Response to the Comments of Referees

We would like to thank to the reviewers for giving constructive criticisms, which are very helpful in improving the quality of the manuscript. We have made minor revision based on the critical comments and suggestions of the referees. The referee's comments are reproduced (*black*) along with our replies (blue) and changes made to the text (red) in the revised manuscript. All the authors have read the revised manuscript and agreed with submission in its revised form.

Anonymous Referee #3

Comment NO.1: *The manuscript presents a highly relevant and ambitious effort to assimilate speciated aerosol vertical profiles from CALIOP-MODIS into the WRF-Chem model using a 4D-LETKF framework. Constraining the vertical distribution of distinct aerosol components is a well-known challenge in the aerosol modeling community, and the authors' approach to explicitly optimizing scattering and absorbing species is commendable. The paper is generally well-written and addresses an important topic within the scope of the journal. I recommend minor revisions before publication.*

Response: We sincerely appreciate the reviewer's positive evaluation and recognition of the scientific value and novelty of our work.

Comment NO.2: *While I understand the engineering necessity of the background correction scheme (Section 3.3) to handle near-zero background spread, replacing the background ensemble directly with the rescaled observations completely bypasses the Kalman update. From a dynamical perspective, abruptly inserting a massive amount of aerosol mass (e.g., the missed 6 km BC plume) into the model state can cause severe "model shock." Since WRF-Chem is running with aerosol-radiation feedback enabled, this sudden mass insertion could generate spurious vertical velocity or artificial buoyancy anomalies in the subsequent integration steps. I recommend that the authors discuss whether any initialization techniques or relaxation timescales were applied to smooth this mass insertion. If not, please evaluate and discuss the potential influences introduced by this hard-replacement strategy.*

Response: To address your concerns, we have conducted a comprehensive statistical analysis of the triggering patterns, performed a new sensitivity experiment (DA_NoBypass) to demonstrate the true capability of the LETKF, and explicitly discussed these limitations in a newly added Section 5.2.

(1) The primary motivation for introducing this scheme was to handle extreme cases, such as when the forward model completely misses an aerosol event. In such cases, the background ensemble spread collapses to near-zero. Under these conditions, the standard LETKF yields a near-zero Kalman gain and effectively ignores the observations entirely. The correction scheme was designed strictly to prevent the assimilation system from failing in these severe underestimation scenarios.

(2) We have calculated the exact triggering statistics. Spatially and temporally, the bypass mechanism was triggered in approximately 28.6% of the total assimilated grid points.

Disaggregated by species, it triggered most frequently for dust (43.6%) and black carbon (24.5%), where the model exhibits the most severe structural missing events. Vertically, the triggering predominantly occurred in the free troposphere (4-7 km) rather than in the boundary layer.

(3) To prove that our system is still a genuine data assimilation scheme, we conducted the DA_NoBypass experiment, which completely turns off this correction scheme and relies solely on the standard LETKF. As shown in Fig. 16, even without this "poor man's update", the pure LETKF experiment significantly improves the vertical structures compared to FR experiment. The correlation coefficient (CORR) for black carbon improved from 0.26 (FR) to 0.79 (DA_NoBypass), and dust improved from 0.98 to 0.99. This demonstrates that the LETKF possesses genuine flow-dependent ensemble covariance skill and effectively captures dynamic aerosol distributions, rather than just performing a numerical rescaling.

(4) We agree that ignoring observational uncertainty during the correction step is a significant simplification. While we implemented a 3-sigma outlier check and required a minimum number of local observations ($N_{obs} > 3$) to mitigate the impact of noisy retrievals, it remains some limitations.

