Summary statement:

This manuscript comes up with a new method, TCKF1D-Var, to retrieve atmospheric profiles from GMWR. This method differs from conventional 1D-Var that employs a ratio-based cost function independent of prescribed background and observation error covariances, and integrate a diagnostic microphysics closure to represent liquid and ice water.

Clear results are shown that lower errors are achieved in cloudy circumstances. However, there are several issues need to be answered before acceptance. I recommend major revision.

Reply:

We sincerely thank the reviewer for the thoughtful and constructive evaluation of our manuscript. We deeply appreciatethe reviewer's recognition of the novelty of the proposed TCKF1D-Var method, particularly its ratio-based cost function and the integration of a diagnostic microphysics closure to represent cloud liquid and ice water. We are also grateful for the positive acknowledgement of the improved retrieval performance in cloudy conditions.

We fully agree with the reviewer that several important issues require clarification and further analysis before the manuscript can be considered for acceptance. In the revised version, we have carefully addressed all concerns raised in the detailed comments.

We deeply appreciatethe reviewer's recommendation for major revision and have substantially improved the manuscript accordingly. We hope that the revised version satisfactorily addresses all concerns and meets the standards required for publication.

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-Var above 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**)

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**)

2. In Section 4.1.4, the potential cause of the water vapor RMSE deficit is analyzed. As a reviewer not specialized in this field, I would deeply appreciateif the authors could clarify why a similar phenomenon does not occur in the temperature vertical profile?

Reply:

We deeply appreciate the reviewer for this insightful question. We deeply appreciatethat, for readers not specialized in microwave radiative transfer or retrieval physics, the asymmetric behavior between humidity RMSE and temperature RMSE may not be immediately intuitive.

To clarify this, we have added an explicit explanation in Section 4.1.4. The key reason a similar RMSE degradation does not occur in the temperature vertical profile is that the temperature retrieval is fundamentally less susceptible to cross-variable leakage for the following reasons: **1.** Temperature channels dominate the information content and have stronger, better-separated weighting functions. The oxygen absorption lines provide well-defined temperature weighting functions that peak at multiple altitudes. These functions overlap less with water-vapor-sensitive channels than vice-versa, which reduces the propagation of humidity-related uncertainties into the temperature retrieval. **2.** Humidity channels have only weak indirect sensitivity to temperature. Although humidity weighting functions are affected by temperature through absorption line broadening, this dependency is substantially weaker compared with the strong temperature sensitivity of oxygen channels. As a result, humidity uncertainties do not significantly contaminate the temperature solution. **3.** Temperature retrieval is more strongly constrained by measurements. The oxygen channels used for temperature have higher signal-to-noise ratios and stronger Jacobians. Their cumulative information content allows the inversion to remain well-conditioned, preventing RMSE amplification in the mid-troposphere. **4.** Humidity retrieval suffers more from vertical-resolution mismatch. The mismatch between humidity and temperature weighting functions (1.5–4.5 km) induces leakage primarily from temperature into humidity, not the other way around. This asymmetry is consistent with prior microwave retrieval theory (Hewison, 2007; Löhnert & Maier, 2012).

To address the reviewer's concern, we have added the following explanation to Section 4.1.4 in the revised manuscript: "A similar RMSE degradation does not occur in the temperature profiles because the oxygen absorption channels provide stronger and more vertically distinct weighting functions, which dominate the temperature information content and are only weakly affected by humidity-related uncertainties. In contrast, water-vapor-sensitive channels exhibit weaker constraints and overlapping sensitivity with temperature weighting functions in the 1.5 – 4.5 km layer, making humidity retrievals more vulnerable to temperature – humidity cross-talk." (**Line 300-306**)

3. 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 that 1D-Var is inherently inferior; indeed, from the standpoint of computational efficiency and operational robustness, 1D-Var offers clear advantages over the proposed TCKF1D-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 favourably. 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."

4. 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. 5g and 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 434 measurements. 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.

5. What criterion is used for the histogram bins in Fig. 9? Are the results sensitive to the division of water vapor ranges?

