Response to Reviewer

We sincerely thank the reviewer for the evaluation of our work and the helpful suggestions to improve the clarity and rigor of the manuscript. Below, we provide detailed point-by-point responses. Comments from reviewers are shown in blue, **our responses in black**, and the corresponding changes made in the manuscript are highlighted in orange.

Comment 1:

The title is vague and doesn't clearly reflect the content of the manuscript. I suggest something like "Retrieval of aerosol component fraction from spectral aerosol optical depth". This describes more clearly what is actually done in the paper. The use of "machine learning" in the title honestly feels like an attempt to add this buzzword in, because the machine learning aspect is in my view not a novel aspect of the work (it's an emulator for MOPSMAP, the work could be done without it, and it's not a particularly complicated or conceptually new use of machine learning).

Response 1:

We appreciate the reviewer's suggestion. However, we respectfully disagree with the characterization that the machine learning component is merely an emulator for MOPSMAP. The aerosol composition is one of the input in the full physics model, MOPSMAP, while aerosol optical properties, such as SSA, AF, and Reff are outputs. In our framework, our machine learning based model adds SSA, AF, and Reff also as inputs. This method enables retrieval of aerosol composition using parameters that are actually observable. We do not simply replicate the MOPSMAP input–output relationship. Instead, we restructure the original model by treating several parameters that are typically outputs of physical simulations (such as SSA, AF, and Reff) as inputs to the machine learning model. Therefore, we believe it is appropriate and necessary to retain the mention of "machine learning" in the title, as it reflects a core part of the paper's contribution.

In the revised version, we have updated the title for clarity, while keeping this essential element: "Retrieval of aerosol composition from spectral aerosol optical depth using a machine learning approach."

Comment 2:

Throughout, especially in the introduction, the term "infrared" is used to describe the FTIR measurements. It is only when we get to the data description that we learn this is about shortwave infrared (SWIR, solar spectral region, 1-3 microns). To me and I suspect to most readers, "infrared" without further specification implies thermal infrared. These are quite different spectral regions with different aerosol behavior. I suggest specifying SWIR throughout when talking about these data, otherwise it is somewhat misleading for the casual reader.

Response 2:

We thank the reviewer for pointing this out. In the revised manuscript, we have replaced all instances of "infrared" referring to the FTIR AOD measurements with "shortwave infrared (SWIR)" for clarity and consistency.

Comment 3

Lines 59-68: there are a lot more examples than these. For example, the GRASP and RemoTAP algorithms use machine learning emulators to replace online radiative transfer calculations and have been widely applied to e.g. POLDER data (and more recently RemoTAP for PACE). There are other approaches (e.g. FastMAPOL, MAPP) which have been extensively published on in recent years, either for airborne polarimeters or more recently also for satellite observations from PACE. So there is already a fairly widespread use of these techniques in routine satellite aerosol data processing.

Response 3:

We thank the reviewer for highlighting these important and widely used approaches. In the revised manuscript (lines 67–74), we have expanded the literature review to include GRASP, RemoTAP, FastMAPOL, and PACE-MAPP as additional examples of retrieval frameworks that incorporate machine learning or emulator-based radiative transfer models. This helps provide a more complete picture of the increasing use of such techniques in operational and research-level satellite aerosol data processing.

L67-74: The FastMAPOL algorithm employs neural network-based forward models within a multi-angle polarimetric retrieval framework, achieving speed-ups of about 1000× with minimal accuracy loss (Gao et al., 2021a). It also includes adaptive view-angle filtering to mitigate retrieval errors from problematic geometries in satellite and airborne data (Gao et al., 2021b). Similarly, the PACE-MAPP algorithm couples atmosphere–ocean vector radiative transfer emulators to jointly retrieve aerosol and ocean optical properties from polarimetric measurements (Stamnes et al., 2023). In addition, algorithms such as GRASP (Dubovik et al., 2011) and RemoTAP (Fu et al., 2020) have integrated radiative transfer emulation strategies and have been widely applied to global aerosol data from POLDER and PACE (Hasekamp et al., 2024).