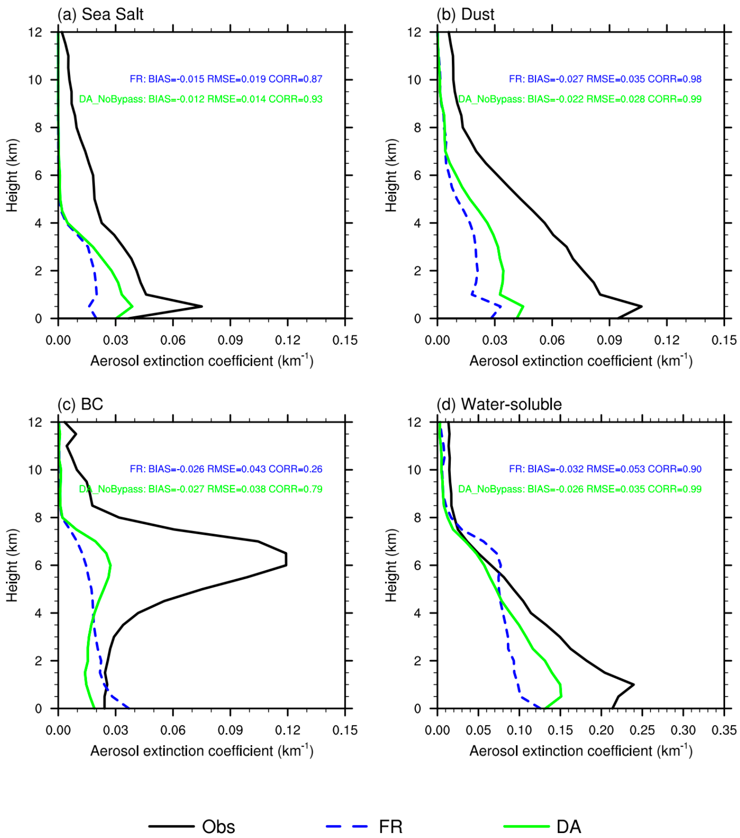

Figure 16. Same as Fig. 4 but for DA_NoBypass experiment.

Changes in Manuscript: Please refer the description in new Section 5.2 in the revised manuscript as follows:

“5.2 Discussion on the DA System Limitations and the Bypass Mechanism

While the component-resolved assimilation system demonstrates significant improvements, it is crucial to objectively distinguish the contributions of the background correction mechanism from the pure LETKF assimilation. It should be acknowledged that the background correction scheme acts as a simplified nudging approach that bypasses the rigorous uncertainty weighting of the standard Kalman filter. Consequently, this direct rescaling becomes less trustworthy when observational uncertainties are large. Although we implemented a 3-sigma outlier check and required a minimum number of local observations to mitigate the impact of noisy retrievals, ignoring observational uncertainty during the correction step remains a limitation.

To demonstrate the true capability of the LETKF, we have conducted a new assimilation experiment (DA_NoBypass) as pure assimilation. Spatially and temporally, the bypass mechanism was triggered in approximately 28.6% of the total assimilated grid points across all assimilation cycles. Disaggregated by species, it triggered in 43.6% of cases for dust, 17.8% for water-soluble aerosols, and 24.5% for BC. Vertically, it predominantly occurred in the free troposphere (e.g., 4-7 km) rather than the boundary layer. As shown in Fig. 16, the DA_NoBypass experiment significantly improves the CORR across different aerosol species compared to the FR experiment. For instance, the CORR for Black Carbon (BC) dramatically improved from 0.26 (FR) to 0.79 (DA_NoBypass). Similar structural improvements are also observed in sea-salt (CORR improved from 0.87 to 0.93) and water-soluble (from 0.90 to 0.99). Because simple bias-rescaling primarily corrects systematic mean errors without significantly altering the temporal or spatial variation patterns, these substantial improvements in CORR particularly for BC, strongly demonstrate that the LETKF system possesses genuine flow-dependent ensemble covariance skill. It effectively captures and corrects the dynamic distribution of aerosols, rather than just performing a numerical rescaling.

Therefore, the extreme local improvements in completely unsimulated plumes highlight a fundamental limitation of ensemble-based DA: empirical corrections remain indispensable when background spread collapses. Conversely, the independent AERONET and AD-Net validations presented in Section 4.2 provide a much more robust and conservative estimate of the system's true predictive skill. Because these sites are often far from CALIOP overpass tracks, the modest but solid improvements observed there reflect the genuine dynamical propagation of the assimilated fields rather than the forced local agreement of the bypass mechanism."

