

Dear Editor,

We thank reviewers for their constructive comments which have greatly improved our manuscript. We have now addressed all comments reviewers raised.

Referee: 2

This is a very interesting study estimating the contribution of regional transport to PM_{2.5} in Beijing. This analysis can support policy-makers in both validating and designing effective policies. In addition, the methodology can be potentially applied in other regions as well, unraveling the contribution of local emissions and regional pollution transport. Nevertheless, there are some points that need further information and clarifications. If these points are addressed, then I would be happy to suggest publication in GMD.

We are thankful for the generally positive comments on our work from the reviewer.

Main comments

- I. A better description of the XGBoost model goal is needed. Is there a separate XGBoost model for ambient and local? Which exactly is the use of term ambient in this study? Please explain in more detail what the difference (in train data) is in building these two models. Which are the features and which are the targets in each case? One should clearly understand the transition from HYSPLIT CWT analysis to the XGBoost models. Which information from the HYSPLIT and CWT analysis are used and in which way to build the XGBoost models, and for which specific reasons (goals).

We thank reviewer to point this out. These are now added in the revision:

Line 123-133: “In this study, we combined CWT analysis with the HYSPLIT trajectory ensemble to obtain hourly-resolution PM_{2.5} source results and used this approach to distinguish between local emissions and regional transport. Solved the problems of traditional CWT methods being unable to obtain hourly time accuracy and models such as NAME consuming a large number of computational resources. Predictive XGBoost models were developed for Beijing using meteorological data and time variables to explain PM_{2.5} concentrations. By training the XGBoost model with PM_{2.5} dominated by local emissions, which are separately distinguished by CWT, and generalizing the findings to all study periods, the concentration of locally emitted PM_{2.5} (local) can be obtained. Similarly, ambient observed PM_{2.5} (ambient) can be determined by training the XGBoost model with ambient PM_{2.5} data. The contribution of regional transport to PM_{2.5} in Beijing can be quantified by comparing the ambient and local PM_{2.5} concentrations.”

II. The year 2020 is used in the analysis and to validate the XGBoost model training. Yet, 2020 is a “special” year due to COVID-19 and the associated restriction measures having direct effect on emissions and air pollution levels. How do COVID-19 restriction measures affect the analysis, and the conclusions raised for the Action Plan? This is something that needs to be well clarified.

We thank reviewer to point this out. In this study, the r^2 for ambient $PM_{2.5}$ and local emissions reached 0.78 and 0.74, respectively. The relatively low r^2 values may be attributed to the extremely low human activities under the same meteorological conditions, which led to a decrease in $PM_{2.5}$ concentration, making it challenging for XGBoost to learn better results. As shown in Figure 2b1 and b2, the significant differences between XGBoost learning results and actual observations are mainly concentrated in the low $PM_{2.5}$ concentration stage. To address this, we have added the following description in the article:

Line 228-232: “The error between XGBoost learning results and actual observed $PM_{2.5}$ values is mainly concentrated in the low concentration stage. This may be attributed to the significant reduction in human activities during the COVID-19 lockdown periods, which led to a decrease in actual $PM_{2.5}$ levels, making it challenging for XGBoost to learn (Fig. 2b1 and b2).”

Line 571-575: “

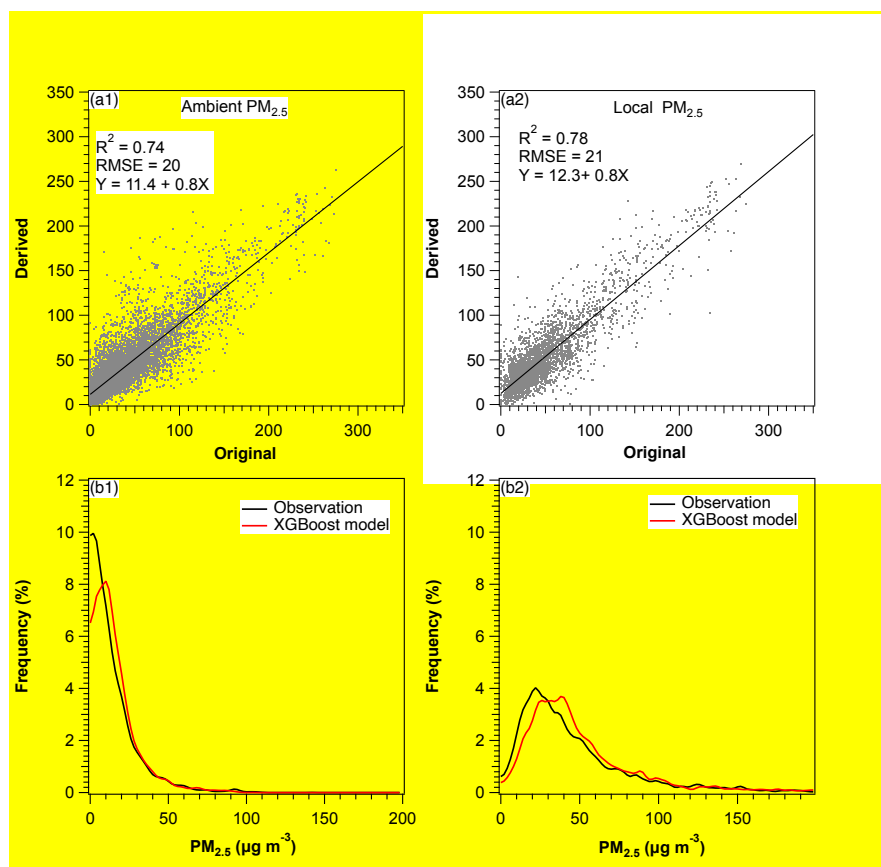


Fig. 2. Comparison of XGBoost model estimates and observations for (a1) ambient $PM_{2.5}$ and (a2) local $PM_{2.5}$ using testing samples from 2020. Frequency distributions of $PM_{2.5}$ observations

(black lines) and XGBoost model predictions (red lines) for (b1) ambient PM_{2.5} and (b2) local PM_{2.5} using testing samples from 2020.”

