

1. In my previous Major Comment 3, I asked the authors to clarify how the mismatch between the internally mixed CALIOP-MODIS absorbing component and externally mixed GOCART BC propagates into the diagnosed heating. Specifically, I asked whether the assimilation compensates by altering BC mass and whether the reported peak heating of 3.0 K d⁻¹ should be regarded as an upper bound, a lower bound, or a bracketed estimate.

I appreciate that the authors have now carried this issue into Section 4.4. However, the response appears to replace the requested assessment with two assumptions. The authors state that the internally mixed CALIOP-MODIS particle has a higher MEE than externally mixed GOCART BC, and then assume $SSA_{int} < SSA_{ext}$, from which they conclude that the 3.0 K d⁻¹ heating is a conservative lower bound.

The algebra that follows is valid only if those optical-property relationships hold for the specific CALIOP-MODIS and CRTM/GOCART representations used here. Neither relationship is demonstrated in the response. In particular, stating that internal mixing enhances BC absorption per unit BC mass does not by itself demonstrate that it has a larger absorption fraction per unit extinction. Once extinction is constrained by the assimilation,

$$\beta_{abs} = \beta_{ext}(1-SSA) = \beta_{ext} \frac{MAC}{MEE}$$

Therefore, the relevant comparison is MAC/MEE, or equivalently 1-SSA, between the two optical representations. The statement in the response that the “underestimation of the mass absorption cross-section dominates” is precisely the point that requires quantitative support rather than assumption.

The lensing effect can certainly enhance BC absorption per unit BC mass, i.e. increase MAC. However, because the DA constrains extinction rather than BC mass, this does not by itself determine the sign of the heating bias. The corresponding change in MEE must be considered simultaneously. For the author’s lower-bound interpretation to hold, the relative enhancement in MAC must exceed the relative enhancement in MEE:

$$\frac{MAC_{int}}{MAC_{ext}} > \frac{MEE_{int}}{MEE_{ext}}.$$

Only under this condition does MAC/MEE increase and $SSA_{int} < SSA_{ext}$ follow.

I therefore suggest the authors demonstrate this relationship directly for the optical schemes actually used in the study. At 532 nm, and preferably for representative RH conditions, please provide the MEE and SSA (or MAC) for the CALIOP-MODIS core-grey-shell representation and for CRTM/GOCART BC. In particular, comparison of $(1-SSA)_{int}$ and $(1-SSA)_{ext}$ in the form

$$R_{abs} = \frac{(1-SSA)_{int}}{(1-SSA)_{ext}}$$

would directly establish the sign of the effect once extinction is constrained. $R_{abs} > 1$ would support the author’s interpretation that the externally mixed calculation underestimates absorption and that 3.0 K d⁻¹ value is a lower bound; $R_{abs} < 1$ would imply the opposite.

Because the internally mixed representation referred to here originates from the Kudo et al. (2023) core-grey-shell retrieval, the comparison should use the optical properties of the actual absorbing component employed in that retrieval rather than a generic statement about internally mixed BC.

2. So upon closely going through the result again, I have a follow up question for the 3.0 K d^{-1} value! In Lines 563-567 state that Figure 13 shows the mean vertical profiles of SW heating rate along the CALIOP tracks and that, relative to FR, the DA experiment produces a “maximum heating rate increase of approximately 3.0 K day^{-1} ” at 4-5 km. The caption likewise defines Fig. 13a as the mean SW heating profiles and Fig. 13b as the vertical profile of the DA-minus-FR heating-rate difference, with both profiles averaged along the CALIOP track. However, the plotted values do not appear to support a mean difference of 3.0 K d^{-1} . In Fig. 13a, the DA heating rate near 4-5 km is approximately 0.8 K d^{-1} , while FR is approximately 0.5 K d^{-1} , implying a difference of roughly 0.3 K d^{-1} . Consistently, Fig. 13b peaks only slightly above $+0.2 \text{ K d}^{-1}$, approximately $0.25\text{--}0.3 \text{ K d}^{-1}$. Thus, the stated 3.0 K d^{-1} value appears to be about an order of magnitude larger than the mean DA and FR difference actually shown in Fig. 13.

A heating-rate increment approaching 3 K d^{-1} does appear in Fig. 14, but there it corresponds to the upper end of the individual assimilation samples, particularly in the strongly absorbing SSA < 0.85 regime, rather than to the mean vertical-profile difference. Figure 14 is explicitly described as the relationship between assimilation-induced extinction increments and heating-rate increments for individual samples.

