REVIEW OF THE MANUSCRIPT:
Impacts of direct CALIPSO Level 1B attenuated-backscatter assimilation on three-dimensional aerosol analyses and forecast over Northern China

GENERAL ASSESSMENT

This manuscript evaluates assimilating CALIOP/CALIPSO L1B attenuated-backscatter measurements directly into the online regional WRF-Chem model to improve 3D aerosol analyses and forecasts over northern China. Direct assimilation of spaceborne lidar observations helps transport models represent vertical aerosol distributions and supports future EarthCARE mission applications. Therefore, the work is very relevant and timely.

Assimilating L1B measurements from a spaceborne LiDAR rather than the retrieved L2 aerosol optical property profiles is the main contribution of this work. However, the authors must clarify its novelty and benefits relative to earlier work by the same group. Ye et al. (2021) assimilated CALIPSO aerosol-extinction vertical profiles (L2 data) into WRF-Chem using the MOSAIC aerosol scheme, while Zang et al. (2022) assimilated Himawari-8 AOD (L2 data) and used CALIPSO L2 profiles for evaluation. The manuscript should distinguish these approaches accurately. On page 6, line 18, the statement implies that both studies assimilated CALIPSO extinction profiles; if so, the authors should correct this description. Finally, the authors should distinguish the present December 2019 case studies from the earlier January and October 2019 cases.

The abstract reports that AOD@532nm correlations rose from 0.51 to 0.58, RMSE fell from 0.28 to 0.24, and mean bias improved from −0.11 to 0.01. While the mean bias reduction is substantial, the improvements in correlation (**R**) and RMSE are modest. The authors should show whether these gains are statistically robust, compare them with the L2 assimilation results of Ye et al. (2021), and specify which vertical aerosol layers improved. These comparisons are essential to demonstrate the added value of assimilating L1B data over existing methods.

Forecast improvements in surface PM2.5 concentrations persist for up to 14 hours, but this duration lacks context without a reference forecast and explicit performance criteria. The authors should define forecast skill, describe its evolution with lead time, and clarify whether this duration varies across cases.

Although the methodology is generally clear and well supported by references, concerns remain regarding measurement and representation error treatment, model–observation residual analysis detail, and unsupported conclusions.

Continuous line numbering would ease manuscript review if journal guidelines permit. Also, please cite published studies directly rather than using "researchers" as a subject to emphasize scientific evidence.

MAJOR COMMENTS

1. Novelty and comparison with previous 3D-Var assimilation studies

Directly assimilating L1B attenuated backscatter is the main advance of this work. State precisely which L2 product limitations this avoids and what new uncertainties it introduces. Directly comparing this approach with earlier WRF-Chem/MOSAIC studies (e.g., Ye et al., 2021; Zang et al., 2022) will establish its added value. Where direct comparison is impossible, explain differences in observations, periods, model settings, and evaluation methods.

2. Independent validation and daytime performance using AERONET

The assimilated CALIPSO observations appear to be from night-time overpasses. Please evaluate whether the improvements persist during the day by comparing modelled AOD with independent daytime AERONET observations at a suitable wavelength, using appropriate spatiotemporal matching. Describe any required spectral adjustment or interpolation. Present results for control and assimilation runs using identical validation samples to confirm whether night-time observations improve aerosol analyses beyond the assimilation window.

3. Measurement uncertainties and PM2.5 evaluation

Please describe the PM2.5 monitoring network, instruments, quality control, temporal averaging, and known uncertainties. Model–observation comparisons must account for measurement uncertainty and, where possible, model and representation

errors—information essential for evaluating the changes in Fig. 11. Studies like van Donkelaar et al. (2013, 2016) offer useful guidance on uncertainty assessment, despite differing in methods and scale.

4. Use of CALIOP measurement channels and optical constraints

CALIOP measures backscatter at 532 and 1064 nm, with parallel- and perpendicular-polarisation channels at 532 nm. The manuscript appears to use only the 532 nm channels. Please explain why the 1064 nm measurements were excluded. Could the additional wavelength and depolarisation information help constrain aerosol type, particle size or cloud screening? Such information cannot be treated as a direct column Ångström exponent without a suitable retrieval or observation operator, but its potential value merits discussion. Please also explain how the chosen measurement channels are combined to form total attenuated backscatter.

