Authors' responses to Referees' comments

Journal: Atmospheric Measurement Techniques

Manuscript Number: egusphere-2025-4237

Title: A Physics-Constrained Deep-Learning Framework based on Long-Term Remote-Sensing Data for Retrieving Vertical Distribution of $PM_{2.5}$ Chemical Components

Authors: Hongyi Li, Ting Yang, et al.

Note:

Comment (12-point black italicized font).

Reply (indented, 12-point blue normal font).

"Revised text as it appears in the text (in quotes, 12-point blue italicized font)".

Anonymous Referee #1

1 General comments:

"A Physics-Constrained Deep-Learning Framework based on Long-Term Remote-Sensing Data for Retrieving Vertical Distribution of $PM_{2.5}$ Chemical Components" by Li et al., proposes a novel method to retrieve concentrations of major aerosol components (sulfate, nitrate, ammonium, organic matter, and black carbon) from ground-based lidar extinction data combined with ERA5 meteorological reanalysis. The authors validate their retrieved concentrations against surface, airborne, and tower observations in the greater Beijing region, reporting generally high correlations. ***Given the importance of aerosols in Earth's radiative balance and air quality, developing methods that leverage lidar's high vertical resolution to determine constituent species concentrations is a valuable endeavor.*** *However, the manuscript in its current form requires significant revisions to adequately describe the methodology and contextualize the results.*

Authors' response:

We sincerely thank the Reviewer for the thoughtful assessment of our manuscript and for recognizing the potential value of our work. In response to the Reviewer's comments, we have undertaken a comprehensive revision of the manuscript.

2 Major Comments:

1) A fundamental issue is how $PM_{2.5}$ concentrations are distinguished from larger particles using lidar extinction data. The lidar dataset presumably provides total aerosol extinction from particles of all sizes, yet the work presented here centers only on $PM_{2.5}$. The authors provide no explanation of how contributions from larger particles (PM_{10}, coarse mode, etc.) are excluded from the extinction signal (if they are). This represents a potentially significant source of error, particularly during dust events when coarse particles may dominate extinction.

Authors' response:

We would like to thank the reviewer for the insightful suggestion. We fully concur that the identification of $PM_{2.5}$ from the total aerosol extinction signal influenced by coarse-mode particles presents a potential source of error. We tackled this issue by employing end-to-end machine learning training that utilizes long-term lidar signals and observations of $PM_{2.5}$ chemical compositions. The detailed explanations are as follows:

a. **The total aerosol extinction from particles of all sizes can serve as a direct indicator for retrieving $PM_{2.5}$**. Earlier studies have revealed a strong correlation between $PM_{2.5}$ and total light attenuation from particles of all sizes, such as aerosol optical depth (AOD) (van Donkelaar et al., 2010; Zhang et al., 2009) and aerosol extinction coefficient (EXT) (Lowenthal and Kumar, 2016; Tao et al. 2012). Therefore, the total light attenuation has been widely served as a crucial indicator for directly predicting $PM_{2.5}$ concentrations in machine learning models (Table 1).

Table 1. Literature review of retrieving $PM_{2.5}$ from the total light attenuation based on machine learning algorithms.

Target feature	Optical input feature	Machine learning algorithm	Citation
Ground-level $PM_{2.5}$	AOD at 550nm	CNN	Park et al., 2020
Ground-level $PM_{2.5}$	TOA reflectance at 460nm, 640nm and 2300nm	Geoi-LSTM	Wang et al., 2021
Ground-level $PM_{2.5}$	AOD at 470nm	XGBoost	Gutiérrez-Avila et al., 2022

$PM_{2.5}$ vertical profile	AOD at 532nm	ET (best)	Chen et al., 2022
$PM_{2.5}$ vertical profile	EXT at 580nm and 590nm	CNN-BiLSTM (best)	Yi et al., 2025
$PM_{2.5}$ vertical profile	EXT at 532nm	CNN-BiLSTM	Wang et al., 2025

(CNN: Convolutional neural network; TOA: Top of atmosphere; Geoi-LSTM: Geo-intelligent Long Short-Term Memory; XGBoost: eXtreme Gradient Boosting; CALIOP: Cloud-Aerosol Lidar with Orthogonal Polarization; ET: Extra Trees; BiLSTM: Bidirectional LSTM)

b. **Machine learning methods can directly identify fine-mode particles from lidar extinction signals through end-to-end learning**. The retrieval framework in our work establishes a nonlinear mapping relationship between the extinction coefficient at 532nm and the chemical compositions of $PM_{2.5}$ to identify fine-mode particles from lidar extinction data. This end-to-end learning prevents the filtering out of lidar extinction signals induced by larger particles (such as PM_{10}), since the learning process specifically captures the signal patterns associated with variations in the chemical compositions of $PM_{2.5}$, which were the data fed into the learning process.