This distinction is important because the same 3.0 K d^{-1} value is subsequently used at Lines 571-578 as the basis for the new “conservative lower bound” discussion.

So is the 3.0 K d^{-1} value actually represents the maximum of the mean DA-FR vertical profile shown in Fig. 13b; or the maximum among the individual samples shown in Fig. 14; or a quantity derived using some other spatial/temporal averaging or normalization that is not currently described?

3. In response to comment 1 and comment 5 of the previous review, I appreciate that the authors have responded constructively by adding the DA_NoBypass experiment and the new discussion in Section 5.2. They report that the bypass is triggered at approximately 28.6% of assimilated grid points, including 43.6% for dust, 17.8% for water-soluble aerosol and 24.5% for BC, predominantly at 4-7 km. They further show that, without the bypass, the BC profile CORR increases from 0.26 to 0.79, and interpret this as evidence of “genuine flow-dependent ensemble covariance skill.” They also acknowledge that DA_NoBypass still underestimates strong aerosol layers, whereas the full DA further reduces BIAS and RMSE, and argue that the independent AERONET/AD-Net improvements reflect dynamical propagation rather than forced local agreement. These additions are useful, but Fig. 16 reveals an important distinction that I think should now be made more explicit.

For the domain-averaged BC profile, BIAS/RMSE/CORR change from approximately $-0.026/0.043/0.26$ in FR, to $-0.027/0.038/0.79$ in DA_NoBypass, and to $-0.021/0.031/0.95$ in DA. Thus, the pure LETKF clearly improves the vertical-profile structure and reduces RMSE, but does not reduce the negative bias of the domain-averaged BC profile. The additional reduction in negative bias and strengthening of the elevated layer appear only in the full DA. This is consistent with the authors’ own statement that DA_NoBypass captures structural variation but still underestimates the magnitude of strong aerosol layers.

I therefore have the following remaining questions.

3.1 The species-specific triggering fractions omit sea salt, although Fig. 16a shows a noticeable divergence between the DA and DA_NoBypass sea-salt profiles, particularly aloft. The

denominator of “28.6% of total assimilated grid points” is also not defined, and the vertical distribution is given only qualitatively as predominantly 4-7 km. Please report the sea-salt triggering fraction, define precisely what constitutes an “assimilated grid point,” and provide the trigger frequency as a function of altitude for each species. A simple temporal summary would also help establish whether triggering is broadly distributed through the assimilation period or concentrated in a small number of major plume events.

3.2 Fig. 16 suggests that the standard LETKF substantially recovers the vertical shape of the BC profile, whereas the bypass provides an important additional correction to the magnitude of the elevated layer when the background spread collapses. I suggest stating this distinction explicitly in Section 4.1, Section 5.2, and where the elevated BC reconstruction is highlighted in the Abstract and Conclusions. At present, statements such as the “successful reconstruction” of the lofted BC plume do not distinguish between these two contributions.

3.3 Lines 665-669 argue that, because simple bias-rescaling should not significantly alter “the temporal or spatial variation patterns,” the increase in BC CORR from 0.26 to 0.79 demonstrates genuine flow-dependent skill. However, the Fig. 16 caption states that these statistics are calculated from domain-averaged vertical profiles over East Asia. The CORR therefore measures similarity in vertical-profile shape; it does not demonstrate temporal variability or horizontal spatial variability. Because the bypass is preferentially triggered at 4-7 km, it can also modify the domain-mean vertical-profile shape itself.

I suggest providing, for DA_NoBypass, the same point-wise N, BIAS, RMSE and CORR diagnostics used in Fig. 3, where all valid space-time-height observations are included. This would show whether the LETKF-only improvement also holds across the full spatiotemporal sample rather than only in the shape of the domain-averaged profile.

3.4 Lines 674-678 state that the AERONET and AD-Net improvements reflect “genuine dynamical propagation of the assimilated fields rather than the forced local agreement of the bypass mechanism.” I do not think the present experiments isolate that conclusion. The independent validations use the full DA experiment, which includes the bypass, and aerosol increments introduced by the bypass can themselves propagate dynamically away from the CALIOP tracks.

