Reviewer 2:

Summary

This study presents a methodology for characterizing vertical heterogeneities in aerosol type/properties. It then quantifies how these heterogeneities impact the accuracy of vertically resolved CCN from the RNCCN method via comparisons with ground-based and airborne in situ CCN observations. This work demonstrates that the vertical aerosol uniformity assumption underlying the RNCCN retrieval method can cause large errors in certain conditions. The method is also presented as a way to enhance or better use/interpret the RNCCN product. The paper is very well written, and I appreciated that the hypothesis and the main goals of the paper were presented up front in the paper and were explained very well. I also felt the data section was written very well and was very easy to follow. I have some minor comments for clarity of various details.

Thank you for the reviewer's positive feedback.

Minor Comments

• Introduction

o Line 58: Somewhere it might be good to mention aircraft-based lidar approaches as well as a third category.

As suggested, we have revised the paragraph to distinguish airborne lidar as a third category of CCN-relevant lidar approaches, complementing the spaceborne and ground-based categories.

"Airborne lidar provides vertically resolved aerosol profiles along flight tracks. For example, airborne high-spectral-resolution lidars (HSRL) directly measure extinction and depolarization, enabling aerosol-type classification and CCN retrieval without assumptions required by elastic-backscatter systems (Burton et al., 2012). Lenhardt et al. (2023) used airborne HSRL extinction collocated with in situ CCN to derive empirical CCN relationships for biomass-burning aerosol over the southeast Atlantic."

o Line 70: Are there other papers in addition to Ghan et al. (2004) that could be cited as examples of this method? This is one recent example I'm aware of that also uses ARM observations: Chen, B., Thompson, S. A., Matthews, B. H., Sharma, M., Li, R., Nowotarski, C. J., Rapp, A. D., and Brooks, S. D.: A new technique to retrieve aerosol vertical profiles using micropulse lidar and ground-based aerosol measurements, Atmos. Meas. Tech., 18, 5841–5859, https://doi.org/10.5194/amt-18-5841-2025, 2025.

We thank the reviewer for the reference. We have added Chen et al. (2025) as a recent example that uses ARM micropulse lidar to retrieve aerosol vertical profiles.

o Line 101: Other recent ML + lidar papers: Redemann & Gao (2024); Ferrare et al. (2025)

Added the suggested two papers. Also, citied these two: del Águila et al., 2025 and Peleg et al., 2025

Peleg, Y., Zeida-Cohen, L., Tzror, I., Bühl, J., Ansmann, A., Chudnovsky, A., and Yakhini, Z.: Cloud Fields and Aerosol Classification with Lidar using Advanced AI Approach, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-5215, 2026.

del Águila, A., Ortiz-Amezcua, P., Tabik, S., Bravo-Aranda, J. A., Fernández-Carvelo, S., and Alados-Arboledas, L.: Aerosol type classification with machine learning techniques applied to multiwavelength lidar data from EARLINET, Atmos. Chem. Phys., 25, 12549–12567, https://doi.org/10.5194/acp-25-12549-2025, 2025.

• Data

o A brief description in this section of the SGP location and what aerosol types/meteorological conditions were observed during the HI-SCALE campaign would provide the reader with a bit more context.

As suggested, we have added a brief description in the revision:

"The SGP site is situated in a rural, mid-latitude continental setting surrounded by heterogeneous land use, including cultivated crops, pastures, grasslands, and forested regions. The campaign comprised two intensive observational periods (IOPs): a spring IOP (24 April–21 May 2016) and a summer IOP (28 August–24 September 2016). Although the SGP site is often regarded as a relatively clean continental location, the aerosol population sampled during HI-SCALE was complex and originated from a mixture of anthropogenic, biogenic, and biomass burning sources (Fast et al. 2019). Meteorological conditions during the campaign were variable. Larger aerosol concentrations were generally associated with warm, mostly sunny conditions and southerly winds that favored photochemistry and the transport of anthropogenic emissions from Oklahoma City and urban areas in Texas, whereas the lowest concentrations occurred under northerly or westerly flow that advected cleaner air over the site (Fast et al. 2019). Although HI-SCALE targeted shallow convective clouds, the campaign also sampled clear-sky days, more complex multi-layer cloud populations, and periods preceding or following deep convection (Fast et al. 2019)."

o What are the approximate average uncertainties for the Raman lidar observables and in situ CCN concentration?

