Point-to-point responses

*We appreciate the reviewers for their valuable and constructive comments, which are very helpful for the improvement of the manuscript. We have revised the manuscript carefully according to the reviewers' comments. We have addressed the reviewers' comments on a point-to-point basis as below for consideration, where the reviewers' comments are cited in **black**, and the responses are in **blue**.*

Referee #1

Jiao et al. present a comprehensive dataset of vertical profile observations of aerosol, NO_2, and HCHO in China from 2019 to 2023 using a hyperspectral remote sensing network. The dataset fills a critical gap in vertical profile monitoring, providing high temporal resolution and wide geographic coverage. The manuscript is well-organized and offers valuable insights into the spatial-temporal variations of atmospheric components, which can inform environmental policies and enhance scientific modeling. However, there are several areas where the manuscript could be improved to enhance clarity, accuracy, and impact.

1) Section 2.2 mentions data filtering criteria (e.g., DOF, relative error thresholds), but the rationale behind these criteria needs to be explained in more detail. In addition, it is recommended to include a specific site example to demonstrate the differences in data distribution before and after filtering.

Re: Thanks for your great comments. In response to the comment regarding the explanation of the data filtering criteria, we have added Section S1 in the supplement, where we provide a detailed description of this process. Additionally, we have selected the CAMS and HNI sites as examples and included Table S3 in the supplement, which demonstrates the differences in the data before and after the application of the data filtering procedures. The content of Section S1 is as follows:

"Section S1. Data Filtering Criteria and Principles

DOF (Degrees of Freedom): In the retrieval process, the averaging kernel matrix reflects the sensitivity of the observation data to the state vector transformation. The sum of the diagonal elements of this matrix represents the Degrees of Freedom (DOF), which quantifies the amount of independent information that can be effectively extracted from the observational data. A higher DOF indicates that more independent information has been extracted, whereas a lower DOF suggests that the observation data is not sensitive to changes in the state vector. Consequently, we exclude data with a DOF value lower than 1.0, as this typically indicates that the retrieval process has not effectively utilized the observational data, leading to potentially unreliable results or higher uncertainty.

χ^2 (Chi-Square): As shown in Equation (1), the χ^2 value is used to assess the model fit, i.e., the difference between the model predictions and the actual observed values. A smaller χ^2 value indicates that the model closely matches the observed values, while a larger χ^2 value suggests a significant discrepancy, which may arise from inaccurate prior information, systematic errors, or other unaccounted factors. In order to ensure the quality and reliability of the analysis, we exclude data with χ^2 values greater than 200. This threshold helps to minimize the influence of outliers and ensures that the results are based on high-quality data.

$$\chi^2 = \left(\boldsymbol{y} - F(\boldsymbol{x},\boldsymbol{b})\right)^T \boldsymbol{S}_{\varepsilon}^{-1}\left(\boldsymbol{y} - F(\boldsymbol{x},\boldsymbol{b})\right) + (\boldsymbol{x}-\boldsymbol{x}_{\boldsymbol{a}})^T \boldsymbol{S}_{\boldsymbol{a}}^{-1}(\boldsymbol{x}-\boldsymbol{x}_{\boldsymbol{a}}) \ , \quad (1)$$

SZA (Solar Zenith Angle): In this study, we focus on the absorption in the troposphere near the Earth's surface. When the solar zenith angle (SZA) exceeds 75°, the predominant scattering transpire within the lower stratosphere and upper troposphere (Song et al., 2023).

In this case, DOAS measurements exhibit heightened sensitivity towards stratospheric absorption, while demonstrating reduced sensitivity to absorption in the proximity of the surface. Therefore, measurements with SZA > 75° are excluded from the analysis.

CI (Color Index): The color index (CI) is defined as the ratio of the spectral intensities at 330 nm and 390 nm, and is used to identify potential cloud interference (Wagner et al., 2016). A polynomial function of time is fitted to the data from clear days without significant cloud cover to establish the diurnal CI variation pattern. Based on this fitted model, a CI threshold is determined for each time point. If the CI value at a given time is below 10% of the threshold, it is assumed that the data are affected by cloud interference, and such data are excluded from further analysis (Ryan et al., 2018).