Comment 4:

Lines 92-95: I am not sure of the purpose of this sentence. Are the authors saying that, in this work, the AOD is used to determine SSA, etc? In that case I think the sentence is not necessary because the method is described later. Or are they saying that AERONET provides these? In that case references should be provided, and note that these derived properties aren't determined by AERONET solely from spectral AOD (and not from all of those channels) but also from sky-scanning radiance measurements.

Response 4:

We have revised the sentence to clarify that SSA, AF, and Reff are not derived from spectral AOD alone, but are part of AERONET's inversion products obtained through sky radiance measurements. And SSA, AF, and Reff are used in input level in ML. We have also replaced the citation with Dubovik (2000) and Giles et al. (2019), which better describe the retrieval process.

L101-105: Standard AERONET sun photometers retrieve AOD at 340, 380, 440, 500, 675, 870, 1020, and 1640 nm, covering the ultraviolet (UV) to shortwave infrared (SWIR) range. In addition to these direct sun measurements, AERONET provides inversion products that include single scattering albedo (SSA), asymmetry factor (AF), and effective radius (Reff), retrieved using sky radiance observations (Dubovik 2000 and Giles et al. 2019). These parameters are useful for aerosol type discrimination and are used in this study.

Comment 5:

Section 2.1 (and tying in to methodology later): to me, it looks like FTIR is providing an AOD at 2 microns but the other bands are the same as or match AERONET values. So the value of the FTIR seems a bit overstated in the manuscript, as it is emphasized repeatedly. None of the analysis shows the value of adding this band in particular. I would rather have seen the use of the UV bands in the network. For an eventual application to AERONET data globally, 2 micron data are not available, while the 340/380 nm band pair more often are. It reads a bit like the choice of wavelengths was motivated by the specific bands available for the case study at Ny-Ålesund, but it isn't justified that this makes sense more generally. It feels a bit like the authors had access to these measurements and then tried to find something to do with them, as opposed to developing the approach and then finding appropriate case studies to look at. Perhaps the history is otherwise but it is not well-justified by the manuscript as submitted. Otherwise why pick Ny-Ålesund and why add the FTIR observation when it isn't present at the hundreds of other AERONET sites in the network?

Response 5:

This study was indeed initiated based on the availability of high-quality, co-located FTIR and AERONET measurements at Ny-Ålesund. In fact, to our knowledge, Ny-Ålesund is one of the few sites globally where such co-located FTIR and AERONET observations exist. Our intent in this study is not to claim broad representativeness, but rather to demonstrate the scientific value of this rare dataset and encourage the aerosol remote sensing community to consider integrating FTIR measurements into more locations in the future.

The key motivation for emphasizing the FTIR measurement is that, based on our experience and prior research, the ultraviolet (UV) channels (e.g., 340/380 nm) offer very limited sensitivity to aerosol composition. While they are useful for constraining aerosol size and absorption, their spectral signatures do not allow robust separation of composition types such as sulfate, sea salt, or dust. In contrast, the shortwave infrared

(SWIR, 1–2.2 µm) region, particularly above 1.5 µm, has higher sensitivity to coarse-mode components, as also seen in Figure 2 of the manuscript.

Comment 6:

Section 2.2, title: It is not accurate to say the satellite provides AOD "measurements", they are retrievals.

Response 6:

We agree and have revised the wording to "retrievals" accordingly.

Comment 7:

Section 2.2: to be clear, is it the case that the analysis takes the VIIRS Deep Blue monthly spectral AOD over ocean and use the Ångström power law to adjust the wavelengths to match the ones chosen for the MOPSMAP-based training? Is that right? I don't know that it makes sense to do the decomposition based on monthly AOD data. To me that seems equivalent to the assumption that aerosol composition is constant over a month, which will not be valid in some locations (including the North Atlantic chosen for the case study later). It is not obvious that calculating composition on an instantaneous or daily basis and then averaging the results rather than calculating composition based on monthly-averaged AOD would give the same results. Most of the error in satellite AOD is not random noise but rather factors related to geometry, surface type, and optical property assumptions at the given location and time – which do not necessary decrease on temporal averaging, so I don't think there is a justification on those grounds. So this choice should be better justified and its advantages and limitations discussed.

Response 7:

Yes, we use VIIRS Deep Blue monthly mean spectral AOD data over ocean, and apply the Ångström power law to interpolate AOD values to the spectral bands used in the MOPSMAP-based training.