c. **Meteorological reanalysis data is used as auxiliary training data to strengthen the EXT-$PM_{2.5}$ relationship**. As the reviewer noted, coarse particles may predominate in extinction during dust events, making the identification of fine particles based solely on the extinction coefficient inaccurate. Following the previous studies (Lee et al., 2011; Xie et al., 2015), we integrated meteorological conditions (such as temperature, relative humidity, wind conditions and vertical velocity) into the machine-learning training process to better capture the spatiotemporal variations in the EXT-$PM_{2.5}$ relationship.

d. **Long-term training datasets allow machine learning models to learn the EXT-$PM_{2.5}$ relationship across various aerosol mode scenarios**. We utilized a 18-month dataset to establish a nonlinear mapping relationship between lidar signals and $PM_{2.5}$ chemical compositions. The training dataset encompasses various seasons and multiple aerosol mode scenarios, including strong dust events reported in Beijing during March 2021 (Gui et al., 2022). This diversity ensures that our machine learning model encounters a range of aerosol mode distributions and learns to make accurate predictions in the presence of coarse particles.

Reference

Chen, B., Song, Z., Pan, F., et al.: Obtaining vertical distribution of $PM_{2.5}$ from CALIOP data and machine learning algorithms, Sci. Total Environ., 805, 150338, https://doi.org/10.1016/j.scitotenv.2021.150338, 2022.

Gui, K., Yao, W., Che, H., et al.: Record-breaking dust loading during two mega dust storm events over northern China in March 2021: aerosol optical and radiative properties and meteorological drivers, Atmos. Chem. Phys., 22, 7905-7932, https://doi.org/10.5194/acp-22-7905-2022, 2022.

Gutiérrez-Avila, I., Arfer, K.B., Carrión, D. et al.: Prediction of daily mean and one-hour maximum $PM_{2.5}$ concentrations and applications in Central Mexico using satellite-based machine-learning models, J. Expo. Sci. Environ. Epidemiol., 32, 917-925, https://doi.org/10.1038/s41370-022-00471-4, 2022.

Lee, H. J., Liu, Y., Coull, B. A., Schwartz, J., and Koutrakis, P.: A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations, Atmos. Chem. Phys., 11, 7991-8002, https://doi.org/10.5194/acp-11-7991-2011, 2011.

Lowenthal, D. H., & Kumar, N.: Evaluation of the IMPROVE Equation for estimating aerosol light extinction, J. Air Waste Manage., 66(7), 726-737, https://doi.org/10.1080/10962247.2016.1178187, 2016.

Tao, J., Cao, JJ., Zhang, RJ. et al.: Reconstructed light extinction coefficients using chemical compositions of $PM_{2.5}$ in winter in Urban Guangzhou, China, Adv. Atmos. Sci. 29, 359-368, https://doi.org/10.1007/s00376-011-1045-0, 2012.

van Donkelaar, A., Martin, R., Brauer, M., et al.: Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application, Environ. Health Perspect., 118(6), 847-855 https://doi.org/10.1289/ehp.0901623, 2010.

Wang, B., Yuan, Q., Yang, Q., et al.: Estimate hourly $PM_{2.5}$ concentrations from Himawari-8 TOA reflectance directly using geo-intelligent long short-term memory network, Environ. Pollut., 271, 116327, https://doi.org/10.1016/j.envpol.2020.116327, 2021.

Xie, Y., Wang, Y., Zhang, K., et al.: Daily Estimation of Ground-Level $PM_{2.5}$ Concentrations over Beijing Using 3 km Resolution MODIS AOD, Environ. Sci. Technol., 49, 20, 12280-12288, https://doi.org/10.1021/acs.est.5b01413, 2015.

Yi, Z., Xiang, Y., Yun L., et al.: Deep learning-driven reconstruction of $PM_{2.5}$ vertical profiles: A fusion of lidar and tower data, J. Clean Prod., 502, 145397, https://doi.org/10.1016/j.jclepro.2025.145397,

2025.

Zhang, H., Hoff, R. M., & Engel-Cox, J. A.: The Relation between Moderate Resolution Imaging Spectroradiometer (MODIS) Aerosol Optical Depth and $PM_{2.5}$ over the United States: A Geographical Comparison by U.S. Environmental Protection Agency Regions, J. Air Waste Manage., 59(11), 1358-1369. https://doi.org/10.3155/1047-3289.59.11.1358, 2009.