Relative Retrieval Error: The relative retrieval error measures the precision of the retrieval results. If the relative error exceeds 50%, the result is considered to have too high an uncertainty and is therefore excluded as unreliable data (Tan et al., 2018).

Due to the peculiarity of the instrument's data collection and storage process, data with SZA and CI values that do not meet the criteria are excluded prior to processing. After filtering based on DOF, χ^2, and relative retrieval error, approximately 10% of the data is discarded."

The content of Table S3 is as follows:

"Table S3. Differences in the data before and after filtering based on DOF, χ^2, and relative retrieval error, using the CAMS and HNI stations as examples. The numbers in the table represent the monthly data integrity at each site, indicating the ratio of days with and without data in a month."

	CAMS	CAMS (filtered)	HNI	HNI (filtered)
2019-01	1.00	1.00	0.00	0.00
2019-02	1.00	1.00	0.00	0.00
2019-03	0.94	0.94	0.00	0.00
2019-04	0.67	0.67	0.07	0.07
2019-05	0.87	0.87	0.06	0.06
2019-06	0.97	0.97	0.97	0.97
2019-07	1.00	1.00	1.00	0.97
2019-08	0.42	0.42	0.48	0.48
2019-09	0.43	0.43	0.00	0.00
2019-10	0.94	0.94	0.00	0.00
2019-11	1.00	1.00	0.97	0.97
2019-12	0.68	0.68	0.94	0.94
2020-01	0.94	0.94	1.00	1.00
2020-02	0.97	0.97	1.00	1.00
2020-03	1.00	1.00	1.00	1.00
2020-04	1.00	1.00	0.93	0.93
2020-05	1.00	1.00	0.94	0.94
2020-06	1.00	1.00	1.00	0.97
2020-07	1.00	1.00	0.39	0.39
2020-08	1.00	1.00	0.00	0.00
2020-09	0.97	0.97	0.00	0.00
2020-10	0.97	0.97	0.00	0.00
2020-11	0.83	0.83	0.93	0.90
2020-12	0.97	0.97	1.00	1.00
2021-01	0.97	0.97	0.94	0.94
2021-02	0.93	0.93	0.79	0.79
2021-03	0.00	0.00	0.52	0.52
2021-04	0.27	0.27	0.27	0.27

2021-05	0.97	0.97	0.77	0.77
2021-06	0.97	0.97	0.80	0.80
2021-07	0.48	0.48	0.84	0.84
2021-08	0.00	0.00	0.81	0.81
2021-09	0.00	0.00	0.43	0.43
2021-10	0.65	0.65	0.16	0.16
2021-11	0.77	0.77	0.67	0.67
2021-12	1.00	1.00	1.00	1.00
2022-01	0.97	0.97	0.81	0.81
2022-02	1.00	1.00	0.96	0.96
2022-03	1.00	1.00	1.00	1.00
2022-04	0.20	0.20	0.67	0.67
2022-05	0.55	0.55	0.00	0.00
2022-06	1.00	1.00	0.00	0.00
2022-07	1.00	1.00	0.00	0.00
2022-08	1.00	1.00	0.00	0.00
2022-09	1.00	1.00	0.00	0.00
2022-10	1.00	1.00	0.00	0.00
2022-11	0.97	0.97	0.00	0.00
2022-12	0.45	0.45	0.00	0.00
2023-01	1.00	0.97	0.00	0.00
2023-02	0.57	0.57	0.00	0.00
2023-03	1.00	1.00	0.00	0.00
2023-04	0.87	0.87	0.00	0.00
2023-05	0.90	0.90	0.00	0.00
2023-06	0.97	0.97	0.00	0.00
2023-07	0.97	0.97	0.00	0.00
2023-08	0.90	0.90	0.00	0.00
2023-09	0.77	0.77	0.00	0.00
2023-10	0.29	0.29	0.00	0.00

2) In Section 3.4 (Validation), the manuscript shows good correlations between the dataset and CNEMC and TROPOMI data. However, there is a lack of a detailed discussion on the potential biases and uncertainties in these comparisons. For example, how do the differences in spatial and temporal resolution between MAX-DOAS and TROPOMI contribute to the observed discrepancies?

