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 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 #1

Comment NO.1: *This paper describes a regional data assimilation experiment for aerosol using a novel height- and species resolved observational dataset. The authors evaluate the analysis against both assimilated and independent observations and conclude significant improvements can be found in the analysis vs the free run. The analysis shows sizeable redistributions of aerosol and an increase of total black carbon which, as the authors show, strongly affects heating rates in atmosphere. This points to important limitations in our forward modelling ability.*

The paper is very much in scope although one of its conclusions (that modelled and observed vertical profiles are very different with a strong impact on heating rates) will not surprise many people. However, the method to correct this is new and the paper describes the issue for a region very important for anthropogenic aerosol. The paper is also very readable.

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

Comment NO.2: *The paper lacks details on the LETKF setup. While observation operators and background correction are described the basic system is not. This is important as it is not even clear what sort of perturbation the LETKF uses. See further my questions below.*

Response: Done.

Thank you for pointing out the lack of details regarding the LETKF setup. Following your constructive suggestions, we have added a new subsection (Section 3.1 LETKF system) at the beginning of Section 3 to explicitly describe the basic system used in the LETKF system:

"To generate the background ensemble for the LETKF system, we apply spatiotemporal correlated perturbations to the aerosol emissions rather than perturbing the aerosol concentrations. A total of 20 ensemble members are generated using perturbation factors drawn from lognormal distributions. Based on known emission uncertainties, the means of the perturbation factors are all equal to one, and the standard deviations are set to 350% for carbonaceous sources (OC and BC) and 130% for the sulfate precursor SO_2 (Yumimoto and Takemura, 2011). Natural aerosol emissions are also randomly perturbed, assuming uncertainties of 100% for dust and 500% for sea salt, consistent with our previous study (Cheng et al., 2019). Furthermore, since OC, BC, and SO_2 often share common combustion sources, their emission perturbations are assumed to be positively correlated with each other, whereas they remain uncorrelated with the natural aerosol species (Schutgens et al., 2010; Pagowski et al., 2014). Regarding the lateral boundary conditions, the meteorological boundaries are driven by the ERA5 reanalysis. For the chemical boundary conditions, we utilize default climatological profiles rather than time-varying outputs from a global chemical transport model. In the current setup, these chemical boundary conditions are not perturbed. "

References:

Cheng, Y., Dai, T., Goto, D., Schutgens, N. A. J., Shi, G., and Nakajima, T.: Investigating the assimilation of CALIPSO global aerosol vertical observations using a four-dimensional ensemble Kalman filter, Atmos. Chem. Phys., 19, 13445–13467, https://doi.org/10.5194/acp-19-13445-2019, 2019.

Pagowski, M., Liu, Z., Grell, G. A., Hu, M., Lin, H.-C., and Schwartz, C. S.: Implementation of aerosol assimilation in Gridpoint Statistical Interpolation (v. 3.2) and WRF-Chem (v. 3.4.1), Geosci. Model Dev., 7, 1621–1627, https://doi.org/10.5194/gmd-7-1621-2014, 2014.

Schutgens, N. A. J., Miyoshi, T., Takemura, T., and Nakajima, T.: Sensitivity tests for an ensemble Kalman filter for aerosol assimilation, Atmos. Chem. Phys., 10, 6583–6600, https://doi.org/10.5194/acp-10-6583-2010, 2010.

Yumimoto, K., and Takemura, T.: Direct radiative effect of aerosols estimated using ensemble-based data assimilation in a global aerosol climate model, Geophys. Res. Lett., 38, L21802, https://doi.org/10.1029/2011GL049258, 2011.

Changes in Manuscript: Please refer the description in new Section 3.1 in the revised manuscript.

Comment NO.3: *The papers suggests an observational uncertainty of 10% (Sect 2.2.1), yet the Kudo et al. paper finds uncertainties of 100% in the original dataset (Table 3), see also my questions below. The difference is a large one with important consequences. Maybe I misunderstand something but the authors need to present a clear narrative (Kudo is co-author on the current paper). I have trouble accepting a 10% uncertainty because already our best satellite AOD has 10% uncertainty (in long-term averages!) but the Kudo dataset is about vertically resolved, speciated extinction (daily!).*

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.

