Reviewer 3

In this study, Pan et al. used the machine learning algorithm LightGBM and meteorological normalization techniques to examine the PM2.5 trend across 36 cities in Northeast China from 2015 to 2025. The observed PM2.5 at 167 stations is decomposed into the meteorologically normalized component (PMemi) and the meteorological modulation term (PMmet), using a LightGBM model and the ERA5 meteorological dataset as predictors. Comprehensive comparisons are performed using the inventory dataset (CEDS emission indices), active fire detections (VIIRS), and CO observations, indicating the association of PMemi with anthropogenic activities in the region. The results of this study provide important information for regional air management and recommendations for future strategies of pollution reduction. Overall, this well-conducted study is within the scope of ACP, I recommend that this work to be published in the journal after addressing and clarifying several concerns in the methodology and results, as follows:

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

Major comment

(1) Line 255–256. Missing and infinite values should be transparently reported in the Supplementary document, and at least mention the percentage of missing values in the main text. Replacing missing and infinite values with the median of the training subset could lead to bias, especially for meteorological data (as it changes by season), please discuss the uncertainty of this analysis.

Response: Thank you for this important comment. We agree that the magnitude of missing data and the treatment of missing or infinite predictor values should be reported explicitly, particularly because median imputation could in principle introduce bias for meteorological variables with pronounced seasonal variability.

For the 36 cities during 2015–2025, the theoretical complete dataset contains 3,471,552 city-hour records. After city-level aggregation and temporal alignment, 3,152,473 valid $PM_{2.5}$ city-hour observations were available for model development, corresponding to an overall data coverage of 90.81% and an unavailable-data fraction of 9.19%. These 3,152,473 observations also constitute the full out-of-fold evaluation dataset. We have now explicitly reported these values in Section 2.1, and annual $PM_{2.5}$ and CO data-completeness statistics are provided in Supplementary Table S3 (Lines 150–166).

We have also clarified the treatment of missing data in Section 3.2. Records with invalid timestamps or missing $PM_{2.5}$ target observations were excluded before model fitting, and missing $PM_{2.5}$ target values were not imputed. The preprocessing pipeline was

configured so that, if missing or infinite predictor values occurred, they would be replaced using median values calculated exclusively from the corresponding training subset, with the same training-derived medians subsequently applied to the validation subset to avoid information leakage.

Because median imputation could potentially distort meteorological variables with strong seasonal variability, we further checked the occurrence of missing and infinite values in the formal modeling dataset. Across all 3,152,473 valid $PM_{2.5}$ city-hour samples, all 10 meteorological predictors were finite, corresponding to 0.00% missing or infinite values. Therefore, no meteorological predictor value actually required median imputation in the formal analysis. This diagnostic information has been reported in Supplementary Table S4 and clarified in Sections 3.2 and 3.4 of the revised manuscript (Lines 280–289 and 346–354).

Accordingly, although median imputation could introduce seasonal bias if it were applied to meteorological variables with substantial missingness, this source of uncertainty did not affect the formal analysis because the imputation procedure was never invoked for the meteorological predictors.

(2) Line 261–262. Please report the parameter range used in hyperparameter tuning and the performance of the best set of parameters.

Response: Thank you for this comment. We have clarified the hyperparameter optimization procedure and now report the parameter search ranges, the selected hyperparameter configuration, and the predictive performance obtained using that configuration.

Hyperparameter optimization was conducted independently for each of the 36 cities. For computational efficiency, up to 50,000 valid hourly samples were randomly selected for parameter optimization and divided into 80% training and 20% validation subsets using a fixed random seed. Eighteen candidate parameter combinations were evaluated. The search ranges were 400–1000 for n_estimators, 0.02–0.07 for learning_rate, 31–127 for num_leaves, 20–200 for min_child_samples, 0.7–1.0 for both subsample and colsample_bytree, 0–1.0 for reg_alpha, and 0–5.0 for reg_lambda, with max_depth fixed at −1. For each city, the parameter combination yielding the highest validation R^2 was selected.

