Review of "**Constraining the Vertical Profiles of Aerosol Component and Heating Rate over East Asia through Assimilating CALIOP observations**" Cheng et al.

Synopsis:

This manuscript presents a component-resolved 4D-LETKF data assimilation system that assimilates CALIOP-MODIS synergistic speciated extinction retrievals (sea salt, dust, BC, water-soluble) into WRF-Chem/GOCART over East Asia, with the explicit goal of correcting the species-misallocation problem inherent to traditional total-AOT or bulk-extinction assimilation. The system is evaluated against the assimilated CALIOP profiles themselves, and independently against AERONET AOT/SSA and AD-Net lidar extinction.

The headline result is the reconstruction of a persistent free-tropospheric BC layer over the Indian subcontinent that the free-running model misses entirely, along with corresponding corrections to single scattering albedo and shortwave heating rate profiles. The problem motivation is sound, the observational dataset is valuable, and the independent AERONET/SSA validation is a genuine strength.

Several aspects of the assimilation design complicate the interpretation of the reported skill. Specifically, the background-correction bypass mechanism, the construction of the ensemble, and the combined water-soluble observation operator make it difficult to distinguish LETKF flow-dependent covariance skill from simpler bias-rescaling. These concerns must be resolved before the central claims can be fully assessed. I recommend major revisions.

Major Comments:

1. The background-correction scheme (Section 3.3) rescales the background ensemble by 1/R_corr when the localized weighted bias ratio falls below 0.2, bypassing the LETKF entirely. Because R_corr is constructed from the ratio of background to observations, this step drives the analysis toward the observations wherever the bypass triggers. This raises a critical concern: the near-perfect domain-mean correlations reported in Figure 4 (CORR = 1.00 for dust and water-soluble; CORR = 0.95 for BC) may partially reflect this forced agreement rather than genuine ensemble-based state-estimation skill.

The issue is most acute for black carbon, where the FR simulation completely misses the elevated 6 km peak (Figure 4c; CORR = 0.26) and background ensemble spread collapses to near-zero (lines 283-285). Under these conditions, the standard LETKF would yield negligible Kalman gain and effectively ignore the observations. The bypass therefore likely dominates the update for these missed plumes, and the resulting DA CORR of 0.95 is achieved primarily by rescaling rather than by flow-dependent covariance. For dust and water-soluble aerosols, where the FR simulation already captures the general vertical structure (CORR = 0.98 and 0.90 respectively), the marginal improvement to 1.00 is similarly difficult to attribute to LETKF performance without knowing how often the bypass intervened.

To resolve this ambiguity and allow readers to evaluate the true LETKF contribution, I suggest the paper report:

- The fraction of grid points, vertical levels, and assimilation cycles where the bypass triggered versus standard LETKF assimilation, disaggregated by species;
- The Figure 4 statistics computed separately for bypass-corrected and LETKF-assimilated subsets, so readers can evaluate whether the dramatic CORR/RMSE improvements reflect ensemble covariance skill or bias-rescaling;
- Ideally, a sensitivity experiment omitting the bypass (or with a more stringent threshold) to isolate the LETKF's contribution under conditions of near-zero background spread.

Without this decomposition, it is difficult to distinguish LETKF flow-dependent covariance skill from simpler bias-rescaling, and the paper's central claim that component-resolved ensemble assimilation improves vertical structure remains incompletely supported.

2. The 20-member ensemble is generated solely by perturbing aerosol emissions(Section 3.3). So the meteorology is running deterministically? This means the ensemble has negligible spread in vertical transport processes (convective lofting, PBL mixing, plume-rise injection heights). But in line 321-323, the authors attribute the persistent underestimation of the 6 km BC peak magnitude to "deficiencies in the model's vertical transport mechanisms or an underrepresentation of the biomass burning plume injection heights".

This creates an internal tension: if the background covariance contains no information about transport uncertainty, the system can only redistribute mass vertically via the prescribed emission profiles, not correct the altitude of lofted plumes through flow-dependent covariance. The fact that DA captures the shape but not the magnitude of the 6 km BC peak (Figure 4c) is consistent with this structural limitation. The authors should explicitly address this constraint.

3. The paper acknowledges that GOCART assumes externally mixed BC, whereas the CALIOP-MODIS retrieval assumes internal (core-shell) mixing (Section 4.1, lines 288-292). However, this discrepancy is not carried through to the radiative heating evaluation in Section 4.4. Since the heating rates are computed using the model's own (externally mixed) optical scheme applied to the DA-analysis BC mass, any bias in assumed mass absorption efficiency propagates directly into the reported 3.0 K/day peak heating anomaly. I would ask the authors to comment, even qualitatively, on the expected sign and approximate magnitude of this effect, since it bears directly on the paper's quantitative radiative. Specifically, the authors should shed some light regarding:

- Is the DA system inflating BC mass beyond its true atmospheric value to match observed extinction;
- Is the diagnosed SW heating rates (Figure 13) then applying this excess mass to the model's lower MAC, creating a non-obvious bias in the 3.0 K day^{-1} peak heating estimate.

The authors should comment on the sign and approximate magnitude of this effect. Is the reported peak heating an upper bound, a lower bound, or bracketed by the two mixing assumptions? Is the reported peak heating an upper bound, a lower bound, or bracketed by the two mixing assumptions? Given recent literature cited in the Discussion (Tiwari et al., 2025; Guan et al., 2026a), this could perturb regional forcing estimates by several W m-2.