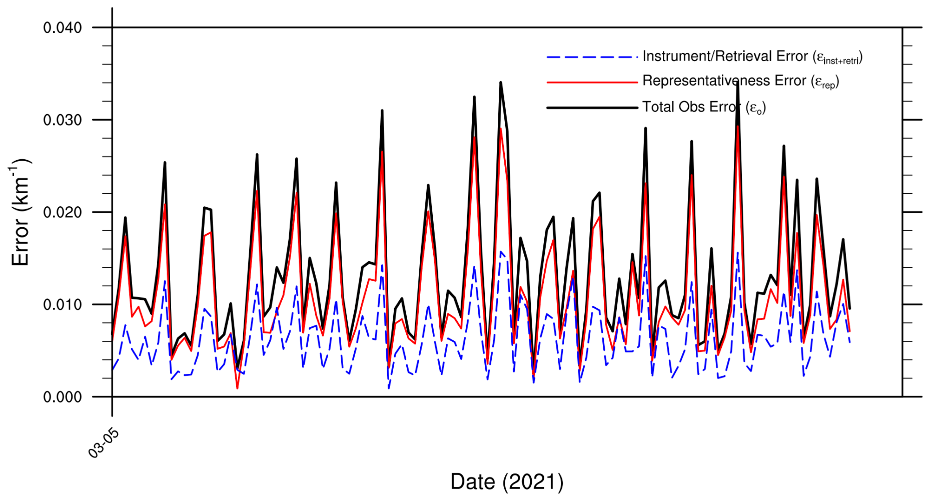

Figure S1. Temporal evolution of the observation errors, including the total observation error (ϵ_o), representativeness error (ϵ_{rep}), and instrumental/retrieval error ($\epsilon_{inst+retri}$), during the assimilation period.

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 ϵ_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: *The paper describes a background correction scheme but provides little information as to how often and when/where this scheme is employed by the LETKF. The scheme is essentially a poor man's analysis update, equating the analysis to the observations without any attention to uncertainties. The presence of this scheme suggests that good agreement between the analysis and assimilated observations will never be in doubt. But is the combined system still a data assimilation scheme or only a form of blunt nudging? Is this background correction acceptable given observational uncertainties and ensemble spread? The authors need to provide more detail and discuss the simplification of data assimilation by their background correction more clearly than they do now. I also want to point out that items 2) and 3) are related: the larger the observational uncertainty, the less truth worthy the background correction.*

Response: Agree. The background correction scheme is indeed a form of "blunt nudging" or a "poor man's analysis update" that effectively forces the analysis toward the observations, bypassing the rigorous uncertainty weighting of the standard Kalman filter. We also fully agree with your point that this direct rescaling becomes less trustworthy when observational uncertainties are large.

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.

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 ensure the aerosol event is totally missed in the model forecast.

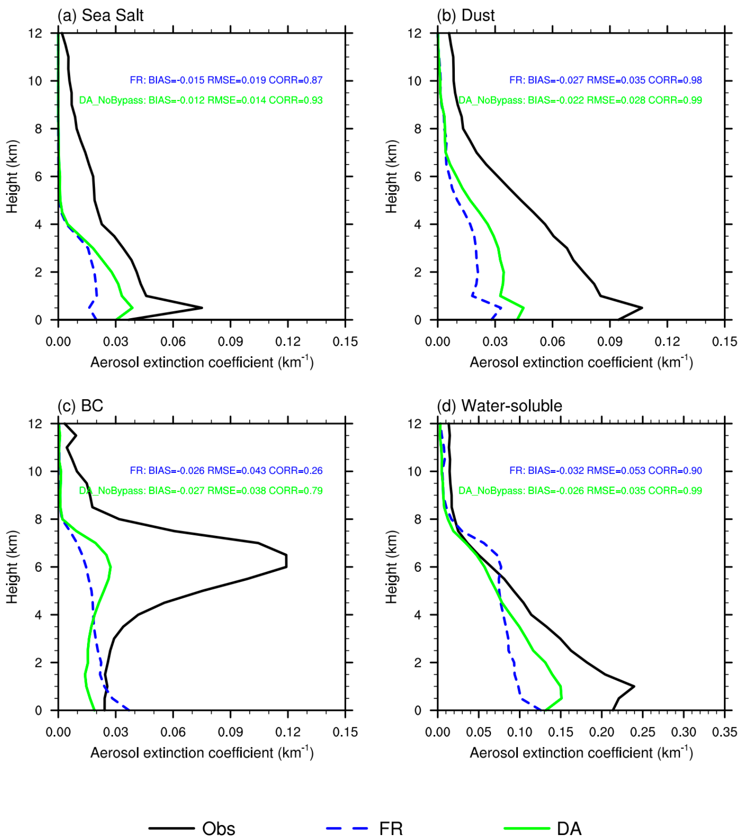

Figure 16. Same as Fig. 4 but for DA_NoBypass experiment.

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.5: *p. 1, l. 31: induces -> induce*

Response: Done. We have replaced the "induces" with "induce".

Changes in Manuscript: Please refer the modification in the revised manuscript, Line 31.