Furthermore, the inconsistent lockdown times of various cities during the COVID-19 pandemic have impacted the calculation of ambient and local PM_{2.5} during 2020. To clarify this, we have added the following explanation:

Line 199-200: “Note that the 2020 analysis results may contain some uncertainties due to the impact of COVID-19.”

Line 312-315: “The highest proportion of regional transport events from the west occurred in 2016, reaching 68%, while the highest proportion of southward transport-induced pollution events occurred in 2017 (with the exception of 2020, which may have been influenced by the COVID-19 pandemic).”

Line 325-328: “However, after the implementation of the Action Plan, the proportion of transport-induced pollution events from the south region gradually decreased to 38%. In 2020, this proportion further declined to 25%, but this may have been affected by the COVID-19 pandemic.”

- III. Do you apply any “feature selection” based on feature importance? What is the rationale for using both month of the year and day of the year? How is overfitting prevented in the XGBoost model training? Several studies (e.g. Akritidis et al., 2021; Zhang et al., 2020) applied an early stopping technique to prevent overfitting. Is there something similar applied here? Please explain and discuss in the manuscript accordingly.

The selection of feature importance parameters in our article primarily referred to the work by Xu et al., published in Nature, who have established mature learnable feature importance parameters for PM_{2.5} (Xu et al., 2023). The main goal of our study is to obtain machine learning PM_{2.5} results that match the observed values; therefore, less attention was paid to the choice of feature importance parameters. Additionally, we have added the contribution of feature importance in the revised manuscript, as follows:

Line 196-197: “Based on the XGBoost learning results, the most sensitive parameters for both local and ambient PM_{2.5} are RH, wind field, surface pressure and PBLH (Fig. S1).”

Line 30-35 in supplement:

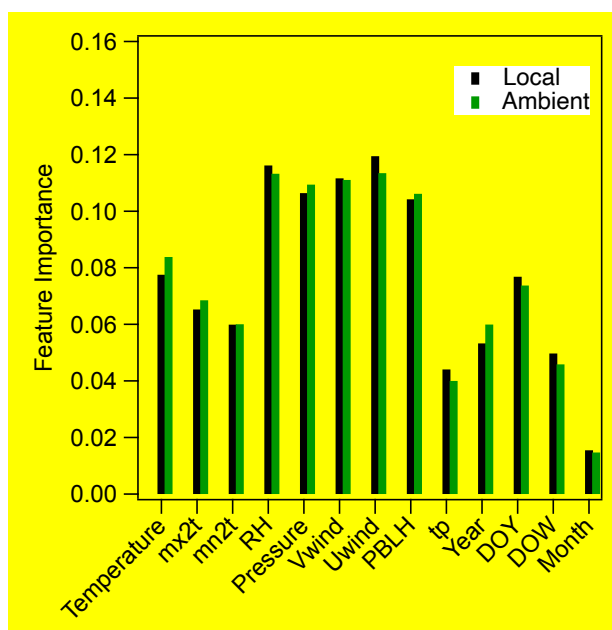


Fig. S1. Feature importance of the XGBoost models in estimating local (black) and ambient (green) PM_{2.5} using the training data. The considered features include temperature, 2-m maximum (mx2t), 2-m minimum temperature (mn2t), relative humidity (RH), surface pressure, V-wind, U-wind, planetary boundary layer height (PBLH), total precipitation (tp), year, day of year, day of week, and month.”

The 'month' feature can explain monthly variation characteristics. However, there is a jump in month values between 1 and 12, which may lead to inconsistencies when directly using 'month' as a numerical feature, affecting the model's learning of periodic patterns. The 'day of year' feature can capture the characteristics of certain small-scale days, such as those before and after the Spring Festival or during a specific summer vacation period. When using both 'month' and 'day of year' features to solve scenarios with significant seasonality (monthly/quarterly changes) and differences within a month, the model can capture cross-month trends and subtle differences in specific dates within the month.

Thanks for pointing out the missing parts in our article. We have added the overfitting prevention methods used during the training process of the XGBoost model, as follows:

Line 212-214: “Hyperparameter optimization and performance evaluation of the model were conducted using fivefold cross-validation (CV), while early stopping with a patience of 10 rounds was employed to prevent overfitting (Akritidis et al., 2021; Zhang et al., 2020).”

Comments

1. Line 53: I believe dust deserves to be included compared to tsunamis and volcanic eruptions.

We thank reviewer to point this out. This is now added in the revision:

Line 52-55: “Ambient fine particulate matter (PM_{2.5}, with particle aerodynamic diameter ≤ 2.5 μm) is influenced by both natural sources, such as dust, volcanic eruptions, tsunamis, and forest fires, and anthropogenic emissions, including fuel combustion, transportation, and industrial production.”

2. Line 59: The more recent studies by Smith et al. (2020) and Kalisoras et al. (2024) based on CMIP6 ESMs can be also included here (see details in References).

This is now added in the revision:

Line 56-59: “Numerous