Comment NO.3: *Setting the observation error to 10% of the retrieved extinction is unrealistically low, especially for speciated retrievals where uncertainties often exceed 50%. In the LETKF framework, an artificially low observation error forces the Kalman gain to heavily favor the observations, which may partially explain the "near-perfect" self-validation scores (CORR = 1.0) alongside the bypass mechanism. The authors should justify this choice from a filter-tuning perspective. Was the 10% threshold chosen specifically to force the analysis closer to the observations due to persistent model negative biases? A sensitivity test using a more realistic observation error (e.g., 30% or 50%) over a short period would greatly strengthen the paper by* *demonstrating how the LETKF behaves when the observation and background errors are more balanced.*

Response: We fully agree with that the native, pixel-level retrieval uncertainties for speciated aerosol extinction (as reported in Kudo et al., 2023) can be substantial, often exceeding 100% for specific components like absorbing aerosols and dust. We apologize for the lack of clarity in our original manuscript regarding how the 10% observational uncertainty was defined and applied within our data assimilation (DA) system.

To clarify, our data assimilation system does not simply assume a flat 10% total uncertainty. As implemented in our system, the actual total observation error fed into the filter is much larger and dynamically calculated based on two key statistical and physical considerations:

(1) The uncertainties reported in Kudo et al. (2023) primarily represent the errors at the native pixel resolution. However, in our DA system, we do not assimilate raw pixel data. Instead, we generate "super-observations" by aggregating the high-resolution retrievals onto the 25 km WRF-Chem horizontal grid and vertical model layers. As demonstrated by previous DA studies, spatial and temporal aggregation significantly reduces the random component of retrieval errors (scaling approximately with $1/\sqrt{N}$, where N is the number of aggregated pixels). Consequently, the retrieval uncertainty of the aggregated super-observation is substantially lower than that of an individual pixel.

(2) As defined in Section 2.2.1, the total observation error (ϵ_o) is formulated as $\epsilon_o = \sqrt{\epsilon_{inst+retri}^2 + \epsilon_{rep}^2}$. The 10% value was solely used as an empirical baseline for the aggregated instrumental and retrieval error ($\epsilon_{inst+retri}$). Crucially, the representativeness error (ϵ_{rep}), which is dynamically calculated as the standard deviation of the sub-grid high-resolution retrievals within each 25 km grid cell, often dominates the total error budget. In regions with heterogeneous aerosol plumes, ϵ_{rep} is large, making the final total observation error (ϵ_o) much larger than 10%. This ensures that the total error realistically reflects the high uncertainty of the speciated observations in complex scenes. To explicitly demonstrate this, we have included a time series of the observation errors during the assimilation period. As shown in Fig. S1, the representativeness error (ϵ_{rep}) clearly dominates the total observation error (ϵ_o), while the baseline instrumental/retrieval error ($\epsilon_{inst+retri}$) remains a much smaller component.

In addition, in the ensemble-based DA frameworks like the LETKF, the specification of observation error is often treated as an empirical tuning parameter to ensure optimal filter performance. Setting the baseline error too high can cause the filter to ignore valuable observational constraints, while setting it too low leads to overfitting. The empirical 10% baseline, combined with the dynamically calculated ϵ_{rep}, provided a reasonable Kalman gain that successfully drew the model state toward the observations without causing filter divergence.