Comment NO.6: *p. 2, l. 41: "Elevated absorbing layers...lower atmosphere" this sentence is not clear to me, it seems that the two parts say almost the same, so why do you use 'whereas"?*

Response: Agree. We sincerely apologize for the confusion. We have rewritten the sentence for clarity as below:

"Over dark surfaces, absorbing aerosols located in the upper atmosphere absorb more incoming solar shortwave radiation than those in the lower atmosphere, leading to stronger warming at higher altitudes. In contrast, over bright surfaces, absorbing aerosols primarily interact with the reflected upward radiation, producing a more pronounced warming effect in the lower atmosphere."

Changes in Manuscript: Please refer the modification in the revised manuscript, Line 41-44.

Comment NO.7: *p. 2, l. 46: I'm a bit surprised by the mention of water vapor absorption. Surely this happens at ()IR wavelengths but your abstarct talks about SW.*

Response: Agree. We sincerely apologize for the confusion and we have deleted this sentence.

Changes in Manuscript: Please refer the modification in the revised manuscript, Line 46.

Comment NO.8: *p. 2, l. 60: I would argue that CALIOP and CATS also suffer from limited spatial coverage. CALIOP has a beam width of 70m? Maybe the authors can be a bit more precise. The spatial coverage here is horizontal spatial coverage.*

Response: Agree. We thank the reviewer for pointing this out. We have rewritten the sentence for clarity as below:

"However, the sparse distribution of these stations restricts their horizontal spatial coverage. Satellite-borne lidars, such as Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) and the Cloud-Aerosol Transport System (CATS), face a similar spatial limitation despite providing global high-resolution vertical profiles (Campbell et al., 2010; Proestakis et al., 2019). Due to their nadir-viewing geometry and narrow footprints, these instruments leave unobserved gaps between adjacent ground tracks. Coupled with their temporal discontinuity, these spatial gaps significantly restrict their spatiotemporal representativeness."

Changes in Manuscript: Please refer the description in the revised manuscript, Line 56-61.

Comment NO.9: *p. 3, l. 86-91: The statement (that the problem is particularly severe over East Asia) is not supported by analysis or references. Please provide some, or be more cautious in your statement.*

Response: Agree. We have modified the statement for clarity as below:

"Addressing this under-constrained problem is especially important in regions with complex aerosol mixtures, such as East Asia."

Changes in Manuscript: Please refer the modification in the revised manuscript, Line 86-87.

Comment NO.10: *p. 4, l. 120-121: "Aerosols are parametrised as external mixtures ..." Are the sizes and standard deviations defined in WRF-Chem identical to those in CRTM? Especially older models tend to be inconsistent about such choices.*

Response: Agree. To ensure microphysical consistency within the forward operator, the mapping of aerosol mass concentrations from WRF-Chem to CRTM strictly aligns with the GOCART scheme definitions. The effective radii and standard deviations assumed in the CRTM aerosol lookup tables are consistent with those defining the aerosol size bins in the WRF-Chem GOCART module, thereby minimizing internal discrepancies during the mass-to-extinction conversion (Lu et al., 2022; Wang et al., 2022).

References:

Lu, C.-H., Liu, Q., Wei, S.-W., Johnson, B. T., Dang, C., Stegmann, P. G., Grogan, D., Ge, G., Hu, M., and Lueken, M.: The Aerosol Module in the Community Radiative Transfer Model (v2.2 and v2.3): accounting for aerosol transmittance effects on the radiance observation operator, Geosci. Model Dev., 15, 1317–1329, https://doi.org/10.5194/gmd-15-1317-2022, 2022.

Wang, S., Dai, T., Li, C., Cheng, Y., Huang, G., and Shi, G.: Improving Clear-Sky Solar Power Prediction over China by Assimilating Himawari-8 Aerosol Optical Depth with WRF-Chem-Solar, Remote Sens., 14, 4990, https://doi.org/10.3390/rs14194990, 2022.

Changes in Manuscript: Please refer the description in the revised manuscript, Line 132-136.

Comment NO.11: *p. 5, l. 39: Can the authors provide a bit more detail about observational uncertainty? They assume 10% retrieval uncertainty but that can't be right. Especially because they break down extinction into different components. Table 3 in Kudo et al. (2019) suggests uncertainties of more than 100% in components like absorbing aerosol, dust aerosol. Assumed uncertainties in components will be important in the data assimilation step so I would like to ask for more detail from the authors.*

Response: Please refer to our detailed response to Comment NO.3.