2) The data processing section, especially the lidar data processing, lacks essential technical details. Key missing information includes lidar instrument specifications, data quality control procedures (e.g, cloud screening), lidar extinction retrieval algorithms/methods, and the methods for reconciling the different vertical resolutions between the lidar data (6m) and the meteorological inputs.

Authors' response:

We sincerely apologize for the omission of technical details related to lidar specification parameters and data processing in the original manuscript. In response to the reviewers' feedback, we have revised the original manuscript and supplemented the content in the supplementary materials. ***Section 2.1*** was exchanged with ***Section 2.2*** for better description. The specific additions are outlined below.

Section 2.1.1: "*The $\sigma_{bsc,532}$ data for deep learning module training and $PM_{2.5}$ chemical component retrieving is obtained from a ground-based dual-wavelength polarization Mie lidar at the Institute of Atmospheric Physics (IAP), Chinese Academy of Sciences (CAS), Beijing (39.98°N,116.38°E). This Mie lidar has consistently detected optical signals since 2017, offering a temporal resolution of 15 minutes and a vertical resolution of 6 m. The lidar specification parameters and data preprocessing are detailed in Text S1 (and Table S1) and Text S2 of the supplement, respectively. The $\sigma_{bsc,532}$ data from February 8-15th, 2021 at 23 lidar sites in the North China Plain (NCP), provided by the China National Environmental Monitoring Center (CNEMC), were utilized to assess the spatial generalization ability. The multi-site data offers a temporal resolution of 5-20 minutes and a vertical resolution of 7.5 m. To generate an hourly resolution lidar dataset, minute-level data were resampled using a simple*

averaging method. Specifically, the arithmetic mean was calculated from all valid minute-level data points within each non-overlapping one-hour window aligned to the start of each hour (e.g., from 00:00 to 00:59)."

Section 2.1.2, Line 111-115: "*…The grid cells of EAC4 and ERA5 that contain the lidar sites were extracted using the k-nearest neighbor search method based on longitude and latitude data (Friedman et al., 1977). The lidar data and the reanalysis data were interpolated onto a preset vertical grid with a height range of 50 m to 3 km using linear interpolation. The preset height information is presented in Text S2 of the supplement.*"

Supplement, Text S1, Lidar instrument specifications: "*As shown in Table S1, the laser emission at wavelengths of 532 nm and 1064 nm relies on a Nd:YAG laser with a second harmonic generator and is corrected by a beam expander before emission. The emitted laser energies at 532 nm and 1064 nm are 30 mJ and 20 mJ, respectively. The laser pulse repetition frequency can reach up to 20 Hz and is set to 10 Hz in practice. The scattered light is collected by a Schmidt-Cassegrain telescope with a diameter of 20 cm and then is collimated and corrected toward a dichroic mirror to separate the received lidar signals at 532 nm and 1064 nm. The lidar signal at 532 nm is separated into horizontal and vertical polarization components and is measured by a photomultiplier tube. The lidar signal at 1064 nm is directly detected by an avalanche photodiode. Finally, the detected lidar signals are recorded by a digital oscilloscope and then are transferred to a computer for data storage.*"

Table S1. The main specification parameters of dual-wavelength polarization Mie Lidar.

Parameter categories		Description
Laser type		Flashlamp pumped Nd:YAG
Laser pulse energy	532 nm	30 mJ/pulse
	1064 nm	20 mJ/pulse
Pulse Repetition Frequency		≤ 20 Hz, 10 Hz used in this work
Telescope Type		Schmidt Cassegrain
Telescope diameter		20 cm
Field of view		1 mrad
Detector type	532 nm	Photomultiplier tube (PMT)
	1064 nm	Avalanche photodiode (APD)
Data acquisition system		Digital oscilloscope

Supplement, Text S2, Lidar data preprocessing: "*A comprehensive data quality control procedure was implemented on the original lidar signals to mitigate issues raised by electrical signal errors and signal offsets caused by background radiation. First, background noise was removed by subtracting the average value of signals within the altitudes of 3-9 km from the original lidar signal. Second, the lidar signal was range-corrected by multiplying by the square of the altitude and corrected for the geometric overlap effect using an empirically determined function derived from lidar profiles under well-mixed atmospheric conditions. Third, a cloud-screening algorithm was applied to identify and remove profiles contaminated by clouds. The algorithm operates by first calculating the vertical gradient of the range-corrected signal. It then identifies potential cloud bases as regions where this gradient exceeds a primary threshold of 4* $\times 10^{-8}$ *for at least 3 consecutive resolution layers. For each candidate cloud layer, the algorithm determines the cloud top and then validates the layer by checking if the maximum signal within it surpasses a secondary threshold of* 5×10^{-6}*. Profiles containing such validated cloud layers were entirely excluded from the subsequent aerosol analysis. Finally, the extinction coefficient at a wavelength of 532 nm was retrieved based on Fernald algorithm (Fernald, 1984).*