Although optimization was conducted independently for each city, the same optimal parameter combination was selected for all 36 city-specific models: n_estimators = 1000, learning_rate = 0.07, num_leaves = 127, max_depth = −1, min_child_samples = 50, subsample = 1.0, colsample_bytree = 0.9, reg_alpha = 0.1, and reg_lambda = 0.1.

After hyperparameter selection, the selected configuration was kept fixed and model performance was evaluated using five-fold random cross-validation. Across the 36 cities, the out-of-fold R^2 values ranged from 0.648 to 0.838, with a mean of 0.790 and

a median of 0.794, and 35 of the 36 cities had R^2 values above 0.7. When all city-level out-of-fold predictions were pooled, the overall performance was R = 0.902, R^2= 0.803, RMSE = 17.221 μg m^{-3}, MAE = 9.647 μg m^{-3}, and MBE = −0.014 μg m^{-3}, based on 3,152,473 observations.

The hyperparameter-search procedure, parameter ranges, and selected parameter configuration have been clarified in Section 3.2 of the revised manuscript, and the corresponding out-of-fold model performance is reported in the Results section and Figure 2 (Lines 290–318 and 490–497).

(3) Line 326–328. Which sensitivity tests did the authors perform? What method was used to identify the stability of estimation and eventually support the conclusion that B should be above 3000?

Response: Thank you for this comment. We agree that the choice of B = 5000 should be supported by an explicit convergence analysis rather than stated without quantitative evidence.

We therefore performed an additional Monte Carlo convergence test using B = 100, 500, 1000, 2000, 3000, and 5000 with five independent random seeds. A fixed temporally stratified sample of 300 target hours was selected for each of the 36 cities, yielding 10,800 target hours for each seed. The same target hours were retained for all B values and seeds. For each seed, a single sequence of 5000 complete meteorological vectors was sampled with replacement from the full valid meteorological record of the corresponding city while the temporal predictors of each target hour were held fixed. Estimates at smaller B values were calculated from cumulative subsets of the same 5000-vector sequence. For each target hour, the mean of the five independent B = 5000 estimates was used as the convergence reference.

Convergence was assessed using the relative MAE with respect to this reference, the Pearson correlation with the reference estimates, and variability among independent random seeds. The relative MAE decreased from 3.652% at B = 100 to 1.523% at B = 500, 1.023% at B = 1000, 0.704% at B = 2000, and 0.535% at B = 3000. Increasing B further to 5000 reduced the relative MAE only modestly to 0.453%. Correspondingly, the correlation with the reference increased from R = 0.9941 at B = 100 to R = 0.99987 at B = 3000 and R = 0.99991 at B = 5000. The mean between-seed standard deviation also decreased from 1.697 μg m^{-3} at B = 100 to 0.253 μg m^{-3} at B = 3000 and 0.229 μg m^{-3} at B = 5000.

These results show that the weather-normalized estimates had largely converged by B = 3000, with only limited additional improvement between B = 3000 and B = 5000. We therefore retained B = 5000 as a conservative choice for the formal meteorological-normalization analysis. The convergence procedure and stability criteria have been clarified in Section 3.4 of the revised manuscript, and the detailed quantitative results are provided in Supplementary Table S6 and Figure S3 (Lines 378–387).

(4) 440–442. When considering the PMemi, the authors found one additional station showing an insignificant trend, indicating that the meteorological parameters significantly affect PM trend at this site. Please add 1 sentence indicating which station and the reason.

Response: Thank you for this comment. The city was Daqing. Its observed $PM_{2.5}$ concentration showed a significant decreasing trend of −1.142 μg m^{-3} yr^{-1} (p = 0.028), whereas the corresponding weather-normalized $PM_{2.5}$ (PM_{emi}) showed a weaker decreasing trend of −0.830 μg m^{-3} yr^{-1} (p = 0.128), which was no longer statistically significant.