Comment NO.12: *p. 5, l. 142: it is very likley that the authors underestimate representation uncertainty with this method as observations tend to cluster and aerosol fields tend to be strongly correlated in space. I understand that better approaches are hard to come by but please discuss this. See also Rijsdijk et al. GMD 2025 for a similar problem with NO_2.*

Response: Agree. We thank the reviewer for this insightful comment and for directing us to the highly relevant work of Rijsdijk et al. (2025). We fully agree with your point. We have added the discussion as below:

"It should be noted that this approach likely underestimates the true representativeness uncertainty. As highlighted by Rijsdijk et al. (2025), satellite observations frequently cluster spatially rather than being randomly distributed. Because aerosol fields exhibit strong spatial correlations, a clustered sample captures less sub-grid variability and is less representative of the entire grid cell. While we maintain the present formulation for computational efficiency, accounting for this systematic sampling effect remains an important area for future refinement."

Reference:

Rijsdijk, P., Eskes, H., Dingemans, A., Boersma, K. F., Sekiya, T., Miyazaki, K., and Houweling, S.: Quantifying uncertainties in satellite NO_2 superobservations for data assimilation and model evaluation, Geosci. Model Dev., 18, 483–509, https://doi.org/10.5194/gmd-18-483-2025, 2025.

Changes in Manuscript: Please refer the description in the revised manuscript, Line 163-167.

Comment NO.13: *p. 5. Sect. 2.2.1.: It seems to me the authors do not discuss how the WRF-Chem aerosol is mapped onto the CALIOP-MODIS components. Can they provide more detail? In particular, how well do WRF-CHEM assumed particle sizes agree with retrieved sizes per component? Same for refractive index per species. Both WRF-Chem and Kudo et al. make* *assumptions on refractive index. Are they the same? I later saw that you partially answer this in Sect. 3.1. Still, my question on sizes and refractive indices still stands.*

Response: Done. We have discussed in Section 3.2 Component-specific observation operator: "However, a fundamental limitation in our observation operator must be acknowledged regarding the microphysical inconsistencies between the forward model and the assimilated retrievals. For the mass-to-extinction conversion in the forward model, the particle sizes and complex refractive indices are strictly defined by the CRTM GOCART-GEOS5 configuration, which treats aerosols as externally mixed pure spherical particles. In contrast, the CALIOP-MODIS retrieval algorithm (Kudo et al., 2023) assumes an internal mixing state for specific components. Most notably, BC is treated using a core-grey shell model, where BC acts as the core internally mixed with water-soluble components as the shell. Consequently, the retrieval employs different assumed size distributions and calculates effective complex refractive indices based on this internal mixing assumption. This fundamental discrepancy in mixing states, sizes, and refractive indices introduces an inherent representativeness error when mapping the simulated externally mixed aerosols to the retrieved internally mixed aerosol extinctions. Nevertheless, it is worth emphasizing that different mixing states primarily affect the aerosol absorption coefficient, whereas their impact on the extinction coefficient is relatively minor. Because our study assimilates the aerosol extinction coefficient rather than the absorption coefficient, the overall impact of this microphysical inconsistency is significantly mitigated. Therefore, despite this theoretical limitation, our assimilation system can still effectively constrain the bulk aerosol mass and vertical structure."

Changes in Manuscript: Please refer the description in the revised manuscript, Line 253-266.

Comment NO.14: *p. 7, sect 3: Can the authors provide more detail on the LETKF system, especially the timing of assimilation? Does it happen every hour, or only at CALIOP overpass times? What is the time-window (4D-LETKF suggests there is a time-window in which obersvations at t will influence state at times before t). Also, how is the ensemble created? Are emissions perturb, or concentrations? What size of perturbations? Any spatial/temporal correlations in the perturbations? Also, how do the authors deal with boundary conditions (inflow from source regions outside domain).*

Response: Done. Please refer to our detailed response to Comment NO.3.

Comment NO.15: *p. 9, l. 206: "local patch center" has not been defined. Please rephrase. Only people familiar with LETKF will understand what it means.*

Response: Agree. We replaced "where r_h, r_v, and Δt represent the horizontal, vertical, and temporal distances of the observations from the local patch center, respectively" with "where r_h, r_v, and Δt represent the horizontal, vertical, and temporal distances between the observations and the specific model grid point being analyzed, respectively."

Changes in Manuscript: Please refer the modification in the revised manuscript, Line 271-272.

Comment NO.16: *p. 9, l. 215: "This strict variable localization... after data assimilation". If the observations of components were independent and the mapping from WRF-Chem to observations* *perfect, wouldn't the authors expect this to be unnecessary? Of course, nothing is perfect but I would appreciate it if the authors explain a bit more why this variable localization is necessary.*

Response: We thank the reviewer for this insightful comment. The reviewer is theoretically correct: if the ensemble size were infinite, the background error cross-covariance between physically independent species would naturally be zero, making variable localization unnecessary. However, in ensemble data assimilation, the finite ensemble size inevitably introduces sampling errors. This leads to spurious cross-correlations in the background error covariance matrix between independent chemical components. If not localized, these spurious correlations would cause unphysical cross-updates (e.g., dust observations erroneously updating black carbon). Therefore, variable localization is strictly necessary to filter out these sampling errors and enforce the theoretical zero-covariance. We have revised the description to explicitly state the necessity of variable localization from the perspective of sampling errors as below:

"Although physically independent species should theoretically exhibit zero error covariance, a finite ensemble size inevitably introduces sampling errors that lead to spurious cross-correlations. This strict variable localization eliminates these spurious correlations, thereby preserving the physical and chemical consistency of the analysis field."