To facilitate data fusion and comply with the input requirements of the machine learning model, all data were vertically re-sampled onto a standardized set of preset height levels ranging from 50 m to 3 km. The high-resolution lidar data and the low-resolution global reanalysis data were interpolated onto this uniform vertical grid using linear interpolation. The preset height grid with logarithmic intervals can be determined by Eq. S1-S2. Logarithmic interval amplifies vertical resolution within the planetary boundary layer, where fine-mode particles and their chemical components are typically most concentrated (Yang et al., 2024).

$$h_i = 10^{log_{10}(Z_{min})+(i-1)\times\Delta Z}, i = 1, 2, \dots, n \quad (S1)$$

$$\Delta Z = \frac{log_{10}(Z_{max}) - log_{10}(Z_{min})}{n-1} \quad (S2)$$

Where h_i *is the height at* i^{th} *vertical layer,* Z_{min} *is the minimum height,* ΔZ *is the logarithmic interval,* Z_{max} *is the maximum height, and* n *is the total number of*

vertical layers."

Reference

Fernald, F. G.: Analysis of atmospheric lidar observations: some comments, Appl. Opt., 23, 652-653, https://doi.org/10.1364/AO.23.000652, 1984.

Yang, T., Li, H., Xu, W., Song, Y., Xu, L., Wang, H., Wang, F., Sun, Y., Wang, Z., and Fu, P.: Strong Impacts of Regional Atmospheric Transport on the Vertical Distribution of Aerosol Ammonium over Beijing, Environ. Sci. Technol. Lett., 11, 29-34, https://doi.org/10.1021/acs.estlett.3c00791, 2024.

3) The validation dataset is insufficient to support the broad conclusions presented. Aircraft validation comprises only four flights (limited to three different calendar months), while tower measurements span just 11 days across two time periods. This is of particular importance because the retrieved aerosol concentrations appear to show similar vertical distributions across different seasons. The limited validation prevents assessment of whether this method captures realistic atmospheric processes or simply learns scaling relationships under specific (mostly wintertime) meteorological conditions.

Authors' response:

We fully agree that robust and comprehensive validation is essential to support the generalizability of our retrieval framework in spatiotemporally varying scenarios. In this work, we designed a multi-faceted validation strategy to thoroughly validate the model's generalizability from different perspectives by using the non-training dataset. The non-training dataset contains spatiotemporal information that the machine learning model has never learned, enabling a validation of whether this method captures realistic atmospheric processes. The validation strategy is outlined as follows.

a. The independent validation/testing sets, which were not used in deep learning, are employed to evaluate whether the nonlinear mapping relationship established by the model is reliable. As presented in ***Fig. 3* of *Section 3.1.1***, the results demonstrated high predictive accuracy on the independent validation/testing sets, with five $PM_{2.5}$ chemical species showing high agreement ($R \geq 0.69$, $RMSE \leq 8.87$ μg m^{-3}) with ground

observations, indicating that our model can learn complex and nonlinear relationships rather than memorizing the training data.

b. We utilized an independent ground observation set of $PM_{2.5}$ concentrations from three years not included in model training to validate the temporal generalization of our method. The sum of our retrieved $PM_{2.5}$ chemical species concentrations was compared against these independent measurements. As presented in ***Fig. 4b, c and d* of *Section 3.1.2***, the results indicate that our method accurately characterizes the changes in mass concentrations of various $PM_{2.5}$ chemical components across all seasons and under diverse meteorological conditions, not just the wintertime conditions.

c. We utilized a ground observation set of chemical component concentrations from 23 independent sites across the North China Plain (NCP) to validate the spatial generalization of our method. As presented in ***Fig. 5* of *Section 3.1.2***, comparisons between our retrieved surface concentrations and measurements from these untrained sites showed moderate agreement, indicating that our method is not site-specific but possesses robust predictive power on a regional scale.

d. We acknowledge the well-noted challenge of acquiring high-frequency vertical profile observation data. The validation results based on 4 aircraft campaigns and 11-day tower-based measurements showed that our retrieved vertical profiles of chemical components showed encouraging consistency with observations in both shape and magnitude. Most importantly, these vertical validation cases must be interpreted within the context of the replies of **Major Comments #3b, c**. The strong performance in spatiotemporal generalization provides a foundational credibility that our method can capture realistic relationships between extinction coefficient and chemical species (**Major Comments #3b, c**). The vertical validation then confirms that these relationships correctly translate into accurate vertical structures (**Major Comments #3d**).