We agree that applying the retrieval method to monthly averaged AOD data implicitly assumes relatively stable aerosol composition over the averaging period, which may not be valid in many locations, including the North Atlantic. We emphasize that our retrieval framework is applicable to individual observations (e.g., daily or instantaneous AOD spectra), and using monthly AOD in this study is a first-step demonstration. For certain regions like the North Atlantic, we believe the monthly mean AOD field can still represent a meaningful background aerosol regime, though we acknowledge its limitations. We plan to explore daily-level retrievals in follow-up work, and may test this in the current case study as the reviewer suggests.

Comment 8:

Asymmetry factor is sometimes summarized as AF and sometimes as Af. This should be made consistent.

Response 8:

All corrected to AF.

Comment 9:

Section 3: I think the whole methodology sections should be rewritten in a different order and with more information. Much of the text and Figure 1 are confusing, and some of the information needed to understand it and section 3.0 is not given until sections 3.1 and 3.2. I think the authors need to be clear what exactly is taken from MERRA2 components: is it just realistic mixing fractions? Or is it also the spectral complex refractive index and size distribution (and therefore also SSA, asymmetry factor)? Are the fractions in terms of mass, volume, area, AOD at some wavelength? What distributions were drawn from to sample these parameters, and how are they justified?

Response 9:

The whole methodology sections have been rewritten to include more information. The machine learning approach presented in this study is not designed to accelerate an existing physics-based model such as MOPSMAP. Rather, it is intended to enable a new kind of model that reorganizes the traditional input–output structure of aerosol optical properties simulations in a way that reflects actual observational capabilities and retrieval needs.

Specifically, as shown in Fig.1 (a new version), MOPSMAP takes as input a variety of aerosol microphysical parameters, including component fractions, particle size distribution, complex refractive index, particle shape, and ambient relative humidity. It then computes corresponding optical properties, such as single scattering albedo (SSA), asymmetry factor (AF), and effective radius (Reff). In contrast, these optical properties are usually observable using instruments such as sun photometers (AERONET). We therefore reorganize the MOPSMAP simulation inputs and outputs according to actual observational conditions. In summary, part of the original MOPSMAP outputs (SSA, AF, and Reff) are repurposed as inputs to a machine learning model. This is not achievable by the traditional forward simulation itself, but can be enabled by data-driven learning.