Changes in Manuscript: Please refer the description in the revised manuscript, Line 283-286.

Comment NO.17: *p. 9, l. 219: "severe model underestimations" does this happen only for component AOD or also total AOD? What causes it? The authors description is a bit vague. Does default WRF-Chem already underestimate AOD? How large is the spread in the emissions and component AOD compared to typical errors? A common solution to this is a pre-tuning step whereby emissions are scaled according to a mismatch between deafult forward run and observations.*

Response: We thank the reviewer for these insightful questions. Below are our detailed responses to each point:

(1) Causes of underestimation:

The underestimation in the default WRF-Chem forward run primarily occurs in specific aerosol components, notably dust and black carbon (BC) extinction coefficients, which consequently leads to an underestimation of the total AOD in regions or periods dominated by these species. One major causes of these underestimations is the underestimation of the emission for dust and BC. Another critical factor is the inaccuracy in meteorological parameterizations, particularly vertical convection and boundary layer mixing, which significantly affect the vertical transport, distribution, and lifetime of aerosols.

(2) The spread in the emissions and component AOD:

Regarding the ensemble spread, we acknowledge that severe model biases can sometimes lead to a mismatch where the ensemble spread is smaller than the typical observation-minus-background errors. To prevent the data assimilation from ignoring the observations under such underestimations, we have introduced reasonable emission perturbations and applied covariance inflation in our assimilation system. This ensures that the ensemble spread in component remains sufficiently large and comparable to the typical errors.

(3) Why pre-tuning was not applied:

We agree that scaling emissions is a common solution for systematic biases. However, in our case, the underestimations are not systematic across all regions, times, and species. Because the errors originate from a complex combination of emission uncertainties and meteorological parameterizations, a simple, uniform scaling of emissions would be physically unreasonable. It could artificially inflate aerosol concentrations in regions where the default model already performs well, leading to spurious spatial distributions. Moreover, even when targeted tuning is applied to specific sources, it cannot resolve all discrepancies. For instance, although we have already tuned the dust emission sources in our model, there are still specific dust events that the model completely fails to capture. Therefore, we rely on the data assimilation system to dynamically correct these flow-dependent errors rather than applying a static pre-tuning.

Changes in Manuscript: We have clarified the description in the revised manuscript: "It should be noted that these underestimations are not systematic but primarily occur in specific aerosol components (e.g., dust and black carbon), driven by uncertainties in both emissions and meteorological parameterizations such as vertical convection. Because these biases are spatiotemporally heterogeneous, a simple global pre-tuning of emissions is inappropriate." Please refer the description in the revised manuscript, Line 290-293.

Comment NO.18: *p. 10, l 230: "outliers" so what do you do with outliers? I suppose they are not assimilated but I can't find this in the text. Please clarify.*

Response: Done. We have added the description about the outliers as below:

"These identified outliers are considered invalid and are completely excluded from both the background correction and the subsequent LETKF assimilation steps."

Changes in Manuscript: Please refer the description in the revised manuscript, Line 303-304.

Comment NO.19: *p. 10, l. 231: "background ensemble factors are rescaled by dividing by this factor." So each member is rescaled with the samefactor. Please clarify this. Also, this implies that you correct the ensemble mean but the ensemblespread remains rather small. I guess that may be why inflation is later on needed? (In my experience, with perturbed emissions, inflation is seldom needed).*

Response: Agree. It should be noted that each ensemble member is rescaled by the identical factor. While this conditional bypass mechanism effectively corrects severe negative biases and prevents negative concentrations, it introduces a mathematical limitation regarding the ensemble spread. Because all members are scaled uniformly, the relative ensemble spread remains artificially narrow, failing to fully represent the physical uncertainty of the newly scaled state. Consequently, to prevent filter divergence in subsequent assimilation cycles, the multiplicative inflation is necessary to provide sufficient background error covariance.