Re: Thank you for your insightful comment. We agree that it is essential to illustrate the potential biases and uncertainties in the comparison between our dataset and the TROPOMI and CNEMC data. To provide a more comprehensive explanation, we discuss the sources of these potential discrepancies as follows:

Firstly, MAX-DOAS typically provides high-resolution, site-specific measurements, whereas TROPOMI offers averaged values over larger spatial grids. This difference in spatial resolution can introduce discrepancies when comparing data from the same geographic area. For instance, localized pollution hotspots detected by MAX-DOAS may be averaged out in TROPOMI data, potentially leading to a dilution of the observed concentrations. However, the comparison between these datasets remains valid, as both capture important atmospheric features, albeit with different spatial representativeness. The robust correlation observed between these datasets suggests that, despite the spatial differences, the overall trends and patterns of atmospheric composition are well captured.

Additionally, although MAX-DOAS data were averaged over a 30-minute window to align with TROPOMI's overpass time, there may still be potential discrepancies caused by instantaneous variations in meteorological conditions and pollution sources. For example, rapid weather changes or localized pollution events could cause significant concentration fluctuations that are more pronounced on shorter timescales, leading to discrepancies between the datasets. Nevertheless, despite these potential short-term deviations, the broad atmospheric trends are accurately represented.

Furthermore, the CNEMC data has a higher temporal resolution and typically aligns well with the MAX-DOAS measurements. However, there are differences in both location and distance between the CNEMC stations and the MAX-DOAS sites, which may introduce some discrepancies. This discrepancy is especially evident in regions with heterogeneous pollution sources. To illustrate this potential difference, we calculated the distance between each MAX-DOAS site and the nearest CNEMC station, as shown in Table S4. The strong correlation between the MAX-DOAS and CNEMC data further supports the reliability of our dataset.

In parallel, it is important to note that the MAX-DOAS sites measure the path-integrated average concentration along the observation line of sight, while the CNEMC stations focus on point-specific concentration measurements at fixed locations, providing highly localized concentration data. Although the two methods differ in data acquisition, our dataset has undergone rigorous quality control and calibration procedures to ensure its accuracy and reliability. It provides valuable insights into the broader distribution of pollutants, demonstrating its effectiveness and credibility in air quality monitoring.

Finally, it is important to note that despite these inherent uncertainties, our data and the related validation demonstrate strong credibility and provide a reliable foundation for the conclusions drawn.

3) While the dataset covers seven major regions, monitoring sites are fewer in regions such as Central and Southwest China. The manuscript should discuss how this uneven distribution might impact regional representativeness and the generalizability of the results.

Re: Thank you for your valuable comments. Regarding the potential impact of uneven site distribution on regional representativeness and the generalizability of the results, while our dataset covers seven major geographical regions of China and provides valuable information on atmospheric composition across the country, we acknowledge that the number of monitoring sites is relatively limited in regions such as Central China and Southwest China (1 and 2 stations, respectively), while the North and East regions have a higher density of stations (9 and 13 stations, respectively). This uneven distribution could indeed affect the regional representativeness and the general applicability of the results. Specifically, the regions with fewer sites may not fully capture the spatial and temporal variations in atmospheric composition, especially considering the complex and diverse topography of these areas (e.g., mountains, basins, and plains). A limited number of sites may struggle to capture all the significant environmental features and pollution dynamics within these regions. For instance, the mountainous and canyon terrain in the Southwest may lead to different pollutant dispersion and accumulation patterns compared to the plains. As a result, we have revised the summary to include a discussion on the limitations of the dataset. The revised section is as follows:

Line 391, "(4) In regions such as Central China and Southwest China, the number of monitoring sites is relatively limited. This uneven distribution may potentially affect the representativeness of the results for these areas."

Although there are certain limitations, we are confident in the overall quality of the dataset presented in the manuscript. First, the dataset provides atmospheric composition observation data across the entire country, encompassing a variety of geographical environments, from urban centers to remote mountainous areas, offering a wealth of valuable information. Second, the high-resolution observations at each station capture subtle variations in local atmospheric composition, which is crucial for understanding inter-regional differences. Additionally, the reliability and accuracy of the data have been further strengthened through comparison with ground-based data from CNEMC and satellite data from TROPOMI.

