Dear Atmospheric Measurement Techniques and Atmospheric Chemistry and Physics Editor,

We are pleased to submit our final manuscript titled "Aerosol layer height (ALH) retrievals from oxygen absorption bands: Intercomparison and validation among different satellite platforms, GEMS, EPIC, and TROPOMI". We have carefully considered the comments provided by the reviewers, and we have made the necessary modifications to our manuscript. Below, we provide a detailed response to each comment, along with the corresponding changes made in the manuscript. Our responses (in blue) for each comment (in black) and updates to the paper (in *italics*) are provided below.

Thank you so much for your time and consideration.

Authors' response to RC1

Synopsis: This is a pretty straightforward paper comparing passive Aerosol Layer Height (ALH) retreivals generated from TROPOMI, GEMS and EPIC over two Asian domains for signciant dust and smoke events. Comparisons are made to AERONET for AOD, CALIPSO CALIOP products for the vertical centroid, and intercomparring heights between each other. Also included are comparisons of the nature of diurnal retrievals and are two example cases. As the authors note, the oxygen band retrievals are coming into their own, and while they have limited measurement degrees of freedom, they do have coverage that cannot be achieved with lidar systems. Overall, the paper is certainly appropriate to AMT and can have importance to the community. This said, I do think the paper requires major revisions. These revisions are in two prime areas.

First, they authors delicately ignore a number of sampling considerations throughout the paper. Notably, when they compare the passive products to CALIOP, implicit in that comparison is excellent viewing conditions. If there was say thin cirrus, there would not be a CALIOP retrieval to compare. For the intercomarison between retrievals without CALIOP, there clearly is a wider and higher distribution of heights. Notable are little "isolate" retrievals with very high height retrievals. The authors need to examine this data population closely. This may require a lot of hand analysis.

We aimed to capture all dust and smoke events across Asia from 2021 to 2023 during CALIOP overpasses. However, these occurrences were limited, spanning only a handful of days. We recognize the concern regarding the scarcity of data points. Throughout the paper, we have underscored our focus on "golden days" marked by ideal viewing conditions and the absence of cloud contamination.

A second concern is the reported diurnal cycle. Certainly we expect the diurnal cycle to be important to retrievals such as GEMS and EPIC that have widely varying scattering angles for retrieval physics if not reality, and thus can be important to include in a paper like this. Indeed,

all of the major AOD retrievals from geostationary have significant diurnal biases. However, the authors gloss through potential artifacts to explain diurnal differences, and are highly suggestive that the diurnal cycle from EPIC (their retrieval mind you) is real. However their explanation of the nature of diurnal aerosol height is flatly wrong-they suggest that the maximum aerosol height should be around solar noon due to PBL mixing. First, mixing will sustain itself through the afternoon and thus we would not expect a solar noon peak. Second, their product uses aerosol index and they are not as sensitive to the PBL anyways. Thus physically their explanation does not make sense. As EPIC stares at the sunny hemisphere of the earth, they may have an unaccounted for diurnal scattering angle issue in their RT code that causes symmetry along solar noon, or they may have a cosine response/resolution issue and perhaps also an associated resolution-based cloud mask bias. Now in full transparency, I am a co-author on a paper from this group that reported a strong diurnal height cycle of dust off of Africa. This said, as part of my contribution I pushed for softer language and examination of biases. As we see the exact same behavior here, I am thinking even more about diurnal bias in their EPIC retrievals. They have to come clean up front, and say they have no current way to evaluate which is correct, noting the possible artifacts, or remove the section all together. I strongly encourage the authors to look at some ground based lidar data in the region to verify their algorithm. There are plenty of Japanese and NASA lidars in this region that they can use.

Thank you for your concern. Firstly, it's important to clarify that our paper does not advocate for any specific product. Instead, we position ourselves as impartial users of several aerosol layer height products. While it's true that some of our co-authors contribute to the development of the EPIC product, we also have team members involved with GEMS and TROPOMI aerosol products. We have taken great care to ensure that our analysis remains objective and unbiased towards any particular product.

We acknowledge your valid concern regarding potential artifacts such as scattering angle issues for GEMS, however, EPIC focuses on backscattering with narrow varying scattering angles given its long distance to the Earth. Still, biases in cloud masking, particularly at lower resolutions near the edges, may influence the diurnal cycle of aerosol height. Additionally, the limited data points obtained from selected case study dates could introduce uncertainty when attempting to generalize aerosol height trends.

In this section, our intention was to illustrate the disparity in diurnal patterns between GEMS and EPIC, rather than definitively concluding that aerosol layer height peaks at local noon time. To reflect this, we have revised the language to adopt a more cautious tone, omitting the statement of a peak at local noon time and emphasizing the inherent difficulty in determining the accuracy of either dataset. Nonetheless, we believe this section remains valuable as it underscores the contrasting diurnal patterns observed between GEMS and EPIC, even if these differences may partially be attributed to measurement biases.

Additionally, we examined three years of MERRA-2 data spanning from 2021 to 2023, focusing on March within the East Asia domain. MERRA-2 AOCH is defined same as the CALIOP AOCH in the manuscript, weighted by optical depth at each vertical layer using aerosol extinction vertical profiles. Orange line indicates the average AOCH where AOCH is lower than the PBLH and the red represents the average of those PBLH values. Furthermore, AOCH calculated considering the aerosol extinction only below the PBLH (illustrated by the green line) show similar diurnal pattern with the PBLH. These similar diurnal variation between AOCH and PBLH indicates that when most aerosols locate within the PBL, diurnal

variation is affected by the PBL process, which changes the PBLH during daytime.

The diurnal variation of EPIC AOCH from Figure 8 is consistent with the MERRA-2 PBLH (grey line) and AOCH calculated by extinction below the PBL (green line), as ascending throughout the morning and descending after 2 pm local time, although EPIC AOCH values are higher than MERRA-2 AOCH, due to its constrain by PBLH. The diurnal variation of GEMS and MERRA-2 AOCH show similarities for the afternoon decrease. However, GEMS showing an overall decrease throughout the day, does not coincide with the MERRA-2 data, which shows in increase until 14 local time.
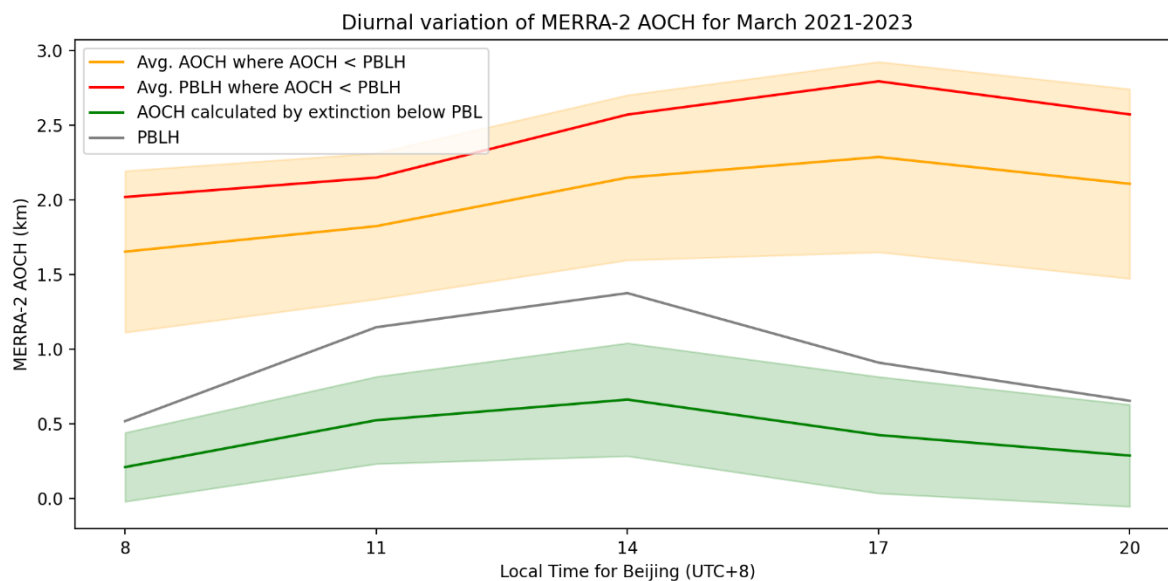
To further support our analysis of diurnal variations, we tried exploring data from MPLNET. However, only one MPLNET site falls within our Southeast Asia domain, it does not coincide with any of the dust or smoke plumes on our selected dates. We remain open to exploring additional ground-based lidar datasets for future studies.

