Review 1

This is a neat paper, a very practical application of data and modeling efforts that the authors and others have been pursuing for years. The study reaches clear recommendations, which is especially gratifying.

We thank the reviewer for taking the time to read and review our paper. We are very pleased to hear their enthusiasm for the findings and recommendations.

Lines 183-184. This is not a complete sentence.

Thank you – now complete.

Clearly the near-surface dust concentration is especially important, and I know that CALIPSO sensitivity tends to diminish within the lowest 75 – 100 m of the surface. Is the reason that CAMS produces lower dose than CALIPSO for near-surface dust concentrations the assumption that the CALIPSO extinction at 100 m is extrapolated to the surface? The confidence with which you can assess the elevation of a near-surface concentration peak seem especially relevant based, e.g., on Figures 4c, 4e, 5c, and 5e, and it comes up again in the discussion of Figure 8. (I now see some discussion in lines 495-502. And I agree it is surprising in light of CALIOP being lower than other measurements. I'm wondering whether there is any EarliNet data that might help here.)

In terms of EARLINET sites, these cover Europe and therefore cannot be used to validate our profiles in this study. However, the suggestion of ground-based lidar validation is valuable. In fact this is the topic of ongoing research, and will be published separately in due course. We discuss the potential application of ground-based lidar data in the conclusion. As the reviewer points out we discuss these differences in the lowest portion of the vertical profile in the discussion. We have added a signposting sentence to the beginning to the results section so the reader knows the differences are discussed later on, "In this section we present the key results and findings of this work. A full discussion of the potential causes for differences between datasets is given in Section 5."

Section 2.3.3. I think you do as good a job as one can with available satellite data in constraining the dust mass concentration. But it might be worth also making some assessment of the uncertainty in mass concentration, e.g., associated with the uncertainty in the MEC. (I realize you compare the model and two lidar estimates with each other, which might represent a rough estimation of uncertainty due to s in Equation 1, but I think you are using the same MEC to obtain concentrations for all three.) (Again, I now see you say a bit more about MEC uncertainty in lines 503-513. Ok, so is the recommendation that we need better measurements of dust MEC?)

Yes, we certainly agree that we need better measurement constraints on dust MEC. Additional text has been added to the conclusions to emphasize this, "(particularly mass extinction coefficient, which is a crucial property in relating model-calculated mass loading to satellite-derived optical retrievals)."

Further, is there any available data on actual engine wear, that might be used to test or constrain the overall results?

The simple answer to the question is no, we don't have sufficient understanding of engine damage rates, relative to the dust dose, to be able to back calculate the dust dose from the amount of damage. Evidence of dust damage exists, but we don't know how much dust it takes to get that level of damage. This is why we want accurate – or as accurate as possible – data on the dust in the atmosphere and thus how much dust dose relates to the damage we observe in engines, and hence the motivation for this study.

Further, engine manufacturers *are* beginning to perform controlled laboratory engine dust tests where we know the dose and can relate this to the damage observed However, these are only in their infancy due to extreme costs involved. There are also challenges involved in reproducing in-service dust conditions, and composition of dust adds an additional complexity.

We have added the following text to the introduction to provide more information:

"Mineral dust can cause engine damage through various different mechanisms. A comprehensive description is provided by Clarkson et al. (2020), who describe how dust particles cause erosion of compressor blades, vanes and of seals between engine components, causing loss of efficiency. Dust particles can melt in hot sections of gas turbines (combustors and core turbines) and deposit on surfaces, reducing aerodynamic efficiency, damaging ceramic thermal barrier coatings, and blocking aerofoil film cooling hole features. Dust particles can also enter secondary engine air systems where they restrict cooling air flow, causing overheating or component deterioration. All these processes lead to reduced efficiency and reduced component lifetime. Resultant losses in efficiency can increase aircraft fuel burn and therefore result in increased aviation emissions of greenhouse gases, thus linking mineral dust to aviation's climate impact (e.g. Lee et al., 2021). The damage done by dust depends on the altitude and power of the engine – in part since this determines the quantity of dust ingested, but also because these factors affect dust particles' behaviour during their transit through the engine."