Future work will focus on expanding the network of monitoring sites, with particular emphasis on increasing the station density in Central China, Southwest China, Northeast China, and Northwest China. This expansion aims to ensure the comprehensiveness and representativeness of the dataset, addressing the current uneven distribution and enhancing the overall coverage of the monitoring network.

4) Technical comments.

Line 26.

'Its sharing would facilitate the scientific community in exploring of source-receptor relationships' -> 'Its sharing would facilitate the scientific community in exploring source-receptor relationships'.

Re: Thanks for your comment. We have rewritten this sentence as "Its sharing would facilitate the scientific community in exploring source-receptor relationships".

Line 42.

'Aerosol, as one of the most complex and critical composition of the atmospheric environment' -> 'Aerosol, as one of the most complex and critical compositions of the atmospheric environment'.

Re: Thanks for your comment. We have rewritten this sentence as "Aerosol, as one of the most complex and critical compositions of the atmospheric environment".

Line 98.

'It helps provide a complete perspective on the vertical distribution of aerosol, NO_2, and HCHO in China' -> 'The diversity of these monitoring sites helps provide a complete perspective on the vertical distribution of aerosol, NO_2, and HCHO in China'.

Re: Thanks for your comment. We have rewritten this sentence as "The diversity of these monitoring sites helps provide a complete perspective on the vertical distribution of aerosol, NO_2, and HCHO in China".

Line 108.

'…which located in China's economically developed and densely populated areas' -> '…which are located in China's economically developed and densely populated areas'.

Re: Thanks for your comment. We have rewritten this sentence as "…which are located in China's economically developed and densely populated areas".

Line 110.

'offer vertical distribution data across different elevation' -> 'offer vertical distribution data across different elevations'.

Re: Thanks for your comment. We have rewritten this sentence as "Sites at Mount Tai (TS) and Taian (TA), located at 1500 and 170 m respectively, offer vertical distribution data across different elevations".

Line 223.

'AECs in spring, summer, autumn, and winter accounts for 23.60%, 24.63%, 24.69%, and 27.08% of the total averaged values of four seasons, respectively' -> 'AECs in spring, summer, autumn, and winter account for 23.60%, 24.63%, 24.69%, and 27.08% of the total averaged values of four seasons, respectively'.

Re: Thanks for your comment. We have rewritten this sentence as "AECs in spring, summer, autumn, and winter account for 23.60%, 24.63%, 24.69%, and 27.08% of the total averaged values of four seasons, respectively".

Line 237.

'High-concentration aerosol with extinction coefficients exceeding 1.0 km^{-1} are primarily distributed below 600 m, while aerosol with extinction coefficients greater than 0.6 km^{-1} are concentrated below 1000 m' -> 'High-concentration aerosols with extinction coefficients exceeding 1.0 km^{-1} are primarily distributed below 600 m, while aerosols with extinction coefficients greater than 0.6 km^{-1} are concentrated below 1000 m'.

Re: Thanks for your comment. We have rewritten this sentence as "High-concentration aerosols with extinction coefficients exceeding 1.0 km^{-1} are primarily distributed below 600 m, while aerosols with extinction coefficients greater than 0.6 km^{-1} are concentrated below 1000 m".

Line 240.

'with one peak occurring before 12:00 BJT and the other between 16:00 and 18:00 BJT' -> 'with one peak occurring before 12:00 Beijing Time (BJT) and the other between 16:00 and 18:00 BJT'.

Re: Thanks for your comment. We have rewritten this sentence as "At many sites, the AEC exhibits a bimodal pattern, with one peak occurring before 12:00 Beijing Time (BJT) and the other between 16:00 and 18:00 BJT".

Line 243.

'located in the central Beijing' -> 'located in central Beijing'.

Re: Thanks for your comment. We have rewritten this sentence as "This bimodal pattern is more pronounced at urban sites such as CAMS, IAP (located in central Beijing) and SH_XH (located in central Shanghai), likely due to the significant contribution of traffic emissions during morning and evening rush hours.".

Line 287.