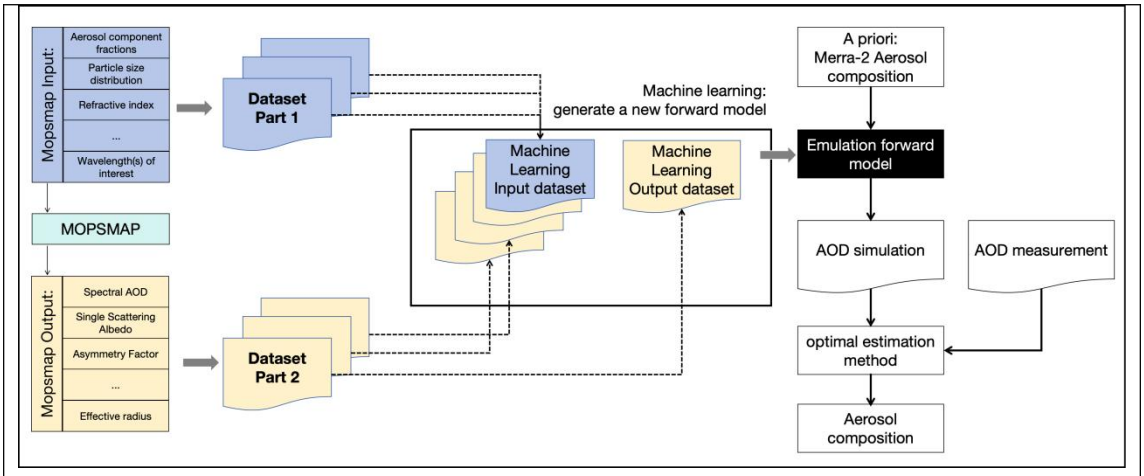

Figure 1. Schematic overview of the model development and retrieval workflow. Left panel: MOPSMAP is used to generate a synthetic database by simulating optical properties from randomly sampled aerosol parameters, including component fractions, size distribution, and refractive index. Based on the subset of optical parameters that are typically available in real observations (e.g., spectral AOD, SSA, AF), we reorganize the simulated database to define the machine learning inputs. The corresponding aerosol component fractions are selected as the machine learning outputs. This effectively inverts the original MOPSMAP input-output structure to train an efficient, observation-driven emulator of the forward model. Right panel: The trained ML-based forward model is then used within an optimal estimation framework. A Prior aerosol composition is taken from MERRA-2 (Gelaro et al., 2017), and AOD is simulated using the learned forward model. By minimizing the mismatch with observed AOD measurements, the posterior aerosol composition is retrieved. This setup allows flexible integration of real-world measurements and efficient inversion without full-physics simulation during retrieval.

L127-157:

3 Method

3.1 Overview of the Methodological Framework

This study presents a hybrid framework that combines physics-based aerosol optical modeling, machine learning (ML), and optimal estimation method (OEM) to retrieve aerosol composition from multi-wavelength AOD observations. The key idea is to reorganize the physical input-output structure of forward model (MOPSMAP) to align with what is actually observable. Instead of directly retrieving aerosol properties from a full-physics model, we first generate a synthetic database using MOPSMAP, in which each sample includes both microphysical inputs and resulting aerosol optical outputs. From this database, we construct a new forward model using machine learning. Specifically, the ML model takes as input the aerosol component fractions (i.e., number concentration of sea salt, sulfate, black carbon, dust, and insoluble aerosols), as well as auxiliary parameters such as single scattering albedo (SSA), asymmetry factor (AF), and effective radius (Reff), parameters that are typically accessible from remote sensing observations. The output of the model is the spectral aerosol optical depth (AOD), originally computed by MOPSMAP. Once trained, this ML-emulated forward model is embedded within an optimal estimation framework.

Observed spectral AOD (e.g., from ground-based measurements or satellite retrievls) is used as the retrieval input, and aerosol component fractions are estimated iteratively by minimizing the mismatch between observed and simulated AOD, under constraints from a prior information and measurement uncertainty.

Specifically, as shown in Fig. 1, MOPSMAP takes as input a variety of aerosol microphysical parameters, including component fractions, particle size distribution, complex refractive index, particle shape, and ambient relative humidity. It then computes corresponding optical properties, such as SSA, AF, and Reff, which are usually observable using instruments such as sun photometers. We therefore reorganize the MOPSMAP simulation inputs and outputs according to actual observational conditions. In summary, part of the original MOPSMAP outputs (SSA, AF, and Reff) are repurposed as inputs to a machine learning model. This is not achievable by the traditional forward simulation itself, but can be enabled by data-driven learning.

To implement the proposed aerosol composition retrieval framework, we follow a structured approach consisting of three main steps. These are outlined as follows and will be described in detail in the following sections:

1. Synthetic Dataset Generation: A large AOD dataset is generated using MOPSMAP by varying aerosol component fractions, as well as four physically-constrained parameters: single scattering albedo (SSA), asymmetry factor (AF), effective radius (Reff), and relative humidity (RH).

2. Machine Learning Forward Model: A neural network is trained to emulate a new forward model, mapping input parameters (the aerosol component fractions, SSA, AF, and Reff) to multi-wavelength AOD spectra.

3. Retrieval via Optimal Estimation: The ML-based forward model is integrated into an optimal estimation framework to retrieve aerosol composition from observed AOD.

Comment 10:

Section 3: I guess one of my issues with the manuscript is I don't really understand the point of adding machine learning to this, as opposed to using MOPSMAP directly, aside from speed. If it is just speed then I think the machine learning nature of this work is a bit overhyped. It might be that there is some detail I am missing about what is done, but the manuscript is not clear enough to say.

Response 10:

We would like to clarify that the role of machine learning in this work goes beyond accelerating MOPSMAP. In fact, we have tested using MOPSMAP directly as the forward model in the inversion framework, but the retrieval results are extremely poor and often fail to converge. As we methoned in last comment, the key innovation in our method is that we restructure the problem: by taking part of MOPSMAP's original outputs (SSA, AF, Reff) as inputs to a machine learning model, we enable

retrieval of the unobservable aerosol composition. This inversion would not be possible using MOPSMAP directly. We have added clarification of this point in the revised Section 3.1.