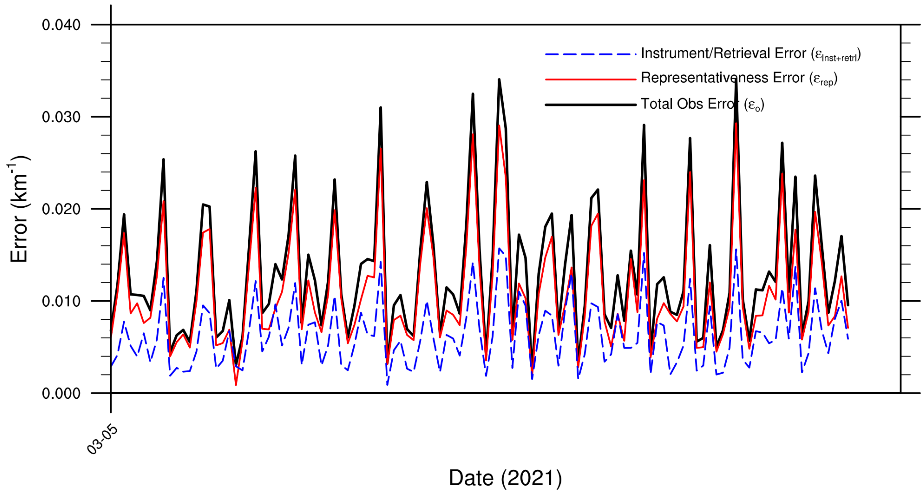

Figure S1. Temporal evolution of the observation errors, including the total observation error (ϵ_o), representativeness error (ϵ_{rep}), and instrumental/retrieval error ($\epsilon_{inst+retri}$), during the assimilation period.

Changes in Manuscript: Please refer the description in Section 2.2.1 (Line 153-167) in the revised manuscript as follows and Fig. S1 in the supplements:

"The observational uncertainty $\epsilon_o$is calculated as $\epsilon_o = \sqrt{\epsilon_{inst+retri}^2 + \epsilon_{rep}^2}$, where $\epsilon_{inst+retri}$ is the aggregated instrumental and retrieval error, and ϵ_{rep} is the representativeness error. It is important to note that while the native pixel-level retrieval uncertainties for speciated aerosol extinction can be substantial (sometimes exceeding 100% for specific components as reported by Kudo et al., 2023), our DA system assimilates aggregated "super-observations" mapped onto a 25 km grid. This spatial aggregation significantly mitigates the random components of the retrieval error. Because the retrieval algorithm does not explicitly provide pixel-by-pixel uncertainty flags that can be directly aggregated, a baseline of 10% of the observed aerosol extinction is employed as an empirical proxy for $\epsilon_{inst+retri}$. This empirical value is chosen to ensure optimal filter performance and avoid filter divergence. Crucially, the representativeness error (ϵ_{rep}), derived from the standard deviation of the sub-grid high-resolution retrievals within the 25 km WRF grid cell, is explicitly calculated. In regions with complex aerosol distributions, ϵ_{rep} typically dominates the error budget (as demonstrated in Fig. S1), ensuring that the total observation error (ϵ_o) realistically reflects the high uncertainties inherent in the speciated observations."

Comment NO.4: *Line 118: The text mentions "aerosol-radiation feedback enabled." Does the FR (Free Run) also include this feedback? If so, the meteorological fields between FR and DA will diverge over the 25-day period due to different aerosol heating. Please clarify if the meteorology was re-initialized or nudged to ERA5 during the cycle to prevent chaotic divergence.*

Response: Yes, the aerosol-radiation feedback is enabled in both the FR and DA experiments. We completely agree that a continuous 25-day free integration with interactive aerosols would lead to significant chaotic divergence in the meteorological fields, obscuring the true impact of aerosol assimilation. To prevent this, our system employs a daily re-initialization strategy rather than continuous grid nudging. In our configuration, the meteorological fields are cold-started and re-initialized every 24 hours using ERA5 reanalysis datasets, while the aerosol chemical states are continuously cycled from the previous day's analysis. We have clarified this meteorological cycling strategy in the revised manuscript.

Changes in Manuscript: We have added the description in the revised manuscript as below:

"To prevent chaotic divergence of the meteorological fields over the simulation period while maintaining the aerosol-radiation feedback, a daily re-initialization strategy is employed. Specifically, the meteorological fields are re-initialized every 24 hours using ERA5 reanalysis datasets, whereas the aerosol chemical states are continuously cycled and carried forward from the previous cycle." Please refer the description in the revised manuscript, Line 124-127.