Uncertainties are added in the revision:

"A direct evaluation of the RLPROF-FEX aerosol extinction profiles against in situ measurements is unavailable. However, comparisons of RLPROF-FEX aerosol optical depth with sun-photometer measurements at the SGP site yielded relative Root Mean Square Error (RMSE) values of approximately 32%–43% and systematic biases of about -3% (Thorsen and Fu, 2015)."

"A CCN closure analysis (Kulkarni et al, 2023a) showed that the in situ CCN errors are generally within 20%."

o Line 189: It is mentioned that the RNCCN dataset excludes observations with RH>85%, such that hygroscopic growth corrections are not included. While (vertical) variations in aerosol type are one major cause of CCN retrieval uncertainty, hygroscopic growth also increases error when retrieving CCN from optical properties. Although not a focus of this study, it may strengthen the paper to add some mention of this as a limitation/source of uncertainty/possible area for future work in the summary, especially since this threshold will exclude observations in very humid near-cloud environments that are relevant for ACI.

We thank the reviewer for this point. We would like to clarify that the RNCCN product does apply a hygroscopic-growth correction, converting ambient to dry extinction via an f(RH) humidification factor. The RH > 85% threshold is applied because this correction becomes large and uncertain near saturation. We agree this is an important limitation, since it excludes humid, near-cloud environments relevant to ACI. In the revision, we have added clarifying statements to the Summary, noting that improved high-RH humidification parameterizations could extend the retrieval into these conditions.

"In addition to vertical variations in aerosol type, humidity-dependent aerosol growth is another source of CCN retrieval uncertainty: as RH approaches saturation, water uptake increases aerosol

extinction and complicates the extinction-to-CCN relationship. The RNCCN product accounts for this effect by converting ambient to dry extinction via an f(RH) correction and excludes observations with RH > 85%, where this correction becomes increasingly uncertain. Improved high-RH humidification parameterizations could broaden the applicability of RNCCN."

• Methodology

o Line 218: On first read, I interpreted the DeLiAn database as a lookup-table approach, which made the role of the kNN classifier somewhat unclear. I think I'm understanding now that 1) it isn't a lookup-table and 2) a model is being trained using the DeliAn dataset as "truth" values so that it can predict an aerosol type classification using only the ARM lidar observed LDR and LR as input. If this is the case, I think a bit more direct explanation of the process/connection here would be helpful up front, because it took me a while to understand this.

The reviewer's understanding is correct. In the revision, we added the clarifications at the beginning of the methodology section:

"The DeLiAn is used here as a labeled training dataset rather than as a lookup table. Each DeLiAn entry provides a pair of optical observables (LR and LDR) together with an expert-assigned aerosol-type label, which we treat as "truth". The kNN classifier learns the relationship between these optical observables and aerosol types from the DeLiAn data, and is then applied to predict an aerosol type for each ARM Raman-lidar observation using only its measured LR and LDR as input. In other words, DeLiAn supplies the training labels, while the ARM lidar supplies the new (unlabeled) observations to be classified."

o Line 241: "cruises"

Changed

o Line 245: May need to remove either "database" or "based"

Changed

o Line 294: This first sentence is a bit repetitive from what was already stated a couple of paragraphs before.

Deleted

• Results and Discussions

o Seeing marine classifications for Oklahoma observations will immediately catch a reader's attention in this section, and the explanatory work shown in Appendix D should be referenced sooner. Right now, I'm only seeing Appendix D referenced in Appendix C.