Comment 11:

Section 3.2: what is the justification for this network architecture? Some more references to the NN methods/code bases should be added as well. Can we see training and validation subset loss functions from the training process? Can the authors demonstrate that 10,000 simulations is enough for a comprehensive sampling and train:val:test subset, given this seems a fairly high dimensional problem, and does not lead to overfitting?

Response 11:The chosen network architecture (two hidden layers with 32 neurons each) follows established design patterns commonly used in aerosol retrieval and radiative transfer emulation studies (Faure et al., 2001; Nanda et al., 2019; Chen et al., 2022). We also conducted sensitivity tests with deeper or wider networks, which did not significantly improve validation accuracy but increased training time. Regarding the justification of our neural network design and training configuration (Section 3.2), we provide the following clarifications:

- Architecture: The network architecture (two fully connected layers with 32 units each, activated by leaky ReLU) was selected based on empirical testing to balance expressiveness and regularization. The relatively shallow depth and small parameter count help avoid overfitting in this moderately sized dataset.
- Sampling: The 10,000 MOPSMAP simulations were designed to cover nearly all feasible combinations of aerosol component fractions and optical parameters, ensuring representative coverage of the entire relevant parameter space. This synthetic dataset includes diverse mixing scenarios that reflect both common and edge-case aerosol compositions observed in the atmosphere.
- Training stability: The training loss (Figure A1) exhibits smooth convergence across five orders of magnitude and stabilizes below 10^{-4}, indicating robust learning without instability or divergence.

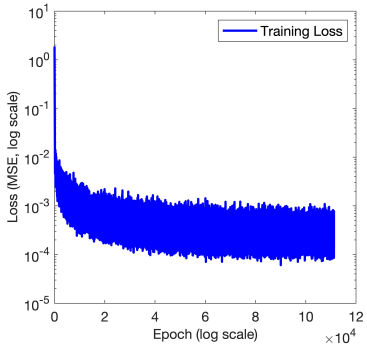

Figure A1: Training loss curve (mean squared error, MSE) plotted on a log–log scale as a function of training epochs. The model shows stable convergence over 10,000 iterations, with final loss values below 10^{-4}, indicating good learning performance and absence of overfitting. The MSE is computed based on normalized AOD spectra simulated from the training dataset.

Comment 12:

Section 3.3: what values, specifically, are taken for x_a and S_a and how are they justified? These are important, particularly for the later discussion about averaging kernels and uncertainty estimates, both of which depend on the strengths of the prior constraints. It is mentioned that S_y is the measurement (input AOD) uncertainty; what numbers specifically are used here? Is it 0.01 or is there a more detailed description of AERONET and FTIR uncertainty? Is there any spectral correlation assumed?

Response 12:

L245- 254: The a prior vector x_a is derived from the MERRA-2 monthly mean aerosol component fractions at the same time and location as the FTIR observations. The a prior covariance matrix S_a is set as a diagonal matrix with variance 0.01 (i.e., standard deviation of 0.1) for each aerosol component. This reflects a relatively loose prior constraint, allowing the retrieval to be primarily informed by the spectral AOD observations while maintaining physical plausibility. For the measurement error covariance matrix S_γ, we distinguish between visible and shortwave infrared (SWIR) wavelengths. For visible bands (AERONET-like observations), we adopt 0.01 as the standard deviation, consistent with the reported uncertainty of AOD retrievals from AERONET. For the infrared bands (SWIR), we adopt 0.02 as the standard deviation, based on reported uncertainties from Barreto et al. (2020) and Alvárez et al. (2023), who applied Langley calibration for FTIR-based AOD measurements in the SWIR region. All uncertainties are assumed to be spectrally uncorrelated, and S_γ is constructed as a diagonal matrix. This assumption has been clarified in the revised manuscript.

Comment 13:

Line 235: are these 1500 cases the "Test" split of the original data, or another randomly-chosen 15%? This should be written more explicitly.

Response 13:

A total of 1500 cases were randomly selected (15% of the full 10,000-sample dataset) for this analysis; they do not represent a designated "test" set.