Reply:

We deeply appreciate the reviewer for raising the important question regarding the binning criterion used in Fig. 9. To enhance transparency and ensure that the histogram-based comparison of cloud liquid water content (CLWC) reflects physically meaningful statistical distributions, we followed the studies (Zhang et al., 2021; Mroz et al., 2023) whic have used histogram and probability–density analyses for LWC and related microphysical parameters. Accordingly, for Fig. 9 (now as Figure 10) we grouped the CLWC retrievals and reference observations into a set of bins whose boundaries were chosen to (1) capture the main modes of the LWC/CLWC distribution and (2) align with typical value ranges reported in cloud microphysics literature, ensuring that each bin contains a sufficiently large sample for meaningful comparison. We have added a statement in the revised manuscript to clarify this, for example in the Methods section: "Histogram bins in Figure 10 are defined to ensure sufficient sample counts in each interval for robust frequency comparisons, following established practice in cloud-microphysics statistical analyses (Zhang et al., 2021; Mroz et al., 2023)." (**Line 333–334**)

To the question "Are the results sensitive to the division of water vapor ranges?" Our current study does not explicitly investigate the sensitivity of the results to the division of water-vapor ranges. Based on the limited physical understanding relevant to our present framework, cloud liquid water content is generally positively correlated with the ambient water-vapor abundance, suggesting that some dependence on humidity partitioning may indeed exist. However, a rigorous quantification of this sensitivity lies beyond the scope of the present analysis. We fully agree that this is an interesting and meaningful direction, and we will explore the relationship between environmental water vapor stratification and cloud-water retrieval performance in our future work.

Reference:

Mroz, K., Treserras, B. P., Battaglia, A., Kollias, P., Tatarevic, A., and Tridon, F.: Cloud and precipitation microphysical retrievals from the EarthCARE Cloud Profiling Radar: the C-CLD product, Atmos. Meas. Tech., 16, 2865 – 2888, https://doi.org/10.5194/amt-16-2865-2023, 2023.
Zhang, Y., Chen, S., Tan, W., Chen, S., Chen, H., Guo, P., Sun, Z., Hu, R., Xu, Q., Zhang, M., Hao, W., and Bu, Z.: Retrieval of Water Cloud Optical and Microphysical Properties from Combined Multiwavelength Lidar and Radar Data, Remote Sens., 13(21), 4396. https://doi.org/10.3390/rs13214396, 2021.

Minor comments:

1. Lines 75 and 78: The abbreviations should be unified—either GMWR or MWR.

Reply:

We thank the reviewer for pointing this out. We have corrected the inconsistency: the abbreviation has been unified to "GMWR" throughout the manuscript (including the instances at **Line 75–78**). The change is reflected in the revised manuscript and in the tracked-changes file.

2. In Figure 1, there are 44 stations, which does not match "43" in Line 75.

Reply:

We thank the reviewer for noting this. The mistyping has been corrected, and the abbreviation has now been consistently unified throughout the manuscript.

3. In Section 2.3, it is recommended to include the accuracy information of CPR_CLD_2A.

Reply:

We have added the accuracy information of the EarthCARE CPR_CLD_2A product in Section 2.3. The revised manuscript now states (**Lines 104–105**): “The active radar observations from EarthCARE provide vertically resolved cloud liquid and ice water content with high sensitivity to optically thick clouds, with a target radar reflectivity accuracy better than 2.7 dB.”

4. What is the underlying reason for the larger differences between TCKF1D-Var and ERA5 during daytime?

Reply:

We thank the reviewer for this insightful question. The manuscript has been revised to clarify the underlying mechanism. In brief, the larger daytime differences between TCKF1D-Var and ERA5 primarily arise from enhanced boundary-layer instability and stronger diurnal variability, which amplify temperature–humidity coupling in passive microwave retrievals. During daytime, solar heating intensifies turbulent mixing and increases the vertical heterogeneity of temperature and water vapor. As a result, TCKF1D-Var — being driven directly by GMWR observations and thermodynamic constraints — responds more strongly to these rapidly evolving features, whereas ERA5 tends to represent smoother background structures due to model diffusion and data-assimilation temporal averaging. This contrast naturally leads to larger daytime departures between the two. Similar daytime–nighttime contrast in vertical thermodynamic heterogeneity and radiometer information content has also been documented in boundary-layer physics and microwave-retrieval studies (Löhnert et al., 2012). Hence, enhanced daytime instability provides a physically consistent explanation for the larger differences between TCKF1D-Var and ERA5 observed in our results.

