Reviewer 2

This manuscript combines hourly PM2.5 observations from 36 cities in Northeast China during 2015–2025 with LightGBM modeling, SHAP analysis, and weather normalization to investigate long-term pollution changes. The long-term, multi-city coverage provides a useful regional perspective, particularly given the combined influences of industry, winter heating, and agricultural burning. I have the following comments.

Response: We thank this reviewer for your comments. We will respond to all comments point by point as shown below.

Comment

(1) Equation (12) defines PMmet as observed PM2.5 minus PMemi. This difference also includes the model residual described earlier, although residuals are stated to be used only for diagnostics. Please clarify this point and avoid attributing the entire difference to meteorology, particularly when discussing exceedance days.

Response: Thank you for this important comment. We agree that the original definition ($PM_{\mathrm{met}} = PM_{\mathrm{obs}} - PM_{\mathrm{emi}}$) contains not only the model-estimated meteorological modulation but also the model residual. We have therefore clarified the interpretation of (PM_{met}) in the revised manuscript. Specifically, (PM_{met}) is now described as the deviation of the observed $PM_{2.5}$ concentration from its weather-normalized baseline, rather than as a purely model-derived meteorological contribution. This clarification has been added in the Methods section. To explicitly evaluate the influence of the model residual, we further decomposed this term as $[PM_{\mathrm{met}} = (PM_{\mathrm{pred,actual}} - PM_{\mathrm{emi}}) + R]$ where ($PM_{\mathrm{pred,actual}}$) is the model prediction under the actual meteorological conditions and ($R = PM_{\mathrm{obs}} - PM_{\mathrm{pred,actual}}$) is the model residual. Thus, ($PM_{\mathrm{pred,actual}} - PM_{\mathrm{emi}}$) represents the residual-excluded, model-estimated meteorological modulation. This decomposition and the corresponding definitions have been added to the revised Methods section (Lines 355–369).

We additionally performed a residual-excluded sensitivity analysis. The annual regional estimates from the original (PM_{met}) series and the residual-excluded meteorological modulation were almost perfectly correlated (R=0.9998), with very similar long-term trends (−0.0777 versus −0.0782 μg m^{-3} yr^{-1}). The trend direction was unchanged in all 36 cities. Importantly, for all 229 regional exceedance days identified in the study, the residual-excluded meteorological modulation remained positive, consistent with the original (PM_{met})-based interpretation. Although the model residual was somewhat larger during exceedance episodes, its mean absolute magnitude

accounted for only 6.62% of the mean absolute (PM_{met}) during these events. These sensitivity results are summarized in Supplementary Table S5 and are described in the revised manuscript (Lines 355–369 and 606-617; see Table S5 for detailed sensitivity statistics).

Accordingly, we also revised the relevant terminology in the main text and figure descriptions so that (PM_{met}) is interpreted consistently as the deviation from the weather-normalized baseline rather than as a purely meteorological contribution.

These results demonstrate that the principal conclusion—that meteorological conditions amplified $PM_{2.5}$ concentrations during the identified regional exceedance episodes—is robust to the exclusion of the model residual.

(2) Please state whether minimum data-coverage requirements were applied to daily and annual means, and whether the set of monitoring stations remained consistent throughout the study. Regional means were calculated from all valid city-hour records, giving greater weight to cities with more valid hours. Please make this weighting explicit.

Response: Thank you for raising these important points. We have clarified the data-coverage treatment, temporal consistency of the monitoring network, and weighting used in the regional aggregation.

First, no additional minimum data-coverage threshold was imposed when calculating daily, monthly, or annual means. These aggregates were calculated from all valid hourly observations available within the corresponding period. The 50% coverage criterion was used only for the station-network diagnostic to determine whether a monitoring station was considered active in a given year, and was not used as an inclusion criterion for concentration averaging. We have clarified this point in Section 2.1, and annual $PM_{2.5}$ and CO data-completeness statistics are reported in Table S3 (Lines 145–161).

The monitoring network was not completely invariant during 2015–2025. For the station-network diagnostic, a station was considered active in a given year when valid hourly $PM_{2.5}$ observations were available for at least 50% of the source-available hours in that year. Based on this criterion, the annual number of active stations across the 36 cities ranged from 163 to 174, and 144 stations remained active throughout the entire study period. Section 2.1 and Table S2 have therefore been revised to explicitly describe temporal changes in station availability rather than assuming a fixed monitoring network (Lines 133–154; see Table S2).