4. The observation operator for water-soluble (WS) extinction combines sulfate and organic carbon (OC) into a single channel: β_WS = α_OC·M_OC + α_SU·M_SU (Section 3.1). While the paper criticizes traditional total-AOT assimilation for being under-constrained with respect to aerosol composition, this WS channel creates a similar problem at the species level: a single optical observation cannot independently constrain two chemically and radiatively distinct species. The LETKF partitions the increment using background ensemble cross-correlations between OC and sulfate, but these species have different sources, lifetimes, hygroscopic growth, and deposition behaviors. The variable localization (Section 3.2) permits WS observations to update both species, but the paper provides no validation of the resulting OC/sulfate partition. Given that both OC and sulfate are predominantly scattering-dominated aerosols, mispartitioning between them likely has much smaller radiative consequences than misallocating between BC and scattering species. However, from the perspective of aerosol composition and chemical transport, the OC/sulfate partition is still a relevant uncertainty. Additionally, if future applications of this system aim to use the analysis fields for process studies or source attribution, the inability to independently constrain OC and sulfate would be a limitation. The authors should acknowledge this under-constrained aspect of the WS observation operator explicitly as a limitation and clarify whether the LETKF's ensemble correlations between OC and sulfate are physically reasonable or whether the system tends to adjust one species preferentially.

5. The DA experiment reports near-perfect domain-mean correlations against the assimilated CALIOP extinction profiles in Figure 4 (CORR = 1.00 for dust and water-soluble; CORR = 0.95 for BC), representing dramatic improvements over the FR experiment (0.98, 0.90, and 0.26, respectively). However, these self-validation metrics stand in striking contrast to the independent AERONET validation in Figure 6, where the column-integrated AOT shows only modest improvement. CORR increases from 0.573 to 0.626, BIAS improves from -0.278 to -0.208, and the fraction of samples within ±0.10 of the observed AOT rises only from 23.36 % to 28.10 %.

While it is expected that self-validation against assimilated data would be stronger than independent validation, the magnitude of the discrepancy raises questions about how much of the self-validation skill reflects genuine state-estimation capability versus forced agreement with the assimilated observations. The near-perfect correlations (1.00 for two species) are unusually high for a 20-member ensemble system with sparse satellite observations and could indicate over-fitting or the influence of the bypass mechanism described in Major Comment 1.

This concern is amplified by the background-correction bypass mechanism (Section 3.3). When the background ensemble spread collapses to near-zero-which occurs precisely for the elevated BC plumes at ~6 km that dominate the FR error (Figure 4c, lines 283-285) (also mentioned in Major Comment 1), the system rescales the background by 1/R_corr and bypasses the LETKF entirely. The resulting CORR = 0.95 for BC may therefore be partially tautologically guaranteed wherever the bypass dominates, rather than earned through flow-dependent covariance. The AERONET validation, being independent and located at sites often far from CALIOP overpass tracks, is not subject to this forced agreement, which may explain why the skill improvement there is more modest.

The spatial distribution of AERONET improvements (Figure 7) further illustrates this point. While the DA experiment reduces absolute BIAS at 91.5% of sites and RMSE at 71.2%, the magnitudes are heterogeneous, and some sites show negligible or even negative improvement. Given that CALIOP tracks are sparse and the horizontal localization radius is 75 km, improvements at downwind sites necessarily come from dynamical propagation of the analysis state rather than direct observation influence. The patchy nature of the improvements is therefore not inherently problematic, it reflects the sparsity of the satellite observations and the complexity of transport pathways.

Nevertheless, the contrast between near-perfect CALIOP correlations and modest AERONET improvements requires explicit reconciliation. I therefore ask the authors to:

1. Address whether the CALIOP CORR values of 1.00 are physically credible or may reflect over-fitting;
2. Discuss the role of the bypass mechanism in generating these near-perfect correlations;

3. Acknowledge that the independent validation provides a more conservative estimate of the system's true performance.

Without this reconciliation, the manuscript presents two seemingly incompatible pictures of performance, and resolving which picture reflects the true capability of the component-resolved LETKF is essential before the central claims can be fully assessed.

6. If the background ensemble mean is effectively zero at a grid point (a completely missed aerosol event), R_corr approaches zero and the correction step requires division by a near-singular value. The paper mentions a 3-sigma outlier check (line 228), but does not describe an explicit numerical floor or regularization for R_corr. Given that missed BC plumes are the primary motivation for the correction scheme, the authors should clarify how this edge case is handled (e.g., is R_corr floored at 0.01? Is a minimum background spread enforced?).

7. The simulation uses HTAPv3 emissions for 2018, while the study period is March 2021- a three-year gap that includes the COVID-19 emission perturbation. The authors should acknowledge that some portion of the diagnosed "model structural deficiency" may reflect stale emission inputs rather than transport or mixing errors.

Minor Comments:

1. The temporal localization scale σ_t = 0.01 is given without units. Please clarify what this value represents physically.

2. Tsikerdekis et al. (2021), already cited, also performed speciated optical-property assimilation (POLDER/PARASOL size and absorption constraints). The introduction should more explicitly state what distinguishes the present approach (e.g., explicit per-species vertical mapping, the background-correction step) from that prior work.

3. In Figure 15, the latitude-height heating-rate cross-section is shown for a single timestamp (06:00 UTC, 7 March 2021), but the accompanying text generalizes the "warming aloft, reduced heating below" structure as a regional feature!!!! Either a composite over multiple events or a more explicit caveat that this is a single case study would be appropriate should be mentioned.