Comment NO.5: *Section 3.2: The vertical localization length is set to 1 model layer. Given that WRF-Chem vertical layers stretch significantly with altitude (from tens of meters in the PBL to over a kilometer in the upper troposphere), a fixed model-layer localization might be too restrictive in the free troposphere and too broad in the boundary layer. Please comment on the impact of this choice.*

Response: Agree. In our configuration, the domain is set up with 34 vertical layers up to 100 hPa. Using a fixed model-layer localization is an intentional design that takes advantage of the terrain-following sigma coordinate's stretching nature. In the planetary boundary layer, where aerosol vertical gradients are sharp, 1 model layer corresponds to a very narrow physical distance, which effectively prevents spurious assimilation updates across strong temperature inversions. Conversely, in the free troposphere, where lofted aerosol plumes are typically well-mixed over broader vertical extents, 1 model layer corresponds to a larger physical distance, allowing the observations to appropriately update the broader plume structure. Therefore, this model-layer-based localization inherently acts as an altitude-adaptive physical localization scale. We have added a brief discussion in Section 3.2 to clarify this physical rationale.

Comment NO.6: *Section 3.1: The authors mention that the GOCART scheme simulates dust in five size bins and sea salt in four size bins. However, the CALIOP-MODIS dataset provides a single bulk extinction coefficient for dust and sea salt, respectively. When the LETKF updates the state vector, how are the analysis increments distributed among these multiple size bins? Are the individual bins updated independently based on their respective background error covariances, or is the total mass increment simply distributed proportionally to the background mass fractions of each bin? Please clarify this technical detail.*

Response: One of the primary advantages of the LETKF system is that it does not rely on simple proportional scaling. Instead, the analysis increments for the individual size bins are updated independently based on their flow-dependent background error covariances. During the assimilation step, the ensemble forward operator calculates the bulk extinction from all bins. The LETKF then computes the cross-covariance between this simulated bulk extinction and the mass concentration of each specific size bin across the ensemble members. Consequently, bins with larger ensemble spread and stronger correlation with the total optical property receive proportionally larger updates. This allows the system to dynamically adjust the aerosol size distribution based on the physical uncertainties captured by the ensemble. We have explicitly clarified this mechanism in Section 3.1.

Changes in Manuscript: We have added the description in Section 3.1 as below:

"Crucially, for multi-bin species like dust and sea salt in the GOCART scheme, the analysis increments are not simply distributed proportionally to their background mass fractions. Instead, the LETKF independently updates the mass of each size bin based on the flow-dependent background error cross-covariances between the simulated bulk extinction and the individual bin concentrations. This ensemble-driven mechanism allows the assimilation system to dynamically adjust the aerosol size distribution based on the physical uncertainties captured by the ensemble spread." Please refer the description in the revised manuscript, Line 219-223.

Comment NO.7: *Section 2.1: The manuscript highlights the successful reconstruction of elevated BC plumes originating from the Indian subcontinent (e.g., Figure 4c). Given that the Indian subcontinent is often located near the southwestern boundary of typical East Asian modeling domains, the treatment of chemical boundary conditions is critical. Please specify what global dataset (e.g., MOZART, WACCM, or CAM-chem) was used to provide time-varying CBCs. If idealized or climatological boundary conditions were used, could the severe underestimation of the mid-tropospheric BC in the FR experiment be partially attributed to missing aerosol inflow at the domain boundaries?*

Response: Agree. we utilized climatological profile boundary conditions rather than time-varying chemical boundary conditions from a global model. The severe underestimation of mid-tropospheric BC in the FR experiment is largely attributed to this missing dynamic aerosol inflow at the domain boundaries, especially given the proximity of the Indian subcontinent to our southwestern boundary. A key finding of our study is that data assimilation successfully compensates for this missing boundary inflow. By assimilating aerosol observations, DA experiment effectively reconstructs the elevated BC plumes that were otherwise absent in the FR experiment due to the idealized boundary conditions.

Changes in Manuscript: We have added the description in Section 3.1 as below:

"we utilized climatological profile boundary conditions rather than time-varying chemical boundary conditions from a global model." Please refer the description in the revised manuscript, Line 217-218.