L268: Instead, a total of 1500 cases are randomly selected (15% of the full 10,000-sample dataset, they do not represent a designated "test" set.)

Comment 14:

Section 4.2 and Figure 4: are these results shown for the "Test" subset? I am not sure that R2 (which the text focuses on) is the relevant metric here; I'd think that the AOD reconstruction RMSE is. Also, as a practical matter, the AODs shown in Figure 4 are

all very high. The lowest 500 nm AOD in figure 4(a) is about 0.7. This is a magnitude rarely seen except in an extreme aerosol event like fire or a dust storm. So an uncertainty analysis of the posterior composition based on this distribution will greatly overstate the actual practical utility of the algorithm, because the AODs are so high that measurement uncertainty is negligible. In practice the true AOD is often likely to be a factor of 3 or so lower, so the relative uncertainty about a factor of 3 higher. In short, the results of this theoretical uncertainty analysis based on simulations are likely to overstate the performance of the method. This will influence the discussion in these sections, including e.g. averaging kernels and relative contributions of different terms to overall posterior uncertainty.

Response 14:

Our retrieval framework operates on spectrally normalized AOD, which reduces dependence on absolute AOD magnitude and emphasizes spectral shape. However, we acknowledge that in low-AOD conditions, the relative uncertainty in normalized AOD increases substantially, which in turn increases the posterior uncertainty and reduces the information content, making the retrieval more dependent on prior assumptions.

Conversely, in high-AOD scenarios (e.g., dust outbreaks, pollution episodes), the relative contribution of observational uncertainty is reduced, and the retrieval becomes more observationally driven. These are often the most scientifically and societally relevant conditions (e.g., during aerosol transport or climate events), where accurate aerosol composition information is most needed.Nevertheless, we agree that the current analysis may underestimate uncertainties in clean-sky conditions, and have added discussion in the manuscript clarifying that the method may be less reliable in remote or pristine environments with very low aerosol loading.

L177- 185: To focus on the spectral shape of AOD rather than absolute magnitude, we scale all wavelengths relative to the 440 nm value. This results in a relative AOD spectrum, with AOD(440 nm) set to 1.0 and all other wavelengths scaled accordingly. The retrieved outputs represent number concentration fractions of aerosol components. Then, in post-processing, these fractions can be combined with observed or modeled AOD magnitudes to calculate component-specific AODs. However, due to differences in size-dependent extinction efficiency among aerosol types, the AOD contribution of each component is not strictly proportional to number fraction. Therefore, while the absolute AOD values of individual components may have some degree of uncertainty, the spatial distribution and relative dominance of each aerosol type remain meaningful, particularly in high-AOD cases.

Comment 15:

Section 4.4/Figure 5: again, are the “component fractions” defined in terms of mass, volume, area, number, AOD at some wavelength, something else?

Response 15:

The retrieved component fractions are defined in terms of number concentration.

Comment 16:

Line 310: I'm not sure we need this many significant figures for the average averaging kernel. It would be good also to show somehow the variability of the averaging kernel matrix between these simulations (e.g. standard deviations of each element). That will help to show whether the information content varies significantly across the ensemble of cases.

Response 16:

We agree that full precision is not necessary, and the averaging kernel matrix is mainly provided to convey the general information content and sensitivity structure of the retrieval. Presenting the mean kernel is a commonly used approach in OEM retrieval studies to provide a first-order assessment. We believe this suffices for the purpose of illustrating typical retrieval behavior in this case.

Comment 17:

Line 314: instead of just the average degrees of freedom, how about also showing the mode? What's the interquartile range or standard deviation or similar? This ties in to the above point.

Response 17:

Not necessary. The degrees of freedom shown here are based on a stable set of test retrievals with consistent uncertainties, rather than on a large ensemble of independent observations. As such, the mean value provides a representative measure, and further distributional metrics (e.g., mode or interquartile range) are not required.

Comment 18:

Section 4.5: the discussion here and figures talk about MODIS, but the introduction to the paper says VIIRS data were used. Which is it, VIIRS or MODIS? My intuition says VIIRS because I don't think the classification shown in Figure 6(b) is provided in MODIS, only VIIRS (though I could be wrong). Also see previous discussion about whether it makes sense to do this on monthly data as opposed to daily then averaging. Also, what assumed satellite uncertainty is taken for this example retrieval, and what is its assumed spectral correlation?