'The averaged near-surface NO_2 concentrations in spring, summer, autumn, and winter accounts for 23.06%, 16.57%, 25.74%, and 34.63% of the total averaged values of four seasons, respectively' -> 'The averaged near-surface NO_2 concentrations in spring, summer, autumn, and winter account for 23.06%, 16.57%, 25.74%, and 34.63% of the total averaged values of four seasons, respectively'.

Re: Thanks for your comment. We have rewritten this sentence as "The averaged near-surface NO_2 concentrations in spring, summer, autumn, and winter account for 23.06%, 16.57%, 25.74%, and 34.63% of the total averaged values of four seasons, respectively".

Line 310.

'transportation' -> 'transportation from'.

Re: Thanks for your comment. We have rewritten this sentence as "HCHO originates from diverse sources, including fossil fuel combustion (Ho et al., 2012; Schauer et al., 2002), biomass burning (Carlier et al., 1986; Lee et al., 1997), transportation from, and industrial activities (Buzcu Guven and Olaguer, 2011)".

References

Buzcu Guven, B. and Olaguer, E. P.: Ambient formaldehyde source attribution in Houston during TexAQS II and TRAMP, Atmospheric Environment, 45, 4272–4280, https://doi.org/10.1016/j.atmosenv.2011.04.079, 2011.

Carlier, P., Hannachi, H., and Mouvier, G.: The chemistry of carbonyl compounds in the atmosphere—A review, Atmospheric Environment (1967), 20, 2079–2099, https://doi.org/10.1016/0004-6981(86)90304-5, 1986.

Ho, S. S. H., Ho, K. F., Lee, S. C., Cheng, Y., Yu, J. Z., Lam, K. M., Feng, N. S. Y., and Huang, Y.: Carbonyl emissions from vehicular exhausts sources in Hong Kong, Journal of the Air & Waste Management Association, 62, 221–234, https://doi.org/10.1080/10473289.2011.642952, 2012.

Lee, M., Heikes, B. G., Jacob, D. J., Sachse, G., and Anderson, B.: Hydrogen peroxide, organic hydroperoxide, and formaldehyde as primary pollutants from biomass burning, Journal of Geophysical Research: Atmospheres, 102, 1301–1309, https://doi.org/10.1029/96JD01709, 1997.

Ryan, R. G., Rhodes, S., Tully, M., Wilson, S., Jones, N., Frieß, U., and Schofield, R.: Daytime HONO,

NO_2 and aerosol distributions from MAX-DOAS observations in Melbourne, Atmospheric Chemistry and Physics, 18, 13969–13985, https://doi.org/10.5194/acp-18-13969-2018, 2018.
Schauer, J. J., Kleeman, M. J., Cass, G. R., and Simoneit, B. R. T.: Measurement of Emissions from Air Pollution Sources. 5. C1−C32 Organic Compounds from Gasoline-Powered Motor Vehicles, Environ. Sci. Technol., 36, 1169–1180, https://doi.org/10.1021/es0108077, 2002.
Song, Y., Xing, C., Liu, C., Lin, J., Wu, H., Liu, T., Lin, H., Zhang, C., Tan, W., Ji, X., Liu, H., and Li, Q.: Evaluation of transport processes over North China Plain and Yangtze River Delta using MAX-DOAS observations, Atmospheric Chemistry and Physics, 23, 1803–1824, https://doi.org/10.5194/acp-23-1803-2023, 2023.
Tan, W., Liu, C., Wang, S., Xing, C., Su, W., Zhang, C., Xia, C., Liu, H., Cai, Z., and Liu, J.: Tropospheric NO_2, SO_2, and HCHO over the East China Sea, using ship-based MAX-DOAS observations and comparison with OMI and OMPS satellite data, Atmospheric Chemistry and Physics, 18, 15387–15402, https://doi.org/10.5194/acp-18-15387-2018, 2018.
Wagner, T., Beirle, S., Remmers, J., Shaiganfar, R., and Wang, Y.: Absolute calibration of the colour index and O_4 absorption derived from Multi AXis (MAX-)DOAS measurements and their application to a standardised cloud classification algorithm, Atmospheric Measurement Techniques, 9, 4803–4823, https://doi.org/10.5194/amt-9-4803-2016, 2016.