"Reports within the aviation industry suggest that engines operating regularly in dusty airports show evidence of accelerated component deterioration (Clarkson et al. 2020). Although this evidence exists, there is a lack of knowledge of the amount of dust required to cause such damage. Although engine manufacturers are beginning to perform controlled tests where dust concentrations are known and damage is observed, these incur extremely high costs and are currently limited."

Figure 3. I know there are some seasonal and altitude differences, but if you could reorder the entries in the legend for this figure so they are generally in the order of dust concentration magnitude, it would be easier to discern the lines associated with the lowest few, which seem to be Phoenix, Hongkong, and Sydney.

We have rearranged the order of the cities in the legend as suggested.

I understand that you are effectively using seasonal background dust levels for these calculations, and I know that for some phenomena such as dust transport in general, extreme events dominate. So, I'm wondering (a) how well the limited CALIPSO sampling and the CAMS model simulations capture sporadic larger dust events, (b) whether the airports in question shut down when dust loading is unusually elevated, and (c) how the dosage for even one elevated event for which the airport might remain open might compare with the typical seasonal averages.

These are good questions.

a) For CAMS climatology (e.g. fig 3), where we represent all model events, temporal sampling is not an issue since we use all model output. In terms of how well the model reanalysis represents larger, sporadic events, this is beyond the scope of the study. However, Errera et al. (2021) find that CAMS reproduces the annual variability fairly well compared to AERONET for dust with correlations between 0.84-0.87. They also show that over the Sahara, the largest DOD underestimations are observed during summer, and related to mesoscale convective events which the model cannot reproduce. Errera et al. (2021) and Bennouna et al. (2023) also state that the CAMS reanalysis reproduced a few recent individual large Saharan dust events reasonably well, though AODs are underestimated.

For CAMS/CALIOP profile and dose comparisons, the temporal sampling potentially becomes important. Figures 4, 5 and supplement figures indicate the number of profiles included in the long-term means shown. For example, taking the Canary Islands for DJF as an example, the CALIOP L3 data includes 146 profiles over 13 years, which is fairly low fraction sampled. Order of magnitude similarities in the sampling also apply to the LIVAS data. Therefore the satellite data are clearly a small sample of the events which occurred. We have not investigated the representativity of CALIOP overpasses as a function of larger dust events. This is beyond the current scope of the work. However, it is reassuring that the CAMS data, also sub-sampled to match the CALIOP overpasses in figure 4/5, shows similar vertical structure and magnitude compared to that using all the data (figure 3), indicating that the climatologies constructed from CALIOP L3/LIVAS sampling are reasonable.

b) Airports and air traffic control have procedures for visibility reductions in place ('Low Visibility Procedures' (LVP) and 'Reduced Aerodrome Visibility Procedures'), which vary depending on how low visibility drops. In extreme events, airports have been reported to have temporarily cancelled or postponed flights due to extreme dust events. When visibility falls beneath around 400m, airports follow these procedures, which generally results in greater spacing between aircraft and therefore reduced airport capacity. We added, "Moderately dusty conditions at airports cause reduced visibility which can require greater spacing between aircraft, and thus reduce airport capacity (ICAO 2023)," to the introduction.
c) Again, this is a very interesting question, but beyond the scope of the current work. Ongoing and future work within our research groups will investigate these questions.

Line 387. The end of the sentence is missing.

Cross-reference to section 5 now included.

Line 576. Might visibility still be an issue in the vicinity of a busy airport, even after an aircraft has left the ground?

No, it isn't a problem. Visibility is a problem on the ground because of the risk of aircraft bumping into other aircraft, ground vehicles and other obstacles. Once the aircraft is in the air the only obstacles to avoid would be other aircraft, but air traffic management procedures keep aircraft a long way apart, whatever the level of visibility. We added, "it is not a concern once aircraft have taken off since air traffic control keep large distances between aircraft, whatever the visibility."

References

Bennouna, Y. et al., 2023, Validation report for the CAMS global reanalyses of aerosol and reactive trace gases: Period 2003-2022,