Response 18:

We thank the reviewer for catching this inconsistency. The aerosol type classification shown in Figure 6(b) is provided by VIIRS and not MODIS. The classification is based on the "Aerosol Optical Model" flag available in the VIIRS AOD product. We have clarified this throughout the manuscript.

Comment 19:

Line 338: another option is taking SSA, asymmetry factor etc from the values used by the algorithm. This would keep consistency with what the retrieval assumed.

Response 19:

Thank you for the comment. We have now revised the manuscript to make this clearer.

L376-379:

1. One approach is to supplement satellite AOD observations with additional physical parameters such as SSA, AF, and Reff, obtained from other satellite products or reanalysis data, as auxiliary inputs to the retrieval algorithm.

2. Another approach is to treat these physical parameters, represented as θ in the forward model $F(x; \theta)$, as part of the state vector x, allowing them to be retrieved jointly with aerosol composition.

Comment 20:

Figure 7: This should be redrawn to use the same map projection and latitude/longitude boundaries for both panels. Having them different makes it difficult to compare the results.

Response 20:

This figure has be redrawn to use the same map projection and latitude/longitude boundaries for both panels. Regarding the comparison itself, it is important to clarify that the VIIRS product only provides aerosol type classification, not actual AOD values. Therefore, a direct quantitative comparison with GEOS-Chem AOD is not possible. Instead, our intent is to qualitatively compare the spatial distribution patterns between the VIIRS-derived aerosol types and the retrieval-based dominant aerosol types from GEOS-Chem. This comparison still provides insight into whether the retrieval framework captures realistic aerosol regimes.

Comment 21:

As a general methodological point: I could not fully judge this study because of the missing information described above. But conceptually, I find the idea that one can take spectral AOD and use this to get at weights of 5 components to seem unrealistic. Since aerosol extinction is spectrally smooth, the different AOD wavelengths are not orthogonal and really there are maybe 3 pieces of information in the AOD spectral (AOD magnitude and maybe two parameters related to spectral curvature as

Ångström exponent is often represented as a log-log quadratic function). So to get 5 components weights out of this seems speculative and it must weigh heavily on the a priori constraints (which are not discussed in detail in the current version of the manuscript). This is borne out somewhat by the averaging kernel analysis which shows the prior is fairly important, especially for black carbon. For realistic aerosol loadings (as mentioned previously, about a factor of 3 lower than in the analysis presented), the uncertainty seems likely to be very high. I think we need a lot more detail on the underlying distributions all these cases were drawn from before we know how robust the results are, and there should be examples of averaging kernels drawn from more realistic aerosol loadings.

Response 21:

We appreciate the reviewer's thoughtful comments. Since several of the concerns raised here can also be detailed explained by earlier points, we provide a concise summary response:

First, our retrieval does not rely solely on spectral AOD. It combines spectral AOD with additional observed parameters such as single scattering albedo (SSA), asymmetry factor (AF), and effective radius (Reff), which help constrain the inversion and improve distinguishability among aerosol types.

Second, we agree that in any optimal estimation framework, the prior plays an important role, particularly when observational uncertainties are large. As shown in the averaging kernel analysis, some components (e.g., black carbon) are more influenced by the prior, which reflects the current limitation in sensitivity. We acknowledge this and note that future improvements, such as incorporating more spectral bands or additional measurement types, could enhance retrieval performance for such components.

Third, we emphasize that our retrieval is based on normalized AOD spectra, where all AOD values are scaled relative to the 440 nm band. This removes sensitivity to AOD magnitude and allows us to focus on spectral shape, which is more stable across aerosol loading conditions. We have clarified this point in the revised manuscript.

Lastly, we agree that retrieval uncertainty will be larger under low-AOD conditions, where observational noise has a relatively greater influence. However, many real-world aerosol episodes of interest, such as dust storms or pollution events, do exhibit elevated AOD levels, and the method is especially suited for such cases. We believe that even with these limitations, the approach provides valuable new insight into aerosol composition from readily available multi-spectral observations.