To further examine this difference, we analyzed the annual time series of observed $PM_{2.5}$, PM_{emi}, and PM_{met} in Daqing (Figure R1). PM_{emi} declined substantially during the earlier part of the study period but showed a partial rebound after approximately 2020. In contrast, this rebound was much less evident in the observed $PM_{2.5}$ series because meteorological modulation was predominantly negative during the later years, particularly during 2022–2025. Thus, favorable meteorological modulation partly offset the increase in the weather-normalized pollution baseline and maintained lower observed $PM_{2.5}$ concentrations.

This explains why the observed $PM_{2.5}$ trend was steeper and statistically significant, whereas the decreasing PM_{emi} trend became weaker and non-significant after meteorological normalization. We therefore interpret Daqing as a city whose estimated long-term $PM_{2.5}$ trend was comparatively sensitive to interannual meteorological variability. The corresponding explanation has been added to the revised Results section (Lines 510–518).

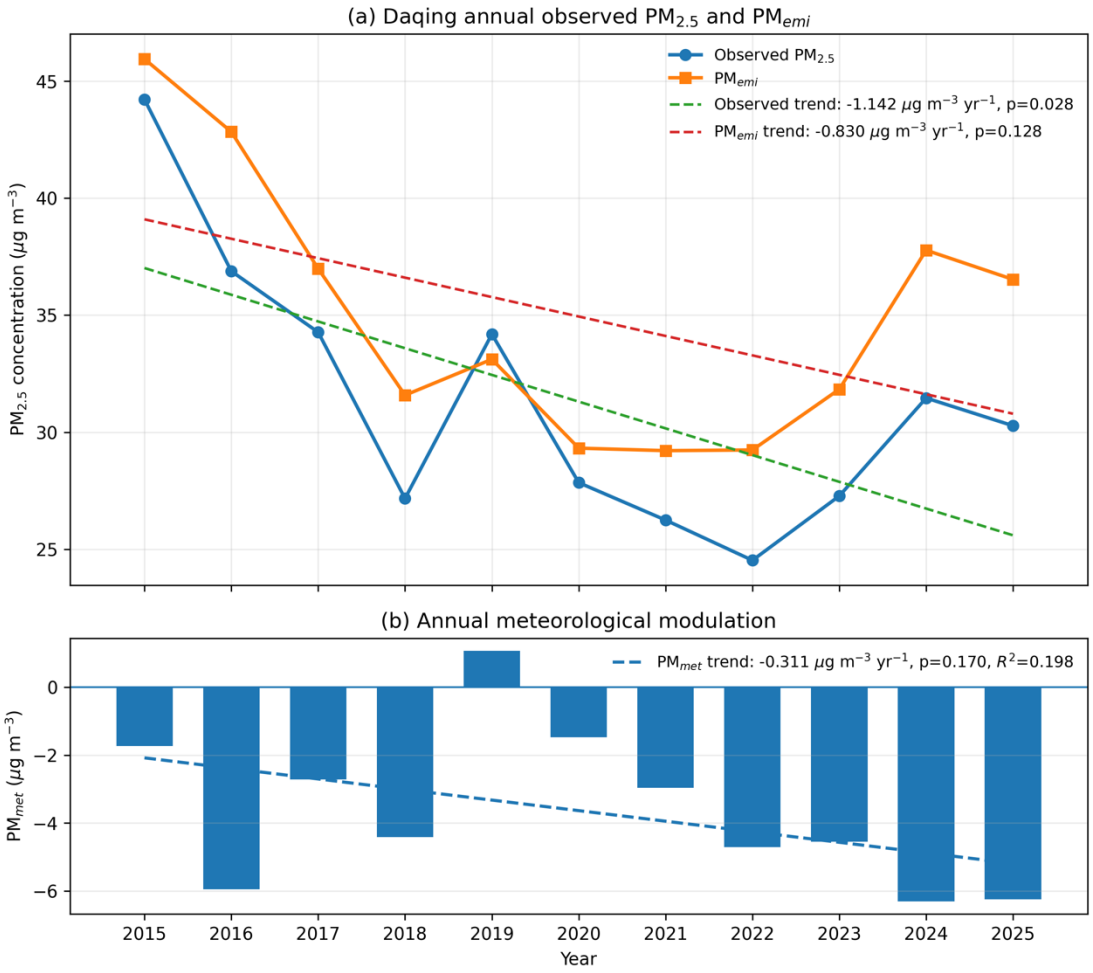


Figure R1. Annual observed $PM_{2.5}$, weather-normalized $PM_{2.5}$ (PM_{emi}), and meteorological modulation (PM_{met}) in Daqing during 2015–2025. Panel (a) shows the annual observed $PM_{2.5}$ and PM_{emi} series together with their linear trends. Panel (b) shows annual PM_{met}, defined as the deviation of observed $PM_{2.5}$ from the weather-normalized baseline. Negative PM_{met} values indicate meteorological conditions that reduced observed $PM_{2.5}$ relative to PM_{emi}.

(5) Line 454–456. Although the R value indicates good agreement between PMemi and CEDS, a significant difference is observed during 2018–2020. In this period, the PMemi exhibit a constant decrease, while the CEDS show stable emissions. Please discuss this difference during 2018–2020.

Response: Thank you for this comment. We have added a specific discussion of the divergence between PM_{emi} and the CEDS $PM_{2.5}$-related emission indices during 2018–2020. Over this period, PM_{emi} decreased by 3.48%, whereas the summed-species and mean-species CEDS indices changed by only −0.90% and −0.27%, respectively.
Several factors may contribute to this difference. First, bottom-up emission inventories are subject to uncertainties in activity data, emission factors, sectoral allocation, and