To assess whether these changes affected the estimated long-term trend, we additionally performed a common-station sensitivity analysis using only the 144 stations that were active in every year from 2015 to 2025. The all-available-station and common-station series were compared using the same paired city-hour samples. Within this paired-sample sensitivity dataset, the two regional annual $PM_{2.5}$ series were nearly identical,

with R = 0.9999, and their linear trends were −1.758 and −1.755 μg m^{-3} yr^{-1}, respectively. City-level trend slopes were also strongly correlated, with R = 0.972, and 34 of the 36 cities retained the same trend direction. The only sign changes occurred in Jixi and YichunHLJ, where the estimated trends were weak and statistically non-significant under both station selections. These results indicate that temporal changes in station composition had little influence on the inferred regional trend.

Second, we agree that the original regional aggregation gives proportionally greater weight to cities with more valid hourly observations. We have now made this weighting explicit in Section 4.1. City-level monthly and annual values were calculated from all valid hourly observations within each city and period, whereas regional means were calculated by pooling all valid city-hour observations within the corresponding period. Consequently, cities with more valid hourly observations contributed proportionally more to the regional mean (Lines 419–431).

We further evaluated the influence of this weighting using an equal-city-weight sensitivity analysis. Annual mean $PM_{2.5}$, PM_{emi}, and PM_{met} were first calculated separately for each city and were then averaged across the 36 cities with equal weight. The annual regional series remained highly consistent with those obtained from the pooled-city-hour approach, with R = 0.9986 for $PM_{2.5}$, R = 0.9993 for PM_{emi}, and R = 0.9995 for PM_{met}. The $PM_{2.5}$ trend changed only from −1.734 to −1.742 μg m^{-3} yr^{-1}, the PM_{emi} trend from −1.656 to −1.661 μg m^{-3} yr^{-1}, and the PM_{met} trend from −0.078 to −0.081 μg m^{-3} yr^{-1}. The estimated contribution of the PM_{emi} trend to the observed $PM_{2.5}$ decline was also essentially unchanged, from 95.52% to 95.33%.

Overall, these sensitivity analyses indicate that neither temporal changes in monitoring-station composition nor the unequal weighting associated with differences in valid hourly coverage among cities materially affected the principal regional conclusions. We therefore retained the original aggregation procedures while explicitly clarifying the data-coverage treatment, monitoring-network variability, and regional weighting scheme in the revised manuscript (Lines 145–162 and 419–431).

(3) Weather conditions were sampled from all valid hours while the target time variables remained fixed. For example, a January record could be paired with weather observed in July, although its month variables still indicate January. Please explain this choice and discuss whether such unusual combinations could affect the weather-normalized estimates.

Response: Thank you for this important comment. The meteorological resampling strategy used in this study follows established meteorological-normalization approaches in which temporal predictors are retained while meteorological conditions are resampled to reduce the influence of meteorological variability on the inferred long-term pollution signal. Similar strategies have been applied in previous studies. For example, Vu et al. (2019) retained the temporal information of each target observation while resampling meteorological conditions, and Wu et al. (2022) evaluated

meteorological-normalization strategies based on resampled meteorological variables with temporal predictors retained. We have clarified this methodological rationale in the revised Methods section.

In our formal implementation, meteorological vectors were sampled from the full valid meteorological record of the corresponding city while the temporal predictors of the target hour were kept fixed. This approach places individual target hours under a common long-term meteorological reference distribution and thereby reduces the influence of interannual meteorological variability. We agree, however, that full-period sampling can generate time–meteorology combinations that are less typical of a particular season.

To evaluate whether such combinations materially affected the weather-normalized estimates, we performed an additional sensitivity analysis in which meteorological vectors were restricted to the same calendar month as the target hour across all study years. The trained city-specific models, target-hour temporal predictors, and 5,000 meteorological resamples were otherwise kept unchanged, so that the meteorological sampling window was the only methodological difference.

The two approaches produced highly consistent long-term results. The regional annual PM_{emi} series were strongly correlated, with $R = 0.974$. The estimated regional PM_{emi} trends were −1.683 and −1.668 μg m^{-3} yr^{-1} for the full-period and same-calendar-month sampling strategies, respectively, corresponding to an absolute difference of 0.016 μg m^{-3} yr^{-1}, or approximately 0.95%. In addition, all 36 cities retained the same PM_{emi} trend direction under the two resampling strategies.