In addition to these two science concerns, the paper has many minor copy editing typos, language irregularities, verb tense, dropping articles etc., Just an example, starting a paragraph on line 84 "The light travels longer path when aerosols locate at lower altitude than those at higher altitude, leading to more absorption from more O2 molecules in longer path (Ding et al., 2016; Xu et al., 2019)." The paper has sentences like this in almost every paragraph, and frequently drops articles (the/a/an). I suggest the authors utilize university copy editing programs before the paper can proceed.

We have addressed grammatical errors and improved sentence structures. Your feedback is appreciated. Thank you.

Specific notes

Abstract Line 15 (And Intro lineS 72-76).   I think from the beginning the authors should be a little more modest about what an aerosol layer height retrieval means.   It does indeed provide information, but it is only a single degree of freedom in what can be a complex aerosol vertical structure. From the beginning, Aerosol Layer Height (ALH throughout) can be a bit misleading, and largely based on a false presupposition-that there is one layer.   I am ok with language like "scale height", or "aerosol centroid", but honestly ALH has always bugged me as a labeled variable. Indeed, as noted in line 23 they have to make adjustments in definitions from the different product lines. I only ask they authors be mindful of this throughout the paper.

We appreciate this comment regarding the terminology of aerosol layer height. Given the diverse definitions and names of aerosol height products across platforms, we used the term "aerosol layer height" (ALH) to encompass all such products, including aerosol optical centroid height (AOCH), or aerosol effective height (AEH). Acknowledging the limitation of ALH in multiple aerosol layer structure, we revised some parts in the manuscript.

(Abstract) Aerosol vertical distribution is crucial for assessing surface air quality and the impact of aerosols on climate. Although aerosol vertical structures can be complex, assuming a shape of the aerosol vertical profile enables the retrieval of a single parameter, aerosol layer height (ALH), from passive remote sensing measurements.

(Introduction) Hence, many algorithms have been developed to extract a single piece of information regarding aerosol vertical distribution, with a primary emphasis on aerosol layer height, which approximates the altitude of aerosols from a presumed aerosol vertical profile shape.

Abstract line 24-25, you may want to mention how the products cross correlate.

We included the cross correlation of the products.

When comparing GEMS ALH to TROPOMI and EPIC, GEMS shows a narrower range and correlation coefficients are both below 0.4 (R < 0.4).

Abstract line 30 "EPIC and TROPOMI tend to overestimate AOD by 0.33 km and 0.23 km, respectively, in dust cases"   I am not sure where the km fits into this, as AOD is unitless.

Thank you for pointing this out. This was a mistake. We removed 'km', since AOD is unitless.

Introduction. Line 52 "Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) on board with the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) platform detects aerosol backscattering extinction profile with fine vertical resolution (Winker et al., 2013)."   This is a bit of unusual language.   What CALIOP measures is attenuated backscatter.   It retrieves aerosol extinction profiles, but these are retrievals and can have uncertainties of their own.

We revised the sentence.

Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), on board with the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) platform, detects

backscatter signal from different altitudes and retrieves aerosol extinction profiles with high vertical resolution (Winker et al., 2013)

Introduction. Line 55. "With the retirement of CALIPSO in August 2023, passive remote sensing will become the only routine technique from space for filling the data gap of measuring aerosol vertical distribution before next lidar dedicated to measure aerosols are launched into space." Technically this is not true, as the Chinese have an HSRL in space. But they don't release the data. We will see if they ever do.

Thanks for the information. We modified the sentence.

With the retirement of CALIPSO in August 2023, passive remote sensing will become the only routine <u>technique accessible for the public at present</u>, from space for filling the data gap of measuring aerosol vertical distribution before next lidar dedicated to measure aerosols are launched into space.

Methods, Section 2.3. I think throughout the paper, it needs to be emphasized that "golden days" are being used in this evaluation. I have no objection to this per say, as long as they make it clear in the abstract and introduction that these results are for ideal viewing conditions, and indeed day to day "mileage may vary considerably" Indeed, the very use of the CALIPSO as a verification dataset implies that they don't have to worry so much about things like cirrus or other cloud contamination.

We have emphasized this in the abstract and introduction.

(Abstract) All analyses are conducted for selected "golden days", which represent ideal viewing conditions for typical dust and smoke cases.

(Introduction) Considering that the spatial coverage of CALIOP is limited, we carefully selected "golden" cases where dust and smoke events favor the retrievals from all three sensors. This selection can maximize the signal to noise ratio for ALH retrieval, and hence, the evaluation can shed light on the future improvement to bring the closure of various types of retrievals. Note that these conditions may differ from those observed on non-selected days.

(2.3 Comparison Approach) Given the availability of EPIC/TROPOMI retrievals for absorbing aerosols, we focus our comparison on a selection of "golden days" characterized by ideal viewing conditions for dust and smoke cases, excluding cloud-covered areas, as observed within GEMS field of regard from 2021 to 2023.

Results

Line 295: I am not entirely sure how meaningful the stats are here "For dust cases, both TROPOMI and EPIC AOD exhibit a positive bias compared to AERONET AOD, with values of 0.23 and 0.33 for TROPOMI and EPIC, respectively." As these are for a distribution of AODs with different densities across AOD values. You can calculate bias as a function of AOD, or if it is linear enough a slope bias.