5. Statistical significance, vertical residuals and forecast skill

The manuscript should present model-minus-observation residuals along the full CALIPSO tracks, before and after assimilation. A consistent colour scale and observation-matched sampling would make changes in the vertical structure easier to assess. Please include uncertainty or significance information where feasible. For AOD and PM2.5, show how forecast skill evolves with lead time and define the criterion used to identify a 9–14-hour benefit. These analyses would provide firmer support for the conclusions about three-dimensional aerosol improvement and forecast performance.

SPECIFIC COMMENTS

The page and line numbers below refer to the manuscript as reviewed.

1. Page 2, lines 2–3. The introduction could make a stronger distinction between passive and active remote sensing. Passive observations often constrain total-column aerosol properties, whereas lidar provides vertically resolved information. Please explain why this vertical information matters for data assimilation.

2. Page 2, line 7. Identify the total attenuated-backscatter observations explicitly as CALIPSO L1B measurements.

3. Page 2, line 8. Please state in the abstract that the study uses the online WRF-Chem model. This will help distinguish the approach from offline chemistry–transport modelling discussed later.

4. Page 2, lines 13–14. Please clarify that the evaluation does not address the full spectral dependence of aerosol optical properties.

5. Page 4, lines 2–3. Please consider referring to the “three-dimensional distribution and temporal evolution of aerosols” rather than only their “vertical representation”.

6. Page 4, lines 7–12. Please provide relevant references to early aerosol modelling and inversion studies at global and regional scales. Possible examples may include Henze et al. (2007), Dubovik et al. (2008), Vautard et al. (2000) and Menut et al. (2000).

7. Page 5, line 1. Please define the Gridpoint Statistical Interpolation (GSI) system briefly and cite its relevant documentation. The phrase “for meteorological applications” is broad. Therefore, please explain more precisely how the system is being extended or applied to aerosol data assimilation.

8. Page 5, line 7. Please provide references for the statement made here.

9. Page 5, lines 23–24. Please explain the nonlinear observation operator; and state what it converts, why it is nonlinear, and how it relates the model aerosol state to the observed lidar signal with suitable references.

10. Page 5, line 27. Please clarify whether the early study cited assimilated single-wavelength satellite AOD. Optical properties are wavelength dependent, and this distinction matters. Collins et al. (2001) may be relevant if it supports the historical claim.

11. Page 6, line 1. Please specify the wavelength of the MODIS AOD observations. Briefly describe the reported improvements and cite the relevant studies.

12. Page 6, line 8. A short discussion of high-altitude aerosol layers would strengthen the motivation for active remote sensing. Volcanic eruptions and intense wildfires can inject aerosols into the upper troposphere and stratosphere. Lidar observations can help constrain their height and vertical evolution. Please keep the discussion relevant to the present regional study.

13. Page 6, lines 24–25. Please moderate the description of L2 products. Their uncertainties arise from measurement noise, calibration, retrieval assumptions and limitations of the forward model. Rather than suggesting that these products contain avoidable or arbitrary artefacts, identify the specific limitations that L1B assimilation is intended to reduce.

14. Page 7, lines 1–5. Several broad statements in this paragraph need supporting references. Please cite the relevant observational and modelling studies. Introduce the term “active remote sensing” earlier in the introduction, where its importance first becomes clear.

15. Page 7, line 7. Please explain what is meant by cloud and surface contamination in the CALIOP measurements. Also describe how these effects are identified or treated. This could be linked to the discussion of the nonlinear observation operator.

16. Page 7, lines 13–14. This sentence repeats an earlier point and could be removed.

17. Page 8, line 6. Please consider replacing “accurately simulating aerosol pollution” with “simulating aerosol concentrations and air quality”. Also acknowledge the limitations reported in the cited WRF-Chem studies, rather than implying that accurate simulation is assured.

18. Page 8, line 7. Please identify the eight aerosol species and explain how they are represented in the four size bins. Give the relevant size ranges and references for the MOSAIC configuration.

19. Page 8, lines 20–21. Please cite the National Meteorological Center (NMC) method and explain why 24- and 48-hour forecasts are selected to estimate background-error statistics.