We acknowledge that acquiring more observations from tower-based and aircraft-based campaigns is essential for adequately strengthening the validation of our method. We openly discuss the limitations regarding the sample size of vertical validation data

in ***Section 3.4 Limitations and uncertainties***. Besides, we emphasize the significant role of temporal generalization validation in addressing concerns about seasonal representativeness in ***Section 2.2.4***.

Section 2.2.4, Line 330-332: "*… (1) We compare the retrieved mass concentrations with the observed values at the surface level during a training year (2021) and three non-training years (2017, 2018 and 2024) to validate the temporal generalization in all seasons and under diverse meteorological conditions…*"

Section 3.4: "*The deep learning module in our retrieval framework can establish a powerful mapping between optical and meteorological features and $PM_{2.5}$ chemical species, and physics-based explicit constraints can enhance the reliability and expandability of the mapping relationships. However, several limitations and sources of uncertainty remain and should be acknowledged when interpreting the results and extending the framework to broader applications.*

First, the spatial scope of the training data is predominantly restricted to the NCP region. Expanding the retrieval framework with data from more diverse geographical locations is necessary to improve its global transferability. Second, the current retrieval framework primarily relies on extinction coefficients at a wavelength of 532 nm, exhibiting dependence on specific lidar instruments. Future retrieval framework should focus on integrating diverse optical features from additional wavelengths to enhancing adaptability and transferability. Third, the auxiliary input data used in both the deep learning module and the physics-constrained optimization are obtained from global reanalysis products, which may not fully capture local atmospheric conditions at specific observational sites, thereby introducing representativeness errors into the retrievals. Acquiring the vertical observational data for these auxiliary features can effectively mitigate the uncertainty induced by the input data. Fourth, the IMPROVE equation applied as an external physical constraint may introduce additional uncertainty into the retrievals due to its systematic estimation biases (Lowenthal and Kumar, 2016). Moreover, since the IMPROVE equation was applied as an external

physical constraint to optimize the retrievals of $PM_{2.5}$ chemical components, the machine learning model itself was not intrinsically constrained by physical principles during its training. Future work could incorporate an internal physical constraint into the machine learning model to improve its physical interpretability by formulating a hybrid loss function for training that combines the traditional data-fitting term with a physical term. Finally, long-term acquisition of independent vertical profiling data from both tower-based and aircraft-based campaigns is essential for a comprehensive assessment of the robustness of the vertical retrievals with respect to varying sites, aerosol types, and seasons."

4) The manuscript would benefit from some extensive editing to improve its readability. Sentences are overly dense, and the model development section would be difficult for most readers to follow. The excessive number of figures (~75 figure/subplots) dilutes the presentation of key results.

Authors' response:

We sincerely thank the reviewer for this critical feedback regarding the readability and presentation of the manuscript. In response to the reviewer's suggestions, we have undertaken a comprehensive revision of the manuscript, including the entire text, all figures, and their captions. Several detailed supporting subfigures have been moved to the supplement materials to maintain a clean and focused flow in the main text while still providing all necessary data for interested specialists. The final revised manuscript can be found in the uploaded file ***manuscript_with track changes.docx***.

5) The manuscript lacks an adequate discussion of the limitations of this retrieval technique. This would be essential for readers considering applying this method in different regions or with slightly different instruments.

Authors' response:

We sincerely thank the reviewer for this critical suggestion. We fully agree that a

thorough discussion of the limitations is essential for effective application in different scenarios. In response, we have added a new section titled "***3.4 Limitations and uncertainties***" of our revised manuscript. This section provides a detailed examination of the constraints and potential uncertainties of our proposed retrieval framework. ***Section 3.4 Limitations and uncertainties*** has been presented in the replies of **Major Comments #3**.

6) The SHAP feature importance analysis raises some questions and methodological concerns. For example, why are specific humidity and relative humidity treated as independent? These are clearly related. There is also lacking a discussion about the definition of and why "geopotential" is so important. Also, it strikes me that the combined SHAP value of extinction, relative humidity, and v-wind being under 50% is relatively low considering they are noted to determine the vertical structure and chemical and physical processes (L410-411).

Authors' response:

We thank the reviewer for these insightful comments and methodological concerns regarding the SHAP analysis. We have revised the manuscript accordingly to address each point.

Q1: Why are specific humidity and relative humidity treated as independent?