Thank you for your feedback. We assessed the difference in AOD between passive sensors with
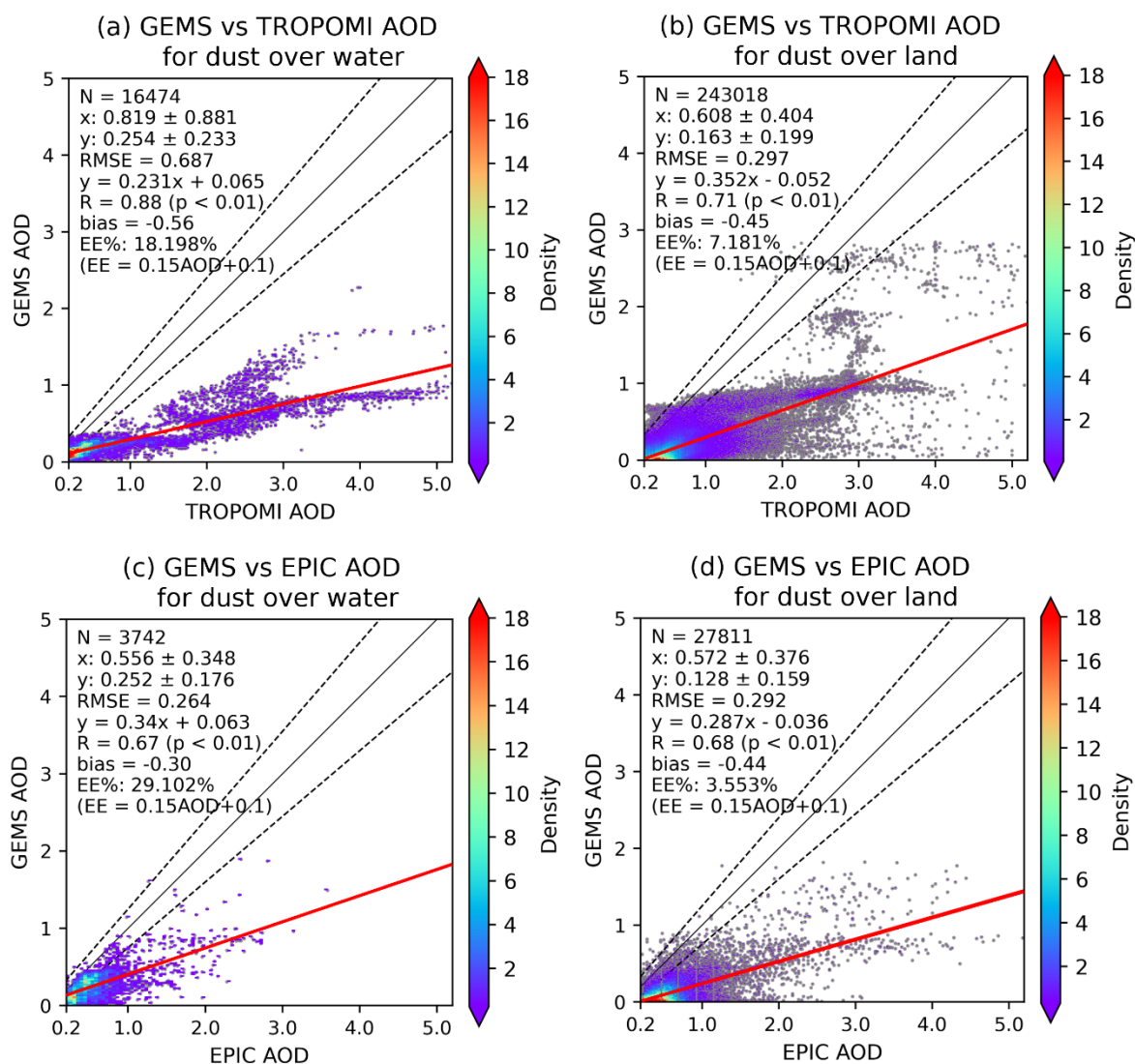
AERONET AOD as a function of different AOD bins. GEMS AOD consistently underestimate AERONET AOD, with the disparity increasing as AOD levels rise. EPIC and TROPOMI show positive bias, which is consistent with Fig. 3, but no significant trend in the difference is observed across different AOD bins. We included this figure below in the Supplementary (Fig. S1).
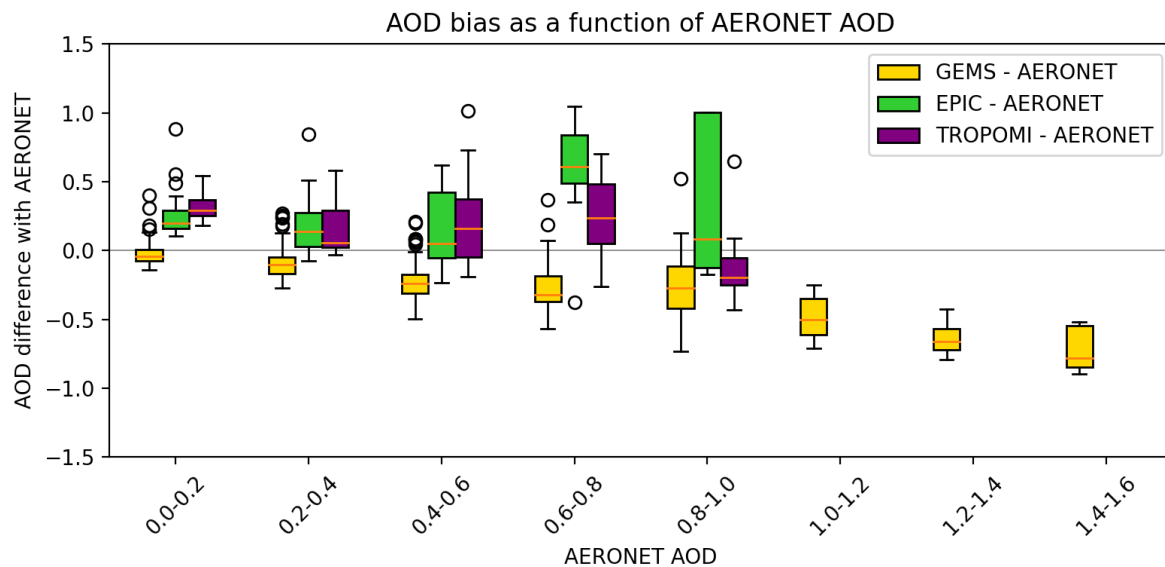
Line 320-323:   The authors should probably use more direct language, that based on Figure 3, the TROPOMI and EPIC dust AOD products were quite high biased (presumably as stated line 323 due to surface reflectance), and GEMS was low perhaps due to the optical model. Thus, when you cross compare, of course there is a massive bias between them. But can you also check this by looking only at data over water?

Thank you for the feedback. The figure below illustrates AOD comparisons distinguished over land and water. Comparison with AERONET data reveals that TROPOMI and EPIC AOD estimations for dust cases tend to overestimate. Specifically, in Figures (b) and (d), which compare GEMS to TROPOMI and EPIC AOD over land, we observe a similarity with AERONET comparisons: for GEMS AOD values close to zero, TROPOMI and EPIC AOD values are higher compared to GEMS.. However, this trend is not seen over water (Figures (a) and (c)). The discrepancy stem from differences in surface reflectance estimation in retrieval algorithms; GEMS employs the OMI surface reflectance climatology data product OMLER v003, whereas TROPOMI and EPIC utilize climatology MODIS surface products. This suggests that surface reflectance estimates from TROPOMI and EPIC over land may require refinement for more accurate retrievals. We have included this figure in the supplementary document.