20. Page 9, line 2. The background-error covariance matrix **B** is described, but the construction of the observation-error covariance matrix R is less clear. Please explain briefly how **R** matrix is specified and whether it includes measurement and representation errors.

21. Page 9, lines 6–8. Please add a short description of the minimisation procedure. How is the Jacobian or gradient evaluated? Is the cost function minimised iteratively, or is a linearised system solved directly? A few sentences would make the method easier to reproduce.

22. Page 9, line 9. Please define what is “minimum solution”. Does this refer to the minimiser of the least-squares cost function or to a particular regularised solution?

23. Page 9, lines 20–30, and page 10, lines 1–13. Please expand the description of aerosol optical properties. Explain how relative humidity affects particle size through hygroscopic growth, how wet radii are calculated, and which refractive indices are assigned to each aerosol species. State whether the complex refractive indices are prescribed from the literature. Although the present application centres on 532 nm, the assumptions governing extinction and absorption should be clear. Fast et al. (2006) is a useful starting reference.

24. Page 12, line 7. Please distinguish between the laser source and the lidar instrument. A laser generates the transmitted light; the lidar system also includes the receiver and measurement electronics. Clarify the intended meaning in this sentence.

25. Page 12, line 26. Please replace the broad phrase “retrieval artefacts” with a more precise account of limitations in the L2 retrievals. Also, state which of these limitations direct L1B assimilation may avoid, and which remain.

26. Page 13, line 3. Please explain how the two polarisation channels at 532 nm are used to obtain total attenuated backscatter. Also, add the appropriate product references.

27. Page 13, line 5. Please clarify the meaning of an altitude of −2.0 km, including the altitude reference and its role in the data-processing procedure.

28. Page 13, lines 5–10. Please describe the L1B preprocessing steps more explicitly. In particular, define “delayed recovery” and explain how affected measurements are identified and handled.

29. Page 13, lines 12–15. The focus on aerosol pollution over northern China gives a reasonable basis for prioritising measurements below 5 km. However, avoid implying that measurements above 30 km are always noise. High-altitude aerosol signals can be real, as shown by the Raikoke and Ulawun eruptions in 2019 and the Hunga Tonga–Hunga Ha‘apai eruption in 2022. Therefore, please state the altitude limits as choices specific to this study and explain the screening criteria.

30. Page 13, line 16. Please justify the assumed 20% measurement uncertainty. Winker et al. (2007) discuss instrument performance, including discrepancies at high altitude. Explain how the quoted uncertainty is derived for the 0–5 km layer used here, and whether it varies with altitude or signal strength.

31. Page 14, lines 6–8. Please explain how data thinning affects the observation-error treatment. Are representation errors included in R or otherwise accounted for in the cost function J(x)? This matters when applying the method at a different model resolution, such as 1 km.

32. Page 14, lines 9–10. Please rewrite the sentence to use consistent variables and units. At present, a threshold expressed in km-1 is compared with a quantity expressed in metres, which makes the intended criterion unclear.

33. Page 14, lines 11–13. Please clarify the role of the surface return and surface type in the lidar measurements. If the statement concerns bright land surfaces, specify the relevant measurement effect and avoid generalising across all surfaces in the study area.

34. Figure 1, page 36. The green circles are identified as surface PM monitoring sites, but the triangles are not defined. Please explain all symbols in the caption or legend.

35. Page 15, lines 16–20. Please describe the reanalysis data used to initialise and constrain the simulations. Include their spatial and temporal resolutions, the lateral-boundary conditions, and any nudging configuration. Also, briefly discuss how these choices may influence the results.

36. Page 16, lines 5–6; Figs. 2–4. Please consider combining the key information into a two-row, three-column figure. For each case, show observed CALIPSO backscatter, control-minus-observation residuals, and assimilation-minus-observation residuals. Use matching scales and sampling. This would make changes in vertical and horizontal structure easier to see and may reduce the need for Fig. 4. Add missing colour-bar units and explain how missing data are displayed; white areas should not be left ambiguous.

37. Page 17, line 12. Please explain the importance of the lowest 100 m for air-quality applications and why the lidar measurements may be limited there. Also, discuss whether near-surface information can be recovered or supplemented by other observations. A brief outlook on the relevance to EarthCARE mission would be useful, provided its measurement capabilities and limitations are represented accurately.