We now introduce the explanation and supporting analysis for the "marine" classification earlier in the main text and refer readers directly to Appendix A (formerly Appendix D).

o Line 389: I may be misinterpreting the figure, but I'm seeing a bigger difference in the RNCCN and G-1 observations in the two scatter points after 19.5 UTC.

The reviewer is correct that the largest RNCCN–G1 differences occur in the two points after ~19:30 UTC. This is consistent with our interpretation: *"large RNCCN–aircraft CCN differences occur in the final two points after ~19:30 UTC, following a period of strong vertical variability in aerosol type (19:00–19:30 UTC; Figure 4c). The ~20-min offset can be explained by the lack of exact collocation: the aircraft sampled within a 30-km radius rather than directly above the lidar,* *so the aircraft and lidar likely sampled different air parcels that advected past the site at slightly different times under heterogeneous conditions."*

o Line 443-447: By "one valid G-1 CCN measurement," does this mean a single 1 Hz observation or one 10-min average?
It refers to at least one valid 1 Hz observation within the 10-minute window. The text has been revised to make it clearer.

o Figure 6: It is not clear to me from the writing that these (I believe) are daily averages. I assume since 20160915 has two points (for two flight segments) that data for other days only comes from one flight segment? It would also be helpful to know how many valid windows are included in each daily average.
Each point represents the mean over valid 10 min windows within one flight segment, rather than a daily average. The two points on 15 September 2016 correspond to two separate flight segments. We revised the Figure 6 caption accordingly and added the number of valid windows in parentheses after each case label.

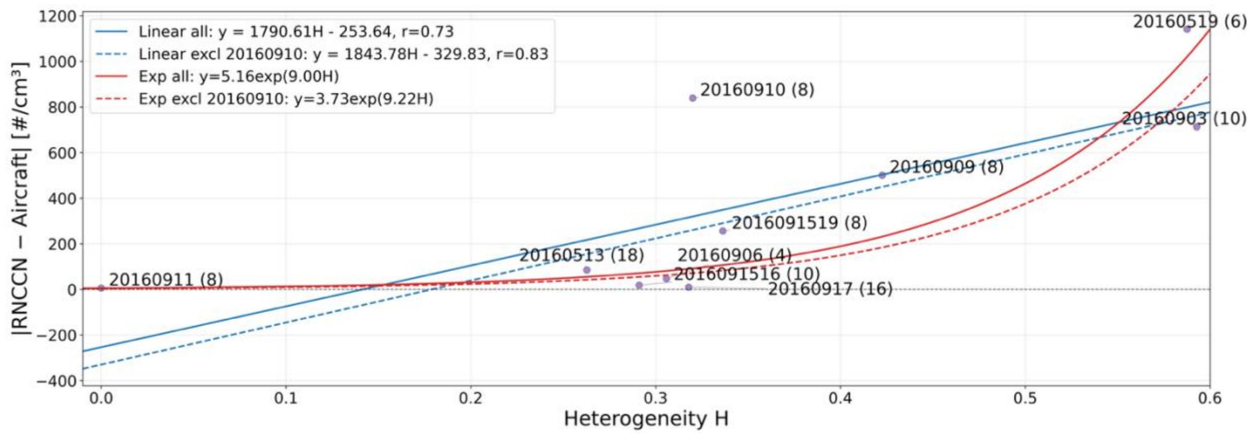

Figure 6. Case-level relationship between aerosol heterogeneity H and RNCCN Aircraft difference |RNCCN−Aircraft|. Cases are labeled by date (YYYYMMDD); when more than one qualifying flight segment occurred on the same day, the central UTC hour is appended (YYYYMMDDHH; e.g., 2016091516 and 2016091519). Solid lines represent linear and exponential fits using all cases, while dashed lines exclude the 20160910-outlier case.

o Line 507: Maybe change "structure" to "variation?"
Changed