spatial distribution. Observation-based evaluations have identified substantial uncertainties in emission inventories over East Asia, including Northeast China, with discrepancies between inventory estimates and atmospheric observations (Choi et al., 2020). Observation-constrained evaluation has also suggested that CEDS may overestimate Chinese BC emissions, particularly those associated with the residential sector in winter (Ikeda et al., 2022).

Second, PM_{emi} represents the weather-normalized ambient $PM_{2.5}$ baseline rather than emission magnitude alone. Its interannual variability may also reflect changes in emission composition, source structure, secondary aerosol formation, and other non-meteorological processes. Exact year-to-year correspondence between PM_{emi} and the CEDS emission indices is therefore not necessarily expected.

Accordingly, we interpret the CEDS comparison as evidence of consistent long-term evolution rather than as direct validation of year-to-year PM_{emi} variability. The discussion of the 2018–2020 divergence and the corresponding references has been added to the revised Discussion section (Lines 747–769).

Reference

Choi, Y., Kanaya, Y., Park, S.-M., Matsuki, A., Sadanaga, Y., Kim, S.-W., Uno, I., Pan, X., Lee, M., Kim, H., and Jung, D. H.: Regional variability in black carbon and carbon monoxide ratio from long-term observations over East Asia: assessment of representativeness for black carbon (BC) and carbon monoxide (CO) emission inventories, Atmos. Chem. Phys., 20, 83–98, https://doi.org/10.5194/acp-20-83-2020, 2020. ACP

Ikeda, K., Tanimoto, H., Kanaya, Y., and Taketani, F.: Evaluation of anthropogenic emissions of black carbon from East Asia in six inventories: constraints from model simulations and surface observations on Fukue Island, Japan, Environ. Sci.: Atmos., 2, 416–427, https://doi.org/10.1039/D1EA00051A, 2022.

(6) In addition, the authors should be aware that even though SO2, NO and NH3 are anthropogenic emissions, they are still in the gaseous phase, and they need to undergo different processes and reactions to transform into the particulate phase. Comparing PM variability with the total change in BC, OC, SO2, NO, and NH3 directly may introduce bias. Consider performing the comparison separately for each chemical.

Response: Thank you for this important comment. We agree that BC and OC are primary particulate emissions, whereas SO_2, NOx, and NH_3 are gaseous precursors that undergo different atmospheric transformation processes before contributing to particulate matter. Following the reviewer's suggestion, we therefore performed species-specific comparisons between PM_{emi} and the annual CEDS emission indices of the five individual species.