Specifically, surface reflectance estimates for land surfaces from TROPOMI and EPIC may need refinement, as GEMS AOD values close to zero tend to be higher in TROPOMI and EPIC over land. However, this trend is not observed over water.

(a) GEMS vs TROPOMI AOD for dust over water

N = 16474
x: 0.819 ± 0.881
y: 0.254 ± 0.233
RMSE = 0.687
y = 0.231x + 0.065
R = 0.88 (p < 0.01)
bias = -0.56
EE%: 18.198%
(EE = 0.15AOD+0.1)

(b) GEMS vs TROPOMI AOD for dust over land

N = 243018
x: 0.608 ± 0.404
y: 0.163 ± 0.199
RMSE = 0.297
y = 0.352x - 0.052
R = 0.71 (p < 0.01)
bias = -0.45
EE%: 7.181%
(EE = 0.15AOD+0.1)

(c) GEMS vs EPIC AOD for dust over water

N = 3742
x: 0.556 ± 0.348
y: 0.252 ± 0.176
RMSE = 0.264
y = 0.34x + 0.063
R = 0.67 (p < 0.01)
bias = -0.30
EE%: 29.102%
(EE = 0.15AOD+0.1)

(d) GEMS vs EPIC AOD for dust over land

N = 27811
x: 0.572 ± 0.376
y: 0.128 ± 0.159
RMSE = 0.292
y = 0.287x - 0.036
R = 0.68 (p < 0.01)
bias = -0.44
EE%: 3.553%
(EE = 0.15AOD+0.1)

Line 345-352: AI can be a can of worms. There are a host of other issues than those noted, including pressure assumption differences, not just land reflectance but altitude models, resolution differences resulting in different cloud effects, etc. probably you need to calculate it yourself consistently.

We acknowledge the multitude of parameters that can impact UVAI, including those you mentioned. Given the variability in UVAI across different products, we adopted a criterion for classification in our paper based on setting UVAI thresholds to achieve similar spatial coverage. To achieve this, we used TROPOMI UVAI as a reference and set the UVAI threshold accordingly, rather than directly calculating UVAI ourselves. This approach ensures consistency in spatial coverage across the datasets under comparison. We have stated this in the manuscript.

Line 354-Figure 5. Why is there no TROPOMI vs EPIC plot? I think it would be good to cross characterize everything.

A figure comparing TROPOMI versus EPIC UVAI has been included in the Supplementary (Fig. S4).

Line 383-Figure 6. Why do the axis go much further than where there is data?   Maybe set to lesser spread and put the results in a table? Also, why not do all of the passive sensors for the different AI ranges?

We reduced the axis size and put the results in two tables. We analysis AOCH for different UVAI ranges only for GEMS since TROPOMI and EPIC AOCH products are already filtered with UVAI values larger than 1 and 3, for TROPOMI and EPIC, respectively, to focus on the retrieval of absorbing aerosols.

Line 408-Figure 7. It would be good to verify what is going on with AI cases bigger than 4, and what role clouds may have in increasing these values.   There are no CALIPSO verification cases with AOCH values over 4-5 km. I think it will be necessary to show that these cases where the retrieved AOCH are above such values are real, and not due to some artifact, such as thin cirrus above or unmasked low clouds below. In their case study (e.g, Figure 9) spot retrievals of very high AOCH are visible.   The authors should look in detail as to what is going on there.

We have examined the areas with high AOCH and found that the issue is a combination of inaccurate cloud detection and inherent sensitivity in the retrieval process. While some areas are influenced by cloud contamination, most of the high AOCH areas tend to have low AOD and not being influenced by clouds. Since AOCH is more sensitive to higher AOD (Xu et al., 2017), there is greater uncertainty in AOCH retrievals in regions with lower AOD. Furthermore, these high AOCH areas do not show high UVAI values greater than 4; instead, they exhibit values around 1-2, using GEMS UVAI as a reference. Therefore, both cloud detection inaccuracies and the low sensitivity of AOCH retrieval to low AOD contribute to the observed high AOCH in these areas, with the latter being more dominant in our selected cases.

Line 430-Figure 8 discussion.   The diurnal version of EPIC   looks pretty symmetrical around noon.   We have talked about this before, that this may be indicative of a retrieval bias. I don't think you can quickly dismiss this as being part of the PBL cycle, especially since the focus is on absorbing aerosol layers above the PBL.   Regardless, the effect of the PBL should maintain its height through the afternoon. The authors really need to incorporate a ground based lidar into these diurnal analyses.   In fact, they might consider dropping this section until that can be done.

We answered in one of the general comments above.

Line 485-Figure 11.   Please add a local solar time to the x-axis.

We added local time for Beijing (UTC + 8) to the x-axis.

Line 518-Figure 13(b).   The crosses for GEMS, TROPOMI and EPIC are very hard to read. The authors may want to change the color scheme and line thickness here and back on Figure 10 to make it consistent.

We changed the color scheme and line thickness. Thank you.

Authors' response to RC2

In this study, Kim et al compared aerosol layer heights products from several satellite instruments such as GEMS, EPIC and TROPOMI using retrieval algorithms all based on oxygen absorption bands. O2-A and B bands are used for TROPOMI and EPIC, O2-O2 band is used for GEMS. To have consistent comparisons, the aerosol layer heights are converted with a similar definition. Cases studies including dust and smoke over several regions in Asia are also discussed. Discrepancies between the products of the three instruments may reveal limitations in assumed aerosol and surface models and shed new lights to improve ALH retrievals. In general, this work fits the scope of AMT, and provide detailed and thorough analysis with useful results. I have several suggestive comments which may help improve the clarity of this work.

General comments:

1. There are many acronyms, and many of them are not defined when mentioned the first time. For example: UVAI (the ultraviolet aerosol index), AEH (aerosol effective height) have been used several times, but only defined in Page 5, Line 146 and 156. I would recommend having a table defining those acronyms.

   Thank you for pointing this out. We included a list of acronyms at the end of the paper (included as appendix) and verified all acronyms when introduced for the first time in the paper.

2. Most conclusions are made by comparing with AERONET and inter-comparisons. However, each product also has uncertainties which relate to the measurement uncertainties and retrieval algorithms. I didn't find discussions on the accuracy of the measurements from these instruments, and corresponding ALH uncertainties. In principle, the ALH uncertainty from each instrument can be validated using the AERONET data too.

   We included measurement and retrieval uncertainties investigated in the level 1 calibration and the algorithm papers of each sensor and product. Then, we compared these theoretical findings with our actual validation results with CALIOP.

   Based on a 2% measurement uncertainty for the EPIC DOAS ratios (Geogdzhayev and Marshak, 2018), the theoretical AOCH retrieval error is shown to remain below 1.25 km for vegetated surface when AOCH exceeds 1 km (Xu et al., 2019). Our analysis shows that the RMSE from all error sources, including measurement and retrieval uncertainties, evaluated between EPIC and CALIOP is approximately 1.25 km, which aligns with the retrieval error.