Comment NO.20: *p. 10, l. 236: "while preserving the physical consistency of the ensemble". I think this background correction stratgey is a poor man's data assimilation tehnique. It ignores uncertainty in both ensemble and observations and just puts the analysis equal to the observations. It seems unsatisfying especially because the authors remain quite vague about how often this happens in the first place. So I would appreciate more discussion and explanation (see also previous questions). How often is this background correction triggered compared to a real Kalman update? Are there any patterns in this triggering (e.g. more over land, more in free troposphere than in BL etc)?*

Response: Done. To address your specific questions regarding how often and where this occurs, we have conducted a new assimilation experiment (DA_NoBypass) as pure assimilation and calculated the exact triggering statistics and patterns:

(1) Overall frequency: Spatially and temporally, the bypass mechanism was triggered in approximately 28.6% of the total assimilated grid points across all assimilation cycles. The remaining 71.4% of the updates were performed by the standard LETKF Kalman update.

(2) Species patterns: 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. It triggered less frequently for water-soluble aerosols (17.8%).

(3) Vertical patterns: Vertically, the triggering predominantly occurred in the free troposphere (e.g., 4-7 km) rather than in the boundary layer.

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

Comment NO.21: *p. 10, l. 236: "physical consistency" please explain, the meaning is unclear to me. What aspect of physics do you refer to? Clearly no law of conservation is involved so I don't understand "physical consistency"*

Response: We agree that "physical consistency" is inaccurate here, as our rescaling method is a mathematical operation and does not explicitly enforce physical conservation laws. Our original intention is to emphasize that by rescaling the background ensemble members rather than replacing them with arbitrary values, we preserve the spatial patterns, multivariate correlations, and the relative ensemble spread generated by the model dynamics. To avoid any confusion, we have replaced the "physical consistency" with "spatial and multivariate consistency".

Changes in Manuscript: Please refer the modification in the revised manuscript, Line 315.

Comment NO.22: *p. 11, l 291: "core_shell" this should have been discussed earlier in Sect 2 on data and methods. See also my previous comments on mapping of aerosol species from obs to GOCART.*

Response: Done. We have discussed in Section 3.2 Component-specific observation operator.

Changes in Manuscript: Please refer the description in the revised manuscript, Line 253-266.

Comment NO.23: *p. 11, fig. 3 I would suggest using logarithmic axes to bring out the region from AOD=0.01-0.1 better*

Response: Accept. We have modified Fig.3 using logarithmic axes as below:

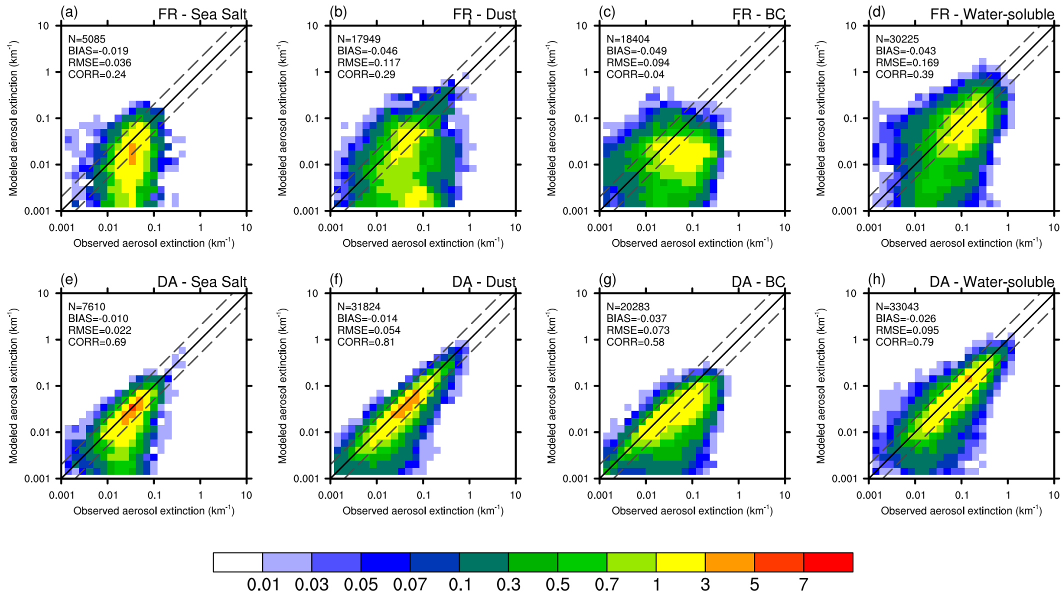

Figure 3. Density scatter plots comparing the modeled extinction coefficients (km^{-1}) against observations for (a, e) sea salt, (b, f) dust, (c, g) black carbon, and (d, h) water-soluble aerosols. The top row (a-d) represents the FR experiment, while the bottom row (e-h) shows the results from the DA experiment. The color scale indicates the data density. The solid black line represents the 1:1 line while the dashed lines represent the 2:1 and 1:2 lines. Statistical metrics (N: sample size, BIAS: bias, RMSE: root mean square error, CORR: correlation coefficient) are provided in each panel. The statistics are calculated using all valid data points within East Asia (shown in Fig. 1) from the surface up to 12 km altitude.