38. Page 17, lines 13–14. The underestimation of attenuated backscatter could arise from aerosol mass concentrations, vertical transport or the optical-property operator. Can the authors examine the changes in aerosol mass-mixing-ratio profiles, by species, between the control and assimilation runs? This would help separate model-state errors from optical assumptions.

39. Page 18, lines 5–8; Fig. 6. Please plot model-minus-observation residuals before and after assimilation at locations and altitudes matched to the assimilated measurements. Also, explain the enhanced modelled signal near 2.2–4.4 km and approximately 120.4–121.4° E if it has no clear counterpart in the observations. A direct residual analysis would support, or qualify, the claim of agreement.

40. Page 18, lines 16–19. The text mentions excluding extreme attenuated-backscatter values, but this filter is not described in Sect. 2.3. Please state the threshold, rationale and number or fraction of observations rejected.

41. Page 18, lines 20–29; Fig. 7. The three selected profiles do not fully characterise performance along the two CALIPSO tracks. Please consider adding full-track residual cross-sections for the control and assimilation runs. They would show where the analysis improves or deteriorates.

42. Page 19, line 2; Fig. 8. Please show absolute, and possibly relative, model–observation residuals before and after assimilation. State the residual convention and use consistent colour scales.

43. Page 19, lines 6–10; Figs. 8 and 11. The assertion that the control run “seriously” underestimates PM2.5 needs support from an uncertainty-aware comparison. Show the bias and its variability, together with uncertainty estimates appropriate to the actual monitoring data and model setup. Uncertainty estimates reported in van Donkelaar et al. (2013) should not be transferred directly to this WRF-Chem experiment without justification.

44. Page 19, lines 28–29. Please discuss uncertainty in both observed and simulated backscatter, including the optical assumptions used to calculate modelled signals. Omar et al. (2010) and Papagiannopoulos et al. (2016) may provide useful context. Full-track residual profiles could help distinguish errors near the surface from those aloft. Possible causes at higher altitudes include transport, aerosol formation and emission errors, but these should be presented as hypotheses unless tested.

45. Page 20, lines 5–8. The text acknowledges that retrieved extinction coefficients need further validation but does not explain what can be concluded from the present study. Please complete the discussion. Also, state the available evidence, its limitations and the validation needed in future work.

46. Page 20, lines 10–13; Figs. 9 and 10. Please consider a more compact presentation of the total-column AOD comparison, including the control and assimilation statistics. Figure 10 suggests improvement, but the statement that assimilation “accurately corrects” AOD is too strong without an explicit accuracy benchmark. Assess performance against a suitable reference or uncertainty requirement. The Global Climate Observing System (GCOS) criteria discussed by Popp et al. (2016) may be useful here.

47. Page 21, line 28. Please explain the reported 9–14-hour duration of forecast improvement. Define the skill threshold, show its evolution with forecast lead time, and discuss what limits the benefit. Could changes in observation coverage, assimilation frequency or model errors extend it?

48. Page 22, lines 6–25; Sect. 3.4. The discussion of PM2.5 is relevant, but the observational data and their uncertainties need fuller documentation. If aerosol–radiation feedback is enabled, changes in aerosol concentrations may affect heating rates and meteorological fields. Please state whether this feedback is active and, if so, show separate changes or residuals for 10 m winds and PM2.5. Agarwal et al. (2024) may offer relevant context, although it does not by itself establish the mechanism in these simulations.

49. Page 23, lines 5–6. The discussion of Mie scattering theory raises questions about particle shape and depolarisation. Does the observation operator use the 532 nm depolarisation information? If so, show how it enters the calculation. If not, discuss the limitations of the assumed particle shape, particularly for non-spherical aerosols, and how these may affect AOD and backscatter.

50. Page 23, lines 22–24. The claim that information from vertically resolved lidar observations is transferred to the model aerosol state needs clearer evidence. Please show full-track residuals or corresponding changes in aerosol-state profiles, rather than relying on a few selected examples.

51. Page 24, lines 16–18. Changes in meteorological fields are discussed but not shown. Either provide before-and-after comparisons with independent observations, or moderate the claim.

52. Page 24, lines 18–20. Please explain the evidence for the proposed cold-air circulation mechanism. Temperature and wind fields before and after assimilation, together with an appropriate reference, would help establish whether this mechanism is supported. Please avoid attributing the aerosol changes to circulation without a direct analysis.

53. Page 25, lines 3–14. This paragraph offers an important discussion of L1B data limitations. Move the essential points to the data and methods section so readers understand these limitations before interpreting the results. Return to them in the discussion when assessing uncertainty and the strength of the conclusions.

REFERENCES CITED IN THE REVIEW

1. Ye, H., Pan, X., You, W., Zhu, X., Zang, Z., Wang, D., ... & Jin, S. (2021). Impact of CALIPSO profile data assimilation on 3-D aerosol improvement in a size-resolved aerosol model. *Atmospheric Research*, *264*, 105877.
2. Zang, Z., You, W., Ye, H., Liang, Y., Li, Y., Wang, D., Hu, Y., & Yan, P. (2022). 3DVAR Aerosol Data Assimilation and Evaluation Using Surface PM2.5, Himawari-8 AOD and CALIPSO Profile Observations in the North China. Remote Sensing, 14(16), 4009. https://doi.org/10.3390/rs14164009
3. Henze, D. K., Hakami, A., & Seinfeld, J. H. (2007). Development of the adjoint of GEOS-Chem. *Atmospheric Chemistry and Physics*, *7*(9), 2413-2433.
4. Dubovik, O., Lapyonok, T., Kaufman, Y. J., Chin, M., Ginoux, P., Kahn, R. A., & Sinyuk, A. (2008). Retrieving global aerosol sources from satellites using inverse modeling. *Atmospheric Chemistry and Physics*, *8*(2), 209-250.
5. Vautard, R., Beekmann, M., & Menut, L. (2000). Applications of adjoint modelling in atmospheric chemistry: sensitivity and inverse modelling. *Environmental Modelling & Software*, *15*(6-7), 703-709.
6. Menut, L., Vautard, R., Beekmann, M., & Honoré, C. (2000). Sensitivity of photochemical pollution using the adjoint of a simplified chemistry-transport model. *Journal of Geophysical Research: Atmospheres*, *105*(D12), 15379-15402.
7. Winker, D. M., W. H. Hunt, and M. J. McGill (2007), Initial performance assessment of CALIOP, *Geophys. Res. Lett.*, 34, L19803, doi:10.1029/2007GL030135.
8. Van Donkelaar, A., Martin, R. V., Spurr, R. J., Drury, E., Remer, L. A., Levy, R. C., & Wang, J. (2013). Optimal estimation for global ground-level fine particulate matter concentrations. *Journal of Geophysical Research: Atmospheres*, *118*(11), 5621-5636.
9. Van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., ... & Winker, D. M. (2016). Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental science & technology*, *50*(7), 3762-3772.
10. Omar, A., et al. (2010), *Extinction-to-backscatter ratios of Saharan dust layers derived from in situ measurements and CALIPSO overflights during NAMMA*, *J. Geophys. Res.*, 115, D24217, doi:10.1029/2010JD014223.
11. Papagiannopoulos, N., Mona, L., Alados-Arboledas, L., Amiridis, V., Baars, H., Binietoglou, I., Bortoli, D., D'Amico, G., Giunta, A., Guerrero-Rascado, J. L., Schwarz, A., Pereira, S., Spinelli, N., Wandinger, U., Wang, X., and Pappalardo, G.: CALIPSO climatological products: evaluation and suggestions from EARLINET, Atmos. Chem. Phys., 16, 2341–2357, https://doi.org/10.5194/acp-16-2341-2016, 2016
12. Popp, T., De Leeuw, G., Bingen, C., Brühl, C., Capelle, V., Chedin, A., ... & Xue, Y. (2016). Development, production and evaluation of aerosol climate data records from European satellite observations (Aerosol_cci). *Remote Sensing*, *8*(5), 421.
13. Agarwal, P., Stevenson, D. S., and Heal, M. R.: Evaluation of WRF-Chem-simulated meteorology and aerosols over northern India during the severe pollution episode of 2016, Atmos. Chem. Phys., 24, 2239–2266, https://doi.org/10.5194/acp-24-2239-2024, 2024