   The TROPOMI AOCH algorithm builds upon the framework established by the EPIC algorithm (Xu et al., 2019), with some adjustments for TROPOMI. Measurement uncertainty for TROPOMI is estimated to be 1 – 2 % (Kleipool et al., 2018). In addition to instrument errors, TROPOMI AOCH algorithm incorporates the convolution of TROPOMI spectral data, introducing potential additional uncertainty. Our study indicates an RMSE of TROPOMI ALH as 1.31 km. Assuming retrieval error similar to EPIC, this uncertainty appears reasonable.

   The GEMS AEH algorithm originates from Park et al. (2016), who performed an error analysis for OMI. The instrument error was indicated to be less than 10 m, stemming from a spectral wavelength error of 0.02 nm, with the total error ranging from 739 to 1276 m depending on aerosol types. Meanwhile GEMS has a spectral calibration accuracy of 0.002 nm (Kang et al., 2020). Our study demonstrates an RMSE of GEMS ALH at 0.75 km, falling within the theoretical retrieval error.

3.   Due to the information content, the ALH uncertainty should have strong dependency with the aerosol loading (or AOD). I don't have a clear understanding on what the AOD range is used in the discussion of ALH from this study, and how that impact the conclusion.
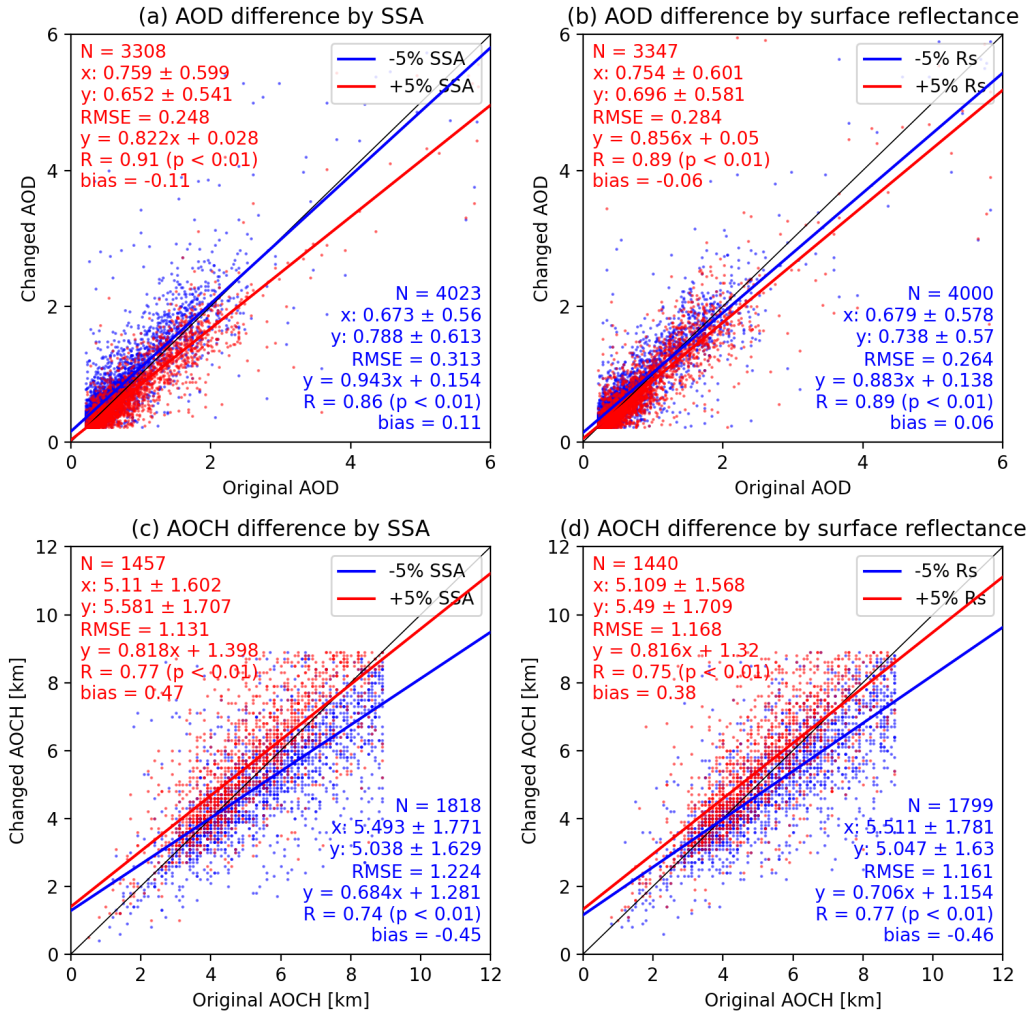
Thank you for your insightful comment. As AOD increases, the sensitivity of the reflectance ratio in the O2 A and B bands to AOCH also increases, as illustrated in Figure 1c-d from Xu et al. (2017). To further understand the AOD dependency of ALH uncertainty in the observational perspective, we analyzed the ALH difference between CALIOP and passive products, according to AOD bins of 0.2. The average difference between CALIOP AOCH and GEMS AOCH within each AOD bin only reaches 0.4 km. For TROPOMI, AOCH difference increase as increasing AOD, at AOD less than 1.2, with TROPOMI consistently overestimating, which resonates in Fig. 6. EPIC AOCH exhibits a linear increase in bias relative to AOD until AOD surpasses 1.2. This trend may stem from the AOD dependency within the aerosol model. We included this figure in the supplementary.
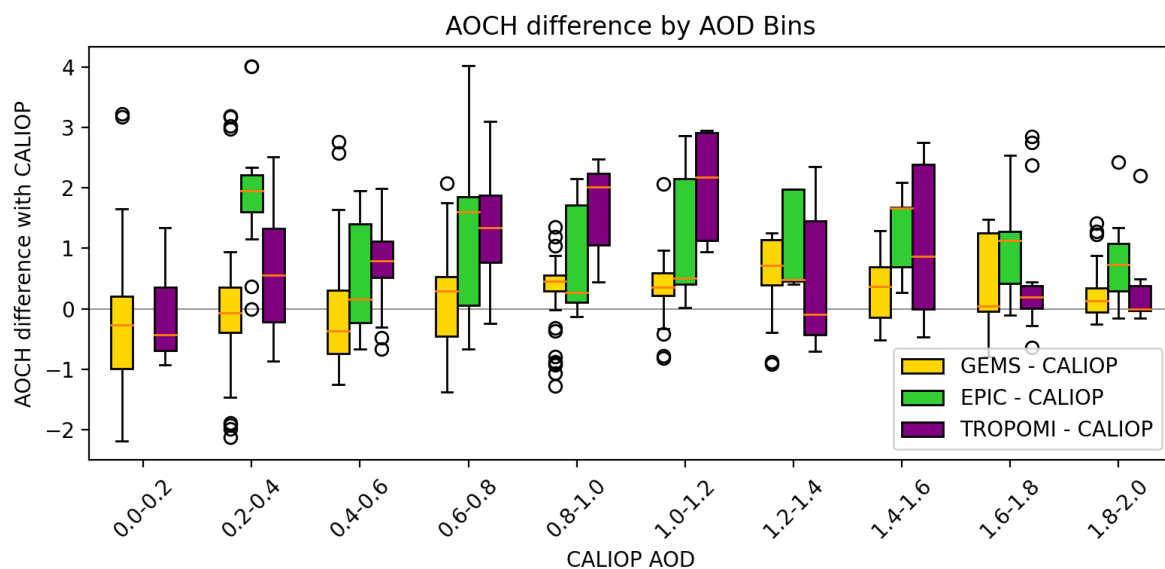
4.   The authors suggested that the aerosol and surface model used in the ALH retrievals may cause the discrepancy between different products. It would be useful to add more discussion on how such models impact the ALH retrievals, and how the authors would recommend to improve based on the results from this study.