Changes in Manuscript: Please refer the description in the revised manuscript, Fig. 3.

Comment NO.24: *p. 13, l. 317: "substantial improvement" please remind the reader that this is a self-validation; the authors are only testing with observations that were used in the assimilation.*

Response: Agree. We have added the sentence to remind the reader that this is a self-validation: "However, it should be noted that this represents a self-validation, as the evaluation is performed against the same observations that were assimilated."

Changes in Manuscript: Please refer the description in the revised manuscript, Line 397-398.

Comment NO.25: *p. 13, l 324: "system is difficult" -> "system has difficulty". Your background correction strategy is designed for this sort of problems, why does your system still have problems correcting the low bias?*

Response: Done. We have deleted this sentence. The system still shows a slight low bias because the background correction mechanism has strict triggering thresholds (e.g., $R_{corr} < 0.2$). It is not triggered everywhere. In regions where standard LETKF is performed, the analysis is a weighted average of the background and observation, meaning the low bias in the background will still partially pull the analysis down.

Comment NO.26: *p. 13, l. 325: "magnitude of lofted plumes" it would be interesting if the authors can supply some analyses of plume structure in obs and FR/DA. Is it possible to show curtain plots (latitude-altitude, averaged over longitude for E Asia)i for BC or something similar? Fig 3 and 4* *provide average statistics so something that tells us more about how good DA reconstructs observed plumes.*

Response: Done. As suggested, we have plotted the latitude-altitude cross-sections of the black carbon (BC) extinction coefficient averaged over the East Asia longitude for CALIOP observations, FR, and DA experiments which is shown in the figure below.

As illustrated in the Fig. S2, the observations exhibit a distinct and strong lofted BC plume between 15°N and 30°N, extending vertically up to approximately 4 km. The FR experiment significantly underestimates the magnitude of this lofted plume, with most of the higher BC extinction values confined near the surface. In contrast, the DA experiment successfully reconstructs the observed lofted plume structure. It captures both the enhanced magnitude and the vertical extent of the BC plume much better than the FR. This cross-sectional analysis clearly demonstrates that the DA not only improves average statistics but also effectively corrects the vertical distribution of aerosols, successfully reproducing the lofted plumes observed in the real atmosphere.

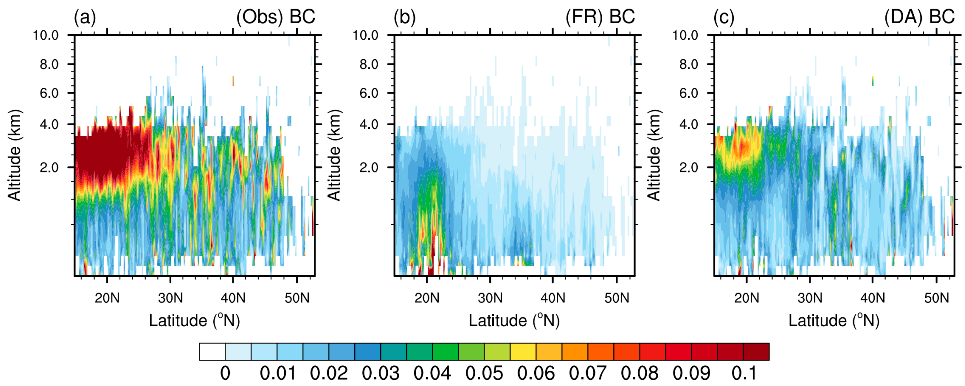

Figure S2. Latitude-altitude cross-sections of black carbon (BC) extinctions (km^{-1}) averaged over the East Asian longitude. The panels show (a) CALIOP observations, (b) FR, and (c) DA experiments.

Changes in Manuscript: Please refer the description in the revised manuscript, Line 401-407 and Fig. S2 in the supplements.

Comment NO.27: *p. 15, sect 4.2: It would be very interesting if authors can also provide an evaluation of the Kudo dataset with AERONET and ground-based LIDAR. This would help understand how much the validation of DA is driven by errors in the Kudo dataset vs errors in the assimilation itself.*

Response: We thank the reviewer for this insightful comment. We agree that whether the errors in the DA validation originate from the assimilated observations or from the assimilation system itself is a very interesting and important perspective. However, a comprehensive evaluation of the Kudo dataset against AERONET has already been thoroughly conducted and documented in the original algorithm paper by Kudo et al. (2023). Their study demonstrated that the retrieved AOTs at 532 nm and 1064 nm agree well with AERONET, although there are inherent uncertainties in SSA and size distributions. Regarding ground-based LIDAR, Kudo et al. (2023) also noted that vertical extinction profiles are subject to retrieval assumptions and are planned for future independent validations.