PM_{emi} was significantly positively correlated with BC, with $R = 0.814$ and $p = 0.008$; OC, with $R = 0.795$ and $p = 0.010$; SO_2, with $R = 0.950$ and $p < 0.001$; and NOx, with $R = 0.820$ and $p = 0.007$. In contrast, NH_3 showed a significant negative correlation, with $R = -0.697$ and $p = 0.037$, indicating a contrasting long-term trajectory relative to PM_{emi} These species-specific comparisons have been added to the revised Results section (Lines 530–545).

We have also expanded the Discussion to address the different atmospheric behavior of these species. BC and OC are emitted primarily in the particulate phase, whereas SO_2, NOx, and NH_3 contribute to $PM_{2.5}$ through secondary sulfate, nitrate, and ammonium formation. Previous studies have shown that the responses of secondary inorganic aerosols to precursor-emission changes can differ substantially and may be nonlinear because of atmospheric oxidation, gas–particle partitioning, and interactions among multiple precursors (Geng et al., 2017; Li et al., 2021; Meng et al., 2022). Although meteorological normalization reduces the influence of meteorological variability on $PM_{2.5}$, the subsequent comparison with CEDS emissions remains statistical and does not explicitly resolve these precursor-to-particle transformation processes. This simplification therefore introduces uncertainty into the interpretation of relationships between precursor-emission changes and PM_{emi}.

Accordingly, the species-specific comparisons are interpreted as complementary evidence of long-term emission-related consistency rather than as a direct quantitative mapping between individual precursor emissions and ambient $PM_{2.5}$. The corresponding methodological clarification, species-specific results, and discussion have been added to the revised manuscript (Lines 196–200, 530–545, and 770–784).

Reference

Geng, G., Zhang, Q., Tong, D., Li, M., Zheng, Y., Wang, S., and He, K.: Chemical composition of ambient $PM_{2.5}$ over China and relationship to precursor emissions during 2005–2012, Atmos. Chem. Phys., 17, 9187–9203, https://doi.org/10.5194/acp-17-9187-2017, 2017.

Li, M., Zhang, Z., Yao, Q., Wang, T., Xie, M., Li, S., Zhuang, B., and Han, Y.: Nonlinear responses of particulate nitrate to NOx emission controls in the megalopolises of China, Atmos. Chem. Phys., 21, 15135–15152, https://doi.org/10.5194/acp-21-15135-2021, 2021. ACP

Meng, F., Zhang, Y., Kang, J., Heal, M. R., Reis, S., Wang, M., Liu, L., Wang, K., Yu, S., Li, P., Wei, J., Hou, Y., Zhang, Y., Liu, X., Cui, Z., Xu, W., and Zhang, F.: Trends in secondary inorganic aerosol pollution in China and its responses to emission controls of precursors in wintertime, Atmos. Chem. Phys., 22, 6291–6308, https://doi.org/10.5194/acp-22-6291-2022, 2022. ACP

(7) In this section, the authors primarily describe and discuss the results rather than interpret them in the context of the existing literature. In my opinion, presenting the results and discussion section without citations to previous studies makes it read more like a report than a scientific paper. I strongly recommend that the authors restructure this section to compare the study's findings with relevant research, especially those reported in China.

Response: Thank you for this important suggestion. We agree that the previous Discussion placed too much emphasis on restating our results and did not sufficiently interpret them in the context of previous studies.

We have substantially revised and reorganized the Discussion to strengthen comparisons with relevant studies, particularly those conducted in China. The long-term $PM_{2.5}$ and weather-normalization results are now compared with previous national and eastern-China attribution studies (e.g., Zhai et al., 2019; Vu et al., 2019; Xiao et al., 2021). We also expanded the discussion of precursor-specific relationships using previous studies of secondary aerosol chemistry in China, and compared our meteorological and biomass-burning results with earlier studies conducted specifically in Northeast China.