A1: Relative humidity exerts a well-established driving role on $PM_{2.5}$ through its influence on hygroscopic growth, aqueous chemistry, and heterogeneous reactions (Chen et al., 2020). In contrast, specific humidity, which represents the total moisture content of a wet air mass, is more closely linked to the vertical diffusion and wet scavenging of $PM_{2.5}$ (Chatfield et al., 2020). Therefore, including both relative humidity and specific humidity as independent features allows the machine model to leverage their complementary roles in governing $PM_{2.5}$ chemical compositions.

Q2: Lacking a discussion about the definition of and why "geopotential" is so important.

A2: Geopotential is an integrated feature that reflects the synoptic meteorological conditions when combined with wind fields and is closely related to $PM_{2.5}$ pollution processes, such as accumulation, transboundary transport and dispersion (Jia et al., 2022; Wang et al., 2021). Crucially, synoptic meteorological conditions identified by geopotential patterns largely determine the development of the planetary boundary layer, influencing the vertical distribution of $PM_{2.5}$ (Miao et al., 2022; Xu et al., 2019).

Section 3.2, Line 473-477: "*…Specific humidity (SH) and geopotential (GEOP) also provided important contributions (13.04% and 12.85%, respectively). SH is related to the vertical diffusion and wet scavenging of pollutants (Chatfield et al., 2020) and GEOP identifies the synoptic meteorological patterns that affect both horizontal process (Jia et al., 2022; Wang et al., 2021) and vertical distribution of pollutants within the boundary layer (Miao et al., 2022; Xu et al., 2019).*"

Reference

Chatfield, R. B., et al.: Satellite mapping of $PM_{2.5}$ episodes in the wintertime San Joaquin Valley: a "static" model using column water vapor, Atmos. Chem. Phys., 20, 4379-4397, https://doi.org/10.5194/acp-20-4379-2020, 2020.

Chen, Z., et al.: Influence of meteorological conditions on $PM_{2.5}$ concentrations across China: A review of methodology and mechanism, Environ. Int., 139, 105558, https://doi.org/10.1016/j.envint.2020.105558, 2020.

Jia, Z., et al.: The impact of large-scale circulation on daily fine particulate matter ($PM_{2.5}$) over major populated regions of China in winter, Atmos. Chem. Phys., 22, 6471-6487, https://doi.org/10.5194/acp-22-6471-2022, 2022.

Miao, Y., et al.: Influence of Multi-Scale Meteorological Processes on $PM_{2.5}$ Pollution in Wuhan, Central China, Front. Environ. Sci., 10, https://doi.org/10.3389/fenvs.2022.918076, 2022.

Wang, X., et al.: Dominant synoptic patterns associated with the decay process of $PM_{2.5}$ pollution episodes around Beijing, Atmos. Chem. Phys., 21, 2491-2508, https://doi.org/10.5194/acp-21-2491-2021, 2021.

Xu, Y., et al.: Two Inversion Layers and Their Impacts on $PM_{2.5}$ Concentration over the Yangtze River Delta, China, J. Appl. Meteor. Climatol., 58, 2349-2362, https://doi.org/10.1175/JAMC-D-19-0008.1, 2019.

Q3: The combined SHAP value of extinction, relative humidity, and v-wind being

under 50% is relatively low considering they are noted to determine the vertical structure and chemical and physical processes.

A3: The SHAP values presented in ***Section 3.2*** were derived from a dataset spanning approximately two years. This long-term perspective reveals that the vertical distribution of $PM_{2.5}$ chemical components is governed by the complex and nonlinear interaction of a multitude of driving features, which contrasts with specific pollution episodes where a single driver may be dominant. From another perspective, our results indicate that the machine learning model effectively captured a complex multi-factorial relationship, rather than relying on an oversimplified representation dependent on a few dominant features. In response to this comment, we have revised the manuscript by replacing "dominant" with the more appropriate term "significant". The revision is as follows.

Section 3.2, Line 459-462: "*Figure 7a1-a5 depicts that the aerosol extinction coefficient at 532 nm (EXT), relative humidity (RH) and v-component wind (VW) are the significant input features for predicting the five $PM_{2.5}$ chemical components with an averaged relative contribution of 14.43 %, 15.84 % and 16.77 %. These features largely affect the vertical structure, chemical and physical processes, respectively…*"

3 Minor Comments:

1) The introduction would benefit from making note of previous work in retrieving $PM_{2.5}$ concentrations from space-based lidar (e.g., Matus et al., 2024; Toth et al., 2022).

Authors' response:

We thank the reviewer for the suggestion. The revised version is as follows.