While the observational dataset inevitably contains uncertainties that propagate into our validation, the primary focus of our current study is the development, implementation, and demonstration of the component-resolved 3D assimilation framework, rather than apportioning the exact sources of validation discrepancies.

Comment NO.28: *p.20, l. 419: if we assume that the AERONET sites are relatively close to sources (closer than CALIOP), this suggests that emissions may be too high but lifetimes too short? If this idea has merit, please discuss it in the paper.*

Response: We sincerely thank the reviewer for this insightful and constructive comment. We completely agree that this is a highly reliable and meaningful explanation.

AERONET sites are predominantly located in or near populated areas and anthropogenic emission sources, whereas CALIOP provides broader regional coverage, including remote and oceanic regions. The contrasting bias pattern observed in our background model, characterized by a local overestimation of absorbing aerosols at AERONET sites and a regional underestimation across the broader CALIOP domain, strongly points to an imbalance in the modeled aerosol lifecycle. Specifically, it implies that the emissions of absorbing aerosols might be overestimated, while their atmospheric lifetime is too short, preventing sufficient long-range transport to the broader domain.

Changes in Manuscript: We have added this discussion into the revised manuscript: "This contrasting bias between the local overestimation at AERONET sites and the regional underestimation compared to CALIOP can likely be attributed to the imbalance in the modeled aerosol lifecycle. Since AERONET sites are generally located closer to anthropogenic emission sources than the broader spatial coverage of CALIOP, this suggests that the emissions of absorbing aerosols might be overestimated, while their atmospheric lifetimes are too short. Consequently, these aerosols are heavily concentrated near the sources but fail to be adequately transported to broader regions." Please refer the description in the revised manuscript, Line 512-517.

Comment NO.29: *p. 21, l 451: "traditional assimilation of total AOT merely scales" This is incorrect and depends strongly on the details of the assimilation technique. Even by assimilating only AOD can one change profiles and speciation because the AOD observations include locations both near and far away from sources and so influence the analysis at different parts of the transport. In addition it is entirely possible (and has been done many times) to include either/both speciation and profile into the state vector. I am not saying that speciated observations (as in this paper) are not important (I think they are) but the authors' statement is simply incorrect.*

Response: Agree. We have modified this incorrect statement as below: "It is worth noting that traditional assimilation of total AOT lacks direct observational constraints on vertical speciation, relying heavily on the model's background covariance to distribute increments, which can sometimes struggle to correct the aerosol composition ratios."

Changes in Manuscript: Please refer the description in the revised manuscript, Line 551-553.

Comment NO.30: *p. 21, fig 12: please explain the downward spikes seen in SSA for the DA results in a), b) and d).*

Response: The occasional sharp drops at Gandhi_College and Dhaka_University are primarily driven by the data assimilation process during specific satellite overpasses. When the satellite captures a dense plume of highly absorbing aerosols (like BC or dust) and the assimilation using this data, the system instantaneously adjusts the scattering-to-absorbing aerosol ratio, leading to a sudden drop in the simulated SSA at that specific time step.

Changes in Manuscript: Please refer the description in the revised manuscript, Line 545-549.

Comment NO.31: *p. 22, l. 467: "warming suggests that the assimilation effectively corrects" This is a really strange conclusion because you have no independent heat profile data to verify against. Also, the Da results have already been evaluated against independent AOD, extinction and SSA data. I suggest removing or completely rephrasing.*

Response: Done. We have removed this sentence.

Comment NO.32: *p. 26, l. 520: "the DA experiment succesfully recosntructed the vertical profiles of all four aerosol components". I think the authors overstate their results. First: there is no independent vertical profile data on speciation that confirms this result. Second: it is not a surprise that self-validation looks so good because the authors use a background correction scheme that makes the analysis identical to the assimilated observations. The authors should rephrarse their statement and clearly describe the limitations I just mentioned.*

Response: Done. We have revised the sentence as below: "In contrast, the DA experiment effectively constrained the simulated vertical structures, yielding analysis fields that are highly consistent with the assimilated CALIOP observations. However, we acknowledge two limitations in this evaluation. First, comparing the analysis field against the assimilated CALIOP observations serves as a self-validation, rather than an independent validation. Second, without independent profile data for specific aerosol species, we cannot fully verify the exact distribution of the four individual components. "

Changes in Manuscript: Please refer the description in the revised manuscript, Line 629-634.

Comment NO.33: *p. 26, l. 535: "internal mixing" while it is important to discuss this limitation in the Conclsuion I think it should already be discussed in Methods.*

Response: Done. Please refer to our detailed response to Comment NO.13.