We conducted a sensitivity test by modifying two key parameters: single scattering albedo (SSA) as a representative of the aerosol model, and surface reflectance as a representative of the surface model. Specifically, we increased and decreased both SSA and surface reflectance by 5% to evaluate their impact on ALH and AOD for one day of EPIC retrieval. This result in Figure below where increasing SSA and surface reflectance leads to lower AOD, but higher ALH, which is consistent with theoretical expectations. However, the complexity of the relationship between these parameters is evident, as not all data points show trends in the same direction. ALH retrieval becomes more complex due to the direct influence of SSA on ALH estimation, compounded by the indirect impact of SSA on AOD, which also affects ALH

retrieval. This underscores the complexity inherent in the relationship within each parameter, making it challenging to quantify their individual contributions.

(a) AOD difference by SSA

N = 3308
x: 0.759 ± 0.599
y: 0.652 ± 0.541
RMSE = 0.248
y = 0.822x + 0.028
R = 0.91 (p < 0.01)
bias = -0.11

-5% SSA
+5% SSA

N = 4023
x: 0.673 ± 0.56
y: 0.788 ± 0.613
RMSE = 0.313
y = 0.943x + 0.154
R = 0.86 (p < 0.01)
bias = 0.11

(b) AOD difference by surface reflectance

N = 3347
x: 0.754 ± 0.601
y: 0.696 ± 0.581
RMSE = 0.284
y = 0.856x + 0.05
R = 0.89 (p < 0.01)
bias = -0.06

-5% Rs
+5% Rs

N = 4000
x: 0.679 ± 0.578
y: 0.738 ± 0.57
RMSE = 0.264
y = 0.883x + 0.138
R = 0.89 (p < 0.01)
bias = 0.06

(c) AOCH difference by SSA

N = 1457
x: 5.11 ± 1.602
y: 5.581 ± 1.707
RMSE = 1.131
y = 0.818x + 1.398
R = 0.77 (p < 0.01)
bias = 0.47

-5% SSA
+5% SSA

N = 1818
x: 5.493 ± 1.771
y: 5.038 ± 1.629
RMSE = 1.224
y = 0.684x + 1.281
R = 0.74 (p < 0.01)
bias = -0.45

(d) AOCH difference by surface reflectance

N = 1440
x: 5.109 ± 1.568
y: 5.49 ± 1.709
RMSE = 1.168
y = 0.816x + 1.32
R = 0.75 (p < 0.01)
bias = 0.38

-5% Rs
+5% Rs

N = 1799
x: 5.511 ± 1.781
y: 5.047 ± 1.63
RMSE = 1.161
y = 0.706x + 1.154
R = 0.77 (p < 0.01)
bias = -0.46

Page 1, line 24 "In comparison with CALIOP ALH, both EPIC and TROPOMI ALH display a high correlation coefficient (R) higher than 0.7 and an overestimation by ~ 0.8 km, whereas GEMS ALH exhibits minimal bias (0.1 km) but a slightly lower correlation with R of 0.64."

Why there is larger bias with higher correlation? Does this indicate limitation to use correlation as a metric?

Using both correlation and bias as metric provides a more comprehensive evaluation of agreement between products, capturing both the strength of the relationship and the magnitude of differences. While GEMS has lower bias, EPIC and TROPOMI show higher correlation, indicating there may be a systematic bias that could be easily adjusted from their retrieval processes.

Page 1, line 25: UVAI not defined, what is its meaning?

Thank you for pointing this out. UVAI means ultraviolet aerosol index. We included a list of acronyms at the end of the paper (included as appendix) and verified all acronyms when introduced for the first time in the paper.

Page 2, Line 63: "…the degrees of freedom for signal (DOFS) increase from 2.1 to 2.8, which becomes sufficient for three parameter retrievals (AOD, aerosol peak height, and aerosol layer thickness)…"

How do you know that the three parameters are the right set of parameters, not other ones, such as SSA, aerosol size, etc?

This is a study from Choi et al. (2021) that focused on aerosol profiling capability. This information shows the impact of measurement factors, such as spectral resolution and coverage, SNR, radiance, and polarization on aerosol profile retrievals. The three key aerosol parameters needed to retrieve aerosol profiles are optical depth, peak height, layer thickness. While information content of other retrieval parameters including surface BRDF, aerosol microphysical properties (i.e., particle size distribution and refractive indices parameters) were also considered in their study, we focused on these three parameters that directly relate to aerosol distribution, which is the main focus of our study and Choi et al.'s work.

Choi et al. (2021) found that the degrees of freedom for signal (DOFS) for a single California Laboratory for Atmospheric Remote Sensing Fourier Transform Spectrometer (CLARS-FTS) measurement becomes sufficient to retrieve three key aerosol parameters—AOD, aerosol peak height, and aerosol layer thickness in the planetary boundary layer (PBL)—when adding a high spectral resolution (with a full-width half-maximum of 3 cm^{-1} or better), polarimetric measurements with SNR of at least 212, and radiance measurements with SNR of 300 for both oxygen (O_2) A and $^1\Delta$ bands.

Page 4. Line 115, although O2-A, O2-B and O2-O2 bands all have sensitivities, I didn't find any discussion on the measurement uncertainties from the three sensors using those bands?

As mentioned above, we included measurement and retrieval uncertainties and compared these theoretical findings with our validation results using CALIOP.

Page 4, line 125, "Accurate retrieval of ALH requires reliable retrieval of AOD, and past studies have shown that ALH and UVAI relationship can change with AOD (Xu et al., 2017)."

Can you elaborate how the relationship will change? And how did that apply to this study?

Thank you for your insightful questions. Since UVAI depends on ALH, AOD, and SSA, its correlation with ALH changes with AOD. From Figure 12 from Xu et al. (2019), the relationship between UVAI and ALH, as well as their correlation, strengthens with increasing AOD. However, our study focuses on comparing UVAI products to select thresholds for each AOCH product that has similar spatial coverage of absorbing aerosols, as well as establishing criteria (e.g., UVAI > 3 in Figure 6,7, and 8) for comparing different AOCH products. *Accurate retrieval of ALH requires reliable retrieval of AOD since the retrieval sensitivity is strongly dependent. Furthermore, ALH algorithms use UVAI to focus on absorbing aerosols or classify aerosol types. Since different sensors have their own UVAI products, we compare these products to select appropriate thresholds for comparing AOCH.*

Page 5, Line 135, "all algorithms assume quasi-Gaussian distribution described by two parameters including centroid height and half width (fixed at 1 km) at half maxima"

It would be useful to show the formula, which can help explain what is a quasi-Gaussian distribution, and half width at half maxima. I feel FWHM (full width at half maximum) is more commonly used. (I saw the formula in later section, you may need to add a reference).