Since DA_NoBypass has already been performed, I strongly suggest evaluating it against the same AERONET AOT/SSA and AD-Net extinction observations. A simple FR / DA_NoBypass / DA comparison, either in the existing figures or in a compact table, would directly show how much independent improvement remains when the empirical bypass is removed. In my view, this is the most informative remaining diagnostic for separating the two mechanisms.

3.5 Please clarify whether the bypass remained active in DA_Obserr. Section 5.3 attributes the relatively weak sensitivity to increasing the observation-error baseline from 10% to 50% partly to the dominance of representativeness error and partly to the robustness of the flow-dependent covariance. However, Section 5.2 explicitly acknowledges that the background correction ‘bypasses the rigorous uncertainty weighting of the standard Kalman filter’ and that ignoring observational uncertainty during the correction step remains a limitation. More importantly, the definition of R_{corr} in Lines 297-300 contains the localization weight f_i , the background ensemble mean and the observed extinction, but no weighting by the prescribed observation-error covariance. The triggering criteria and subsequent rescaling likewise depend on R_{corr} , the number

of local observations, the 3-sigma screening and the empirical thresholds; when triggered, the rescaled ensemble directly serves as the analysis and the LETKF step is bypassed.

Thus, for a given bypass event, the correction magnitude is not directly affected by increasing the prescribed observation-error baseline from 10% to 50%. The altered observation error could still affect subsequent background states through earlier LETKF updates and thereby indirectly change when the bypass triggers, but the bypass update itself does not contain the observation-error weighting being tested in DA_Obserr.

If the bypass remained active in DA_Obserr, then the weak overall sensitivity of that experiment cannot be attributed solely to the robustness of the flow-dependent covariance. I therefore suggest qualifying this interpretation and reporting the bypass triggering frequency in DA_Obserr so that the contribution of this error-insensitive update pathway can be assessed.

4. In response to Minor comment 1, the authors clarify that the temporal localization scale $\sigma_t=0.01$ is expressed in units of assimilation time slots and that this value restricts the update to observations at the exact current time step, effectively excluding asynchronous observations within the assimilation window (Lines 274-278). I appreciate this clarification; however, I have one question regarding the description of the assimilation framework.

The defining feature of a 4D-LETKF, compared with a 3D-LETKF, is that observations distributed within an assimilation window are compared with ensemble members at their corresponding observation times and their information is propagated through the time-evolving ensemble trajectories to constrain the analysis state. With the current temporal localization setting, observations separated from the analysis time receive negligible weight, meaning that the asynchronous observation component appears to be effectively disabled in the present experiments. This interpretation also appears inconsistent with Lines 329-330, where the authors attribute improved aerosol-field evolution to asynchronous assimilation. If asynchronous observations are effectively excluded by the chosen temporal localization, this statement should be reconsidered or clarified.

Also, If the present configuration intentionally uses only observations at the analysis time, then the assimilation behaves effectively as a 3D-LETKF for these experiments, despite being implemented within a 4D-LETKF framework, isn't it?

5. The authors state that the representativeness error ϵ_{rep} is derived from the standard deviation of sub-grid retrievals within each 25 km grid cell and that it typically dominates the observation-error budget, allowing the total observation error to realistically represent large uncertainties (Line 159-162). This point is central to the interpretation of the DA_obserr experiment in Section 5.3. However, the screening procedure described in Lines 147-150 may affect the magnitude of the estimated ϵ_{rep} .

The authors remove pixel retrievals outside one standard deviation of the grid mean and retain only grid cells with $(N > 3)$ and coefficient of variation < 0.5 . If ϵ_{rep} is calculated from the retained retrievals, the estimated sub-grid variability will be reduced because the extreme values contributing to the representativeness error have already been removed. This is a separate potential source of underestimation from the spatial-clustering issue discussed in Lines 163-168.

Could the authors clarify whether ϵ_{rep} is calculated before or after these screening steps? If it is calculated after screening, please discuss how this filtering influences the estimated

observation uncertainty and the statement that ϵ_{rep} ensures a realistic representation of high uncertainty. A comparison of ϵ_{rep} before and after screening for representative grid cells would help quantify whether this effect is negligible or significant.

The revised manuscript has addressed the major concerns raised during the previous review and presents a valuable contribution to aerosol data assimilation and radiative impact assessment. The remaining issues mainly concern clarification of several methodological assumptions and ensuring that the interpretation of some key results is fully supported by the presented analyses. Addressing these points would improve the transparency and robustness of the manuscript.