Introduction, Line 45-47: "*…Continuous remote-sensing lidar detection technologies with high temporal and vertical resolution serve as robust pathways for the constant identification of $PM_{2.5}$ and its components across all altitudes (Matus et al., 2024; Toth et al., 2022; Wang et al., 2022) …*"

Reference

Matus, A. V., Nowottnick, E. P., Yorks, J. E., and da Silva, A. M.: Enhancing surface $PM_{2.5}$ air quality estimates in GEOS using CATS lidar data, Earth and Space Sci., 12, e2024EA004078, https://doi.org/10.1029/2024EA004078, 2025.

Toth, T. D., Zhang, J., Vaughan, M. A., Reid, J. S., and Campbell, J. R.: Retrieving particulate matter concentrations over the contiguous United States using CALIOP observations, Atmos. Environ., 274, 118979, https://doi.org/10.1016/j.atmosenv.2022.118979, 2022.

2) L23: Specify the Chinese megacity (Beijing-Tianjin-Hebei region)

Authors' response:

We thank the reviewer for the suggestion. The revised version is as follows.

Abstract, Line 21-23: "*…Finally, a dataset of vertical mass concentration profiles of these components over six years in a Chinese megacity (Beijing) was generated by the retrieval framework…*"

3) L41-42: Citing papers for examples of tower, aircraft, balloon, and UAV campaigns is not necessary. These are very common platforms for atmosphere remote sensing.

Authors' response:

We thank the reviewer for the suggestion, and we have removed these citations in the revised manuscript.

Introduction, Line 41-42: "*Field campaigns are widely conducted to obtain vertical profiles of $PM_{2.5}$ chemical components by mounting observation instruments on meteorological towers, aircraft, tethered balloons and unmanned aerial vehicles…*"

4) L263-264: Specify how these data are averaged.

Authors' response:

We thank the reviewer for the suggestion. The revised version is as follows.

Section 2.1.1, Line 96-98: "*…To generate an hourly resolution lidar dataset, minute-level data were resampled using a simple averaging method. Specifically, the arithmetic*

mean was calculated from all valid minute-level data points within each non-overlapping one-hour window aligned to the start of each hour (e.g., from 00:00 to 00:59)."

5) Figure 3: Specify the observations used in the figures. All altitudes from the tower and airplane?

Authors' response:

We thank the reviewer for the suggestion. To clarify, ***Fig. 3*** exclusively uses ground-level observations to independently validate the ground-level predictions from our machine learning model, as the model itself was trained and tested exclusively on surface data. The tower-based and aircraft-based measurements were reserved for the independent validation of the retrieved vertical profiles, which is presented in ***Fig 6***. In response to the suggestion, the revised version is as follows.

Section 2.2.4, Line 322-327: "*An hourly multivariate dataset with extensive temporal coverage was employed to train and evaluate the deep learning module. To maintain temporal independence, the training (and validation) set was constructed from a 1-year (2021) time-series dataset obtained from a Beijing site (Fig. S1), while the testing set contains an independent 6-month (Jan 1-Mar 31 and Jun 1 to Aug 31, 2022) time-series dataset obtained from the same site. A 10-fold time-series cross-validation (CV) scheme was designed for the training (and validation) set to preserve its temporal order and prevent future information leakage, which is detailed in Text S3 and Fig. S2 of the supplement. The iteration number of Bayesian optimization is set to 20.*"

Section 3.1.1, Line 345-347: "*The 10-fold CV sets and a testing set with temporal independence are utilized to evaluate the predictive performance of the deep learning module, which is quantified by the discrepancies between simulations and observations at ground level for NH_4^+, SO_4^{2-}, NO_3^-, OM and BC…*"

6) Figure 4: There are many cases where the retrieval is not particularly close to the observations. A broader discussion about these cases would be valuable.

Authors' response:

We thank the reviewer for the suggestion and have added a broader discussion in the revised manuscript as follows. Besides, the discussion in retrieval uncertainties has been presented in the replies of **Major Comments #3**.

Section 3.1.2, Line 385-392: "*...These results indicate that the retrieval framework roughly interprets the changes in concentrations of various chemical components across different periods, exhibiting fundamental temporal generalization capabilities. However, the retrieved concentrations show some overestimation cases during autumn in 2018 and spring in 2024, potentially associated with the uncertainties induced by the training data. The training data may lack a sufficiently diverse spectrum of meteorological conditions and pollution patterns, which limits the temporal generalizability of the retrieval framework across all complex and dynamic atmospheric scenarios. Future efforts should enhance retrieval accuracy by augmenting the training data with observations spanning a wider range of temporal conditions.*"

7) Figure 5: Subplot "a" needs a better explanation. What is conveyed differently in the histograms versus the dots? Subplots in the "b" and "c" rows would benefit from a better map. Readers from outside China may be lost without other context (coloring the ocean/seas, highlighting major cities, etc.)