We removed terms "quasi-Gaussian distribution" and "half width" since this part is just an introduction for upcoming sections. Instead, a more detailed explanation has been included in section 2.2.

Page 5, Line 156, "aerosol types are classified by the ultraviolet aerosol index (UVAI) and visible aerosol index derived from GEMS observations"

How UVAI is used to classify aerosol types?

UVAI and VisAI are calculated using the following equation:

$$\text{AI} = -100 \left[log \left(\frac{N_{\lambda_1}}{N_{\lambda_2}} \right)_{meas} - log \left(\frac{N_{\lambda_1}(LER_{\lambda_1})}{N_{\lambda_2}(LER_{\lambda_2})} \right)_{calc} \right]$$

In this equation, N_{λ_1} and N_{λ_2} represent the normalized radiances at the wavelength pairs 354/388 (477/490) nm for UVAI and VisAI, respectively. The subscripts "*meas*" and "*calc*" indicate the measured and calculated normalized radiances, respectively (Cho et al., 2024). We added the definition and meaning of UVAI in the introduction when it was first mentioned.

Three aerosol types were classified using UVAI and the Visible Aerosol Index (VisAI), which, similar to UVAI but with visible channels, categorizes aerosols into highly absorbing fine (HAF), dust, and non-absorbing (NA) aerosols. NA aerosols are selected when UVAI yields a negative value, the dust type is determined when both UVAI and VisAI are positive, and HAF is selected when UVAI is positive but VisAI is negative (Cho et al., 2023).

Page 5, Line 159. "For LUT generation, aerosols are assumed to be spherical and their particle size distribution, refractive index and fine mode fraction for each aerosol types are derived from global AERONET inversion climatology."

Can you confirm that whether AERONET aerosol inversion already considered non-spherical aerosols? I believe there are products used non-spherical aerosol model.

AERONET's aerosol inversion provides sphericity factor, but GEMS uses Mie theory for its LUT calculation. This limitation is because GEMS uses a computationally intensive spectral binning method, as noted by Cho et al. (2023). While spectral binning enhances stability by averaging across wavelengths and improves observation reliability by reducing random errors, it requires significant computational resources. To manage this, GEMS calculations use the Mie theory, which simplifies by ignoring the non-sphericity of dust.

For LUT generation, aerosols are assumed to be spherical due the computationally intensive spectral binning method, as noted by Cho et al. (2023). In addition, the particle size distribution, refractive index and fine mode fraction for each aerosol type are derived from global AERONET inversion climatology.

Page 6, Line 172, what O2AB-UI algorithm stands for?

The O2AB-UI algorithm was initially used to indicate EPIC and TROPOMI's AOCH algorithm. However, we realized use throughout the paper might lead to confusion, prompting us to remove this term entirely.

Page 6, Line 180, UVAI is defined in previous page.

Thank you for pointing this out. We defined all acronyms when introduced for the first time in the paper.

Page 6, Line 183, "only those pixels covered by lofted layer of absorbing aerosols with UVAI larger than 1.5 and AOD larger than 0.2 (at 680 nm) are analysed."

Is this the case for all following analysis? Fig 3, seems include AOD as small as 0.1 for all sensors.

This sentence applies for EPIC retrievals, and in Fig 3 (d) EPIC AOD does not have data below 0.2.

Page 6, Line 191, "However, the hyperspectral measurements from TROPOMI, unlike the EPIC measurement in narrow channels, prevent us to applying the EPIC AOCH algorithm in TROPOMI L1B data directly"

So what is the band width for TROPOMI?

TROPOMI consists of a high-resolution spectrometer system functioning across a range from ultraviolet to shortwave infrared. It incorporates seven distinct spectral bands: UV-1 (270-300nm), UV-2 (300-370nm), VIS (370-500nm), NIR-1 (685-710nm), NIR-2 (745-773nm),

SWIR-1 (1590-1675nm), and SWIR-3 (2305-2385nm). However, we removed this sentence as it was unnecessary.

Page 7, Line 210, Eq (1), what is the reason not choosing a Gaussian distribution but choose the current form? If a Gaussian distribution is used, FWHM or half width at half maximum can be easily expressed by the standard deviation.

Does the choice of 1km as half width at half maximum impact the ALH results?

Thank you for a valuable comment on the quasi-Gaussian distribution. As mentioned in a previous comment, while the full width at half maximum (FWHM) is more commonly used for Gaussian distribution, in the context of our study, half width is more often used as we are following the quasi-Gaussian distribution defined by Spurr and Christi (2014). Quasi-Gaussian distribution is a generalized distribution function (GDF) profile that offers easier integration in a simple closed form without requiring error functions.

The assumption of a 1 km half width is based on typical Lidar observations for dust and smoke aerosols, as noted by Reid et al. (2003). This same value has also been employed in the derivation of Aerosol Optical Depth (AOD) from ultraviolet (UV) observations by both TOMS (Total Ozone Mapping Spectrometer) and OMI (Ozone Monitoring Instrument), as highlighted in the work by Torres et al. (1998). Now, it is a commonly used parameter value as seen in products from EPIC, TROPOMI, and GEMS aerosol layer height retrievals.

Xu et al. (2019) illustrated the sensitivity of the DOAS ratio (ρ) to the half-width parameter (σH) in the Figure A1, showing that while the DOAS ratios exhibit a negative sensitivity to σH for aerosols at higher altitudes, the sensitivity becomes positive for ALH values below 1.5 km, and an error of 0.5 km in σH could result in a retrieval error of up to 0.3 km for ALH.

Page 8, Line 265 "Hence, the accuracy of each AOD product also influences corresponding ALH retrieval, which will be validated here by the ground-based Aerosol Robotic Network (AERONET) inversions as well."

Similar to a few previous comments, the aerosol loading itself also impacts ALH retrieval. One example can be found from polarimetric retrievals, such as Gao et al 2023, (https://doi.org/10.5194/amt-16-5863-2023). It would be useful to make it clear how AOD impacts the conclusion in this study.

We acknowledge that we should give more information on how AOD influences ALH retrieval. We included information about retrieval sensitivity of ALH to AOD as mentioned in a previous answer.

Page 8, Line 274 "Since TROPOMI and EPIC AOD products are retrieved at the wavelength of 680 nm whereas GEMS AOD products are retrieved at 354, 443, and, 550 nm,…"

Is there any estimation of the AOD accuracy derived from these bands? Which one is more accurate?

GEMS AOD is retrieved at 443 nm, with AOD at 380 nm and 550 nm subsequently calculated as part of their products. Consequently, there may be error propagation for GEMS AOD at 380 nm and 550 nm if the Angstrom exponent of aerosol types used to calculate AOD at different

wavelengths is inaccurate. In contrast, TROPOMI and EPIC AOD products are retrieved at 680 nm. We clarified that the GEMS AOD is retrieved at 443 nm.