Reference:

Löhnert, U. and Maier, O.: Operational profiling of temperature using ground-based microwave radiometry at Payerne: prospects and challenges, Atmos. Meas. Tech., 5, 1121–1134, https://doi.org/10.5194/amt-5-1121-2012, 2012.

5. Lines 237–239: It seems that Figs. 6i and 6m (rather than 6n) are being analyzed. Moreover, the statement “TCKF1D-Var also exhibits reduced temperature errors below 5 km compared to ERA5 and 1D-Var, while above 5 km its performance is comparable to 1D-Var” corresponds to Fig. 6i, and “ERA5 shows similar errors to 1D-Var below 3 km but becomes less accurate above this level” corresponds to Fig. 6m. Please separate these analyses to avoid confusion.

Reply:

We fully agree with your comment regarding the mismatch in figure references and the mixing of two separate analyses. Following your suggestion, we have revised the text accordingly and separated the two discussions to avoid ambiguity. The corrections have been implemented in the revised manuscript at Line 269.

6. Please ensure that the title and content of Table 1 appear on the same page.

Reply:

We have adjusted the layout, and the table title and the table now appear on the same page in the revised manuscript (**Line 320**).

7. Correct the repeated “Figure 8” in the title of Fig. 8.

Reply:

We deeply appreciate the reviewer for pointing this out. The duplicated Figure 8 has been deleted (**Line 329**).

8. Although Taylor et al. (2007) and Garcia-Carreras et al. (2010) are cited to justify using the temporal moving anomaly of virtual potential temperature as an early-warning indicator, it is recommended to briefly clarify the underlying mechanism.

Reply:

We deeply appreciate the suggestion. We have clarified the underlying mechanism for using the temporal moving anomaly of virtual potential temperature as an early-warning indicator. In the revised manuscript (**Lines 355–357**), the text now reads: "Following the approach proposed by Taylor et al. (2007) and Garcia-Carreras et al. (2010), we adopt the temporal moving anomaly of virtual potential temperature as an early-warning indicator, which removes slowly varying background signals associated with large-scale processes and diurnal variations."

9. The results in Figs. 10 and 11 are somewhat repetitive. It is recommended to either combine these figures and the corresponding analysis, or present the results without Fig. 11 for conciseness.

Reply:

We deeply appreciate the suggestion regarding the potential redundancy between Figures 10 and 11. In response, Figure 11 has been moved to Appendix A ("Sensitivity of Virtual Potential Temperature Anomaly to Temporal Averaging Window") and is now labeled as Figure A1.

Correspondingly, the related discussion (lines 348–356) has been rewritten for clarity and conciseness. The revised text now reads: "Using the same methodology, we recalculated the time–height evolution of the virtual potential temperature anomaly with a reduced temporal averaging window (Figure A1 in Appendix A), and the gradients of the anomaly variations become weaker compared to those in Figure 11, owing to the shorter averaging window. Nevertheless, both ERA5 and TCKF1D-Var profiles still exhibit the characteristic transition of the anomaly from positive to negative about 7–8 h prior to rainfall onset. Although the warm anomaly tongue intrusion remains detectable in both products, its intensity is reduced. When adopting −0.75 K as the early-warning threshold, the signal becomes indistinct under the 4.5-hour averaging window, whereas it is enhanced and temporally stabilized within about 2 hours of the precipitation onset when using 6.0-hour and 7.5-hour windows. Consistent with the previous findings, the 1D-Var (Figure A1 c, f, and i) profiles fail to extract effective early-warning signals for heavy rainfall."