The revised Discussion now focuses more clearly on how our findings agree with, differ from, and extend existing knowledge, while also emphasizing the region-specific implications of winter heating, unfavorable meteorology, secondary aerosol formation, and agricultural biomass burning in Northeast China (Lines 746–834).

Minor comment

(1) Line 139. Please add the total number of PM2.5 samples.

Response: Thank you for the suggestion. We have added the total number of $PM_{2.5}$ samples to Section 2.1 of the revised manuscript. After city-level aggregation and temporal alignment, a total of 3,152,473 valid $PM_{2.5}$ city-hour observations were retained for model development. The corresponding annual and overall data-completeness statistics are also provided in Supplementary Table S3 (Lines 150-158).

(2) Line 181–183: Please provide the method to calculate the annual emission indices of BC, IC, SO2, NO, and NH3.

Response: Thank you for the suggestion. We have clarified the calculation of the annual CEDS emission indices in Section 2.3 of the revised manuscript. We note that "IC" in the comment corresponds to OC in our analysis, and the nitrogen oxide species used in this study is NOx.

Specifically, monthly regional emissions for each species were first obtained by spatially integrating the CEDS grid-cell emissions over the 36 study cities. For each

species, BC, OC, SO_2, NOx, and NH_3, the 12 monthly regional emissions within each calendar year were summed to obtain the annual regional emission, ($E_{s,y}$). The annual emission index was then calculated as ($I_{s,y} = \frac{E_{s,y}}{E_{s,2015}} \times 100$), with 2015 set to 100.

We also clarified the calculation of the two composite indices used in Figure 4. For the summed-species index, the annual regional emissions of the five species were first summed and the resulting total was then normalized to its 2015 value. For the mean-species index, each species was first normalized separately to its own 2015 value, and the five normalized indices were then averaged with equal weight (Lines 201–211).

(3) Line 270–271. Please indicate the number of iterations performed in cross-validation

Response: Thank you for the suggestion. We have clarified the number of iterations in the cross-validation procedure in Section 3.2 of the revised manuscript. After hyperparameter selection, five-fold random cross-validation was performed, corresponding to five iterations. In each iteration, four folds were used for model training and the remaining fold was used for validation. Thus, each valid sample served as a validation sample exactly once, and predictions from the five validation folds were combined to generate the complete out-of-fold prediction dataset (Lines 305–313).

(4) Line 302–303: How do the meteorological parameters' SHAP importance change across sites? Is temperature always the most important feature?

Response: Thank you for this comment. We examined the city-specific SHAP importance of the meteorological predictors and clarified this point in the revised Results section. Among the 10 meteorological predictors, 2 m air temperature was the most spatially consistent variable, ranking first in all 36 cities. In contrast, the relative importance of the remaining meteorological predictors varied among cities, indicating spatial heterogeneity in the secondary meteorological influences represented by the city-specific models (Lines 644–653).

(5) Line 362–363. Please add the definition of εy.

Response: Thank you for pointing this out. We have added the definition of ε_y in Section 4.1 of the revised manuscript. It represents the regression residual for year y in the ordinary least-squares trend analysis (Lines 429–431).

(6) Line 394–403 and 440–442. The regression slope and the p-value may be insufficient to show the trend. Please add the coefficient of determination of the regression.

Response: Thank you for this suggestion. We have added the coefficient of determination, R^2, for the city-level linear trend regressions in both sections of the

revised manuscript. For observed $PM_{2.5,}$ the median R^2 across the 36 city-level regressions was 0.664 (Lines 465-466). For PM_{emi}, the corresponding median R^2 was 0.669 (Lines 520–521).

(7) Figure 1b. Please adjust the color bar of the PM2.5 trend (the range may be −4 to 2 μg m−3 year−1), the downward trend will be more visible.

Response: Thank you for this suggestion. We have adjusted the color-bar range in Figure 1b to −4 to 2 μg m^{-3} yr^{-1} to better visualize the spatial differences in city-level $PM_{2.5}$ trends (Figure 1b).