar issue occurs between Figs. 5gand 5h. Please check these results.

   Reply:

   We deeply appreciate the reviewer for pointing out this issue. To address the concern, we have added confidence intervals to Figure 4 (now Figure 5) and Figure 5(now Figure 6) to better illustrate the variability of the statistics. However, we acknowledge that this addition alone does not fully resolve the specific concern regarding the apparent discrepancy between mean bias and RMSE at certain altitudes.

   We also recognize that the evaluation was conducted using radiosonde observations from seven sites on 7 July 2025, with two launches per day, yielding a total of 434measurements. While this satisfies the traditional statistical definition of a large sample, it remains insufficient to fully address variations under certain conditions, for example around ~1700 m above ground level, where the retrieval results exhibit relatively large fluctuations. In future work, we plan to increase the sample size for radiosonde verification to improve statistical robustness under such specific conditions.

   From another perspective, although the TCKF1D-Var results show fluctuations in retrieval accuracy around ~1700 m, the classical 1D-Var maintains relatively stable performance in this layer. This observation is consistent with the reviewer's suggestion that additional attention is required in this altitude range and supports the value of carefully interpreting the variability seen in limited-site composites.

Re:

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 e_i =model - observation, the root mean square error (RMSE) and mean bias (MB) satisfy RMSE $= \sqrt{\overline{e^2}}$, MB $= \overline{e}$, and therefore RMSE ≥ | MB| (because $\overline{e^2} - \overline{e}^2 = \mathrm{Var}(e) \geq 0$). If the plotted MB exceeds the RMSE, that suggests an inconsistency. Please check and clarify the corresponding analysis.

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.