Since TROPOMI and EPIC AOD products are retrieved at the wavelength of 680 nm whereas GEMS AOD products are retrieved at 443 nm, we estimated GEMS AOD at 680 nm from its AOD at 443nm ...

Page 8, Line 290 "The observed underestimation of GEMS AOD at 680 nm can be in part due to an overestimation of the Angstrom Exponent (AE), which can be affected from inaccurate particle size or refractive index in the wavelength-dependent aerosol model."

What aerosol model is used?

Details regarding the aerosol models employed in each algorithm are provided in Table S1 of the supplementary document

Table S1. Aerosol model comparison used in AOD/ALH retrievals for GEMS and TROPOMI/EPIC

		HAF	Dust
mr		1.46	$0.00428 \ln \tau + 1.55$ (675 nm)
mi		0.02044	$0.00197 \ln \tau + 0.00268$ (675 nm)
Reff	Fine mode	0.0854	$0.0152\tau + 0.122$
	Coarse mode	1.4115	-
Veff	Fine mode	1.5421	$0.156\tau + 0.227$
	Coarse mode	1.7630	-
fmf		0.99994	$-0.0696 \ln \tau + 0.37$
SSA			Coarse mode: $0.0214 \ln \tau + 0.949$ (675 nm)
Phase function		Mie	Fine: Mie Coarse: Dynamic AERONET climatology

τ is the AOD at 680 nm.

Page 11, Line 302, what is the surface model used here?

The surface estimations are included in the data section (2.1) as well as here.

(2.1 Satellite data) The surface reflectance data used for TROPOMI/EPIC ALH retrievals involves two sources: Land surface reflectance is obtained from the MODIS surface bi-

directional reflectance climatology, while water surface reflectance is derived from the GOME-2 surface Lambert-equivalent reflectivity (LER) database.

For TROPOMI and EPIC retrievals over land, climatological surface reflectance data from MODIS is employed. Additionally, unlike GEMS AEH retrieval algorithm that uses GEMS level 2 surface reflectance data, Cho et al. (2023) developed a new method for GEMS AOD retrieval, employing a novel hourly surface reflectance database generated through the minimum reflectance method, which integrates climatological minimum reflectance values for each pixel within a ±15-day window over a two-year period, along with monthly background AOD data. This novel surface reflectance estimation from GEMS AOD retrieval is shown effective.

Page 12, Fig 3. "Satellite data points only with a standard deviation less than 0.3 are shown for spatial consistency."

How is the standard deviation derived?

This is because we average satellite retrievals to collocate with AERONET data. In this process of averaging for a fixed radius around AERONET sites (here we use 0.2º) we might incorrectly average spatially inconsistent satellite retrievals. Also, data points with a standard deviation exceeding 0.3 constitute approximately 3% of the total data points; consequently, we regarded them as anomalies and removed them.

Page 13 Line 318, how the dust and smoke cases are separated?

Distinguishing between dust and smoke cases is based on the predominant aerosol type observed in the CALIOP lidar data.

(https://www-calipso.larc.nasa.gov/products/lidar/browse_images/std_v451_index.php)

Page 13, Line 321, what dust aerosol model is used, in terms of size, refractive index etc?

Aerosol models are presented in Table S1 of the supplementary document.

Page 14, Fig 4, it seems GEMS AOD has a boundary constraint which make its less than 0.5 most of the time, at least for dust? But it seems smoke case don't have such constraint.

The AERONET validation in Fig.3 provides a more comprehensive understanding of the apparent disparity for dust cases between GEMS and EPIC/TROPOMI. Specifically, when AERONET AOD is above one, GEMS AOD tends to be lower, whereas for AERONET AOD below 1, there are many instances where GEMS AOD exceeds the one-to-one line. This indicates that a higher disparity with EPIC/TROPOMI can be expected for higher GEMS AOD values, which may appear like a boundary constraint.

Page 16, 374-375, UVAI are used to categorize GEMS aerosol retrievals. What is the meaning for UVAI? Does different category relate to different aerosol types?

UVAI computes the difference between measured and calculated (as in purely molecular atmosphere) near-UV spectral dependence (Torres et al., 1998). UVAI values approaching zero occur when the atmosphere is devoid of aerosols, or when significant non-absorbing aerosol

particles and clouds are present. Aerosols capable of absorbing UV radiation, such as carbonaceous aerosols, volcanic ash, and desert dust, are the main factors leading to positive UVAI values (Torres et al., 2007). As mentioned previously, GEMS use UVAI and VisAI to classify aerosol types. TROPOMI and EPIC ALH focus solely on retrieving UV absorbing aerosols by setting a UVAI threshold, ensuring retrieval only occurs for pixels surpassing this threshold.

Page 19, Line 425, how the boundary layer height relates to the ALH? Is there any quantitative relationship?

Although we don't have results to show quantitative relationship with the boundary layer and ALH, here are some previous studies that showed the diurnal variation of aerosol layer height.

Xu et al. (2017) found that higher aerosol layer heights (ALH) were observed in the afternoon, possibly indicating a relationship with the diurnal evolution of tropospheric convection. Lee et al. (2019) observed that aerosol heights tend to rise in the afternoon and early evening, likely due to the development of the boundary layer's mixed layer. Lu et al. (2023) conjecture that the diurnal cycle of Saharan dust plume height is a consequence of the diurnal variation in solar heating, which leads to thermal buoyancy lifting the dust layer, combined with the diurnal evolution of the boundary layer.

Page 19, Line 443, are these UTC time? Can you also provide local time?

The time shown here is similar to local time, but constructed with the relative solar local noon times, which marks the moment when the solar zenith angle reaches its minimum value at a specific location. Using these times as the central reference point, GEMS and EPIC products for the day are adjusted to relative solar local time accordingly. This adjustment was considered necessary due to the wide geographical area covered by our study and the selected cases spanning from March to August. These months exhibit seasonal variations, resulting in notable changes in the sun's position.

Page 20, Fig 9, can you provide local time too?

Yes, we added local time for Beijing (UTC + 8) to the x-axis.

Page 25, Line 513, "…This suggests that EPIC and TROPOMI ALH retrievals exhibit a systematic positive bias for aerosols over Southeast Asia, indicating the potential need for tuning in the related smoke model, including surface reflectance and aerosol properties."

Can you elaborate what can be tuned in the smoke model?

"…This suggests that EPIC and TROPOMI ALH retrievals exhibit a systematic positive bias for aerosols over Southeast Asia, indicating the potential need for tuning in the related smoke model, including aerosol properties like size distribution, refractive index, single scattering albedo, and phase function."

Page 27, Line 554, "Both EPIC and TROPOMI consistently overestimates ALH in comparison to CALIOP, with an approximate bias of 0.8 km."

At what AOD range this conclusion is made?

All retrieved AOD from EPIC, and TROPOMI are considered. However, it is noted that EPIC and TROPOMI only retrieves AOD and ALH when AOD is higher than 0.2.