These results indicate that although full-period meteorological sampling allows a broader range of time–meteorology combinations, it does not materially affect the inferred long-term PM_{emi} trend or its spatial trend pattern. We have clarified both the methodological rationale and the same-calendar-month sensitivity analysis in the revised Methods section (Lines 345–377 and 388–395).

Reference

Vu, T. V., Shi, Z., Cheng, J., Zhang, Q., He, K., Wang, S., and Harrison, R. M.: Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique, Atmos. Chem. Phys., 19, 11303–11314, 2019.

Wu, Q., Li, T., Zhang, S., Fu, J., Seyler, B. C., Zhou, Z., Deng, X., Wang, B., and Zhan, Y.: Evaluation of NOx emissions before, during, and after the COVID-19 lockdowns in China: A comparison of meteorological normalization methods, Atmos. Environ., 278, 119083, 2022.

(4) Please specify the CEDS release and clarify the calculation of the summed-species and mean-species indices in Figure 4b–c. Because these indices combine primary particles and gaseous precursors, “PM2.5-related emission indices” would be more accurate than “PM2.5 emission indices.” Please also note that the high correlations with PMemi may partly reflect shared downward trends.

Response: Thank you for this comment. We have added the CEDS release information and clarified the calculation of the two composite indices in the revised Methods section. The CEDS data for 2015–2020 were obtained from the CEDS v2021_04_21 dataset distributed through the GEOS-Chem/HEMCO v2021-06 archive, whereas the data for 2021–2023 were obtained from the 2025 release of CEDS (Lines 187–195).

For the summed-species index, the regional annual emissions of BC, OC, SO_2, NOx, and NH_3 were first summed, and the resulting total was then normalized to its 2015 value. For the mean-species index, each species was first normalized to its own 2015 value, and the five normalized species indices were then averaged with equal weight. Both indices were referenced to 2015 = 100. Following the reviewer's suggestion, we have also renamed these quantities as "$PM_{2.5}$-related emission indices" throughout the manuscript and Figure 4 to distinguish them from direct $PM_{2.5}$ emission inventories (Lines 196–212 and 587–597).

To examine whether the high correlations with PM_{emi} partly reflected shared long-term downward trends, we additionally performed a first-difference analysis, in which each annual value was replaced by its year-to-year change, $\Delta X_y = X_y - X_{y-1}$. This removes much of the shared long-term trend and provides a more direct comparison of year-to-year co-variability. The annual summed-species and mean-species indices were strongly correlated with PM_{emi}, with R = 0.925 and R = 0.882, respectively. After first differencing, however, the corresponding correlations were substantially weaker, with R = 0.347, p = 0.399 for the summed-species index and R = 0.292, p = 0.484 for the mean-species index. These results indicate that the agreement between PM_{emi} and the CEDS indices is stronger at the long-term scale than for individual year-to-year changes. The first-difference procedure and the corresponding interpretation have been added to the revised Methods and Results sections (Lines 212–215 and 530–545).

(5) Section 2.1 states that CO was used to assess PMemi. However, Figure 7c (all months) and Figure 7d (burning months) appear to show comparisons with observed PM2.5 rather than PMemi. Please confirm which series was used and make the stated purpose consistent with the analysis shown. A comparison with observed PM2.5 does not directly assess the weather-normalized component.

Response: Thank you for this comment. We agree that the original Figure 7c–d, which showed the relationship between observed $PM_{2.5}$ and CO, was not fully consistent with our stated use of CO as an independent combustion-related indicator for evaluating the weather-normalized component PM_{emi}.

We therefore recalculated the analysis using PM_{emi} directly. For all months during 2015–2025, regional monthly PM_{emi} was strongly correlated with CO, with R = 0.889, $p < 0.05$, and n = 132. A similarly strong relationship was observed during the main burning months of March, April, October, and November, with R = 0.791, $p < 0.05$, and n = 44. These results support the use of CO as an independent combustion-related indicator for evaluating variability in PM_{emi}.

Accordingly, we replaced the original $PM_{2.5}$–CO relationships in Figure 7c–d with PM_{emi}–CO relationships and revised the corresponding Results text, Discussion, Conclusion, and figure caption to ensure consistency throughout the manuscript (Lines 721–729, 735–745, 820–834, and 857–860; revised Figure 7).