Authors' response:

Figure 5a highlights the data distribution properties derived from the retrieved and observed surface mass concentration of NH_4^+, NO_3^-, SO_4^{2-}, OM and BC at 39 non-training BTH lidar sites over a period of February 8-15th, 2021. Figure 5a combines dotplots, boxplots and Kernel density. Kernel density highlights the overall shape of the data distribution. Dotplots, while similar in Kernel density, also convey information about the exact number of datapoints across the distribution.

In response to the Reviewer's comments, we have removed the dotplots. Because the high density of data points resulted in a cluttered presentation, and the information

they convey about the distribution is effectively captured by the kernel density, which provides a smoother and more interpretable representation. Then we added geographic basemaps into Figure 5b, c for better presentation.

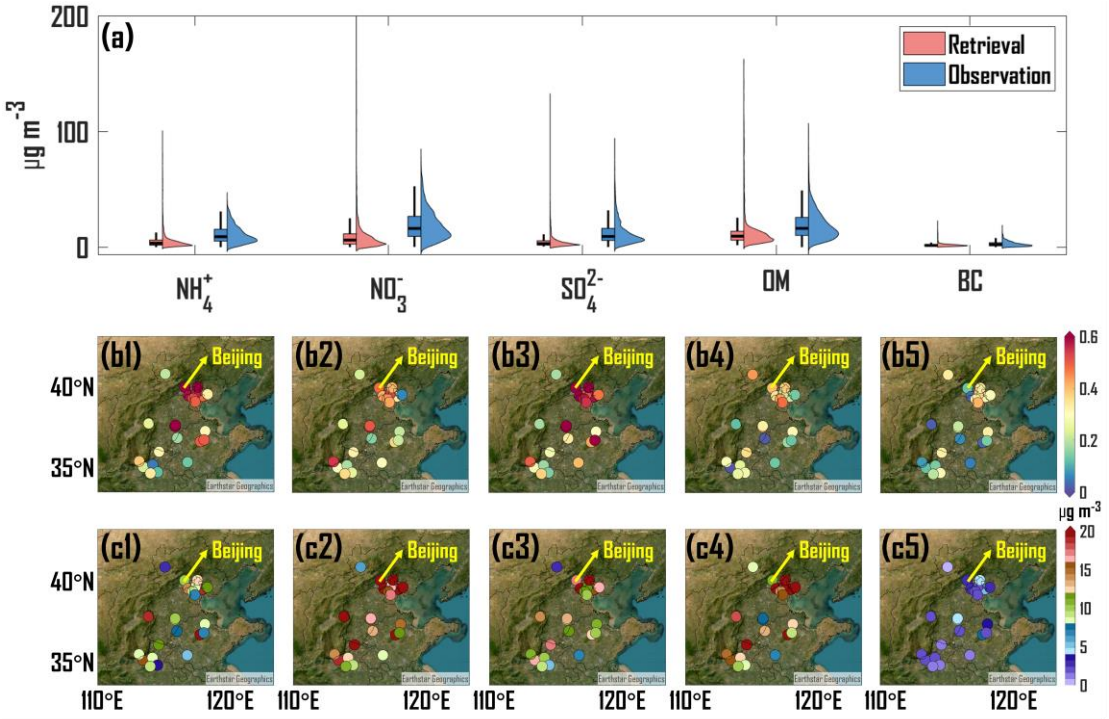

***Figure 5**: Data distribution properties of retrieved and observed surface mass concentration (μg m^{-3}) of NH_4^+, NO_3^-, SO_4^{2-}, OM and BC at 23 non-training NCP lidar sites over a period of February 8-15th, 2021, presented by a combination of boxplots and kernel density (a). Spatial distribution of Pearson correlation coefficient (CORR) between retrieved and observed surface mass concentration of NH_4^+ (b1), NO_3^- (b2), SO_4^{2-} (b3), OM (b4) and BC (b5). (c1-c5) Same as (b1-b5) but for root mean square error (RMSE, μg m^{-3}). The geographic basemap is hosted by Esri (https://www.esri.com/en-us/home).*

8) L418: Define "upper atmosphere" in this context

Authors' response:

We thank the reviewer for the reminder. The term "upper atmosphere" is inaccurate, and we have replaced with "upper planetary boundary layer" in the revised manuscript.

Section 3.2, Line 469-470: "*…The aerosol content in the upper planetary boundary layer is relatively low…*"

Section 3.2, Line 471-472: "*…Conversely, pollution transport in the upper planetary boundary layer is less affected by interference from complex underlying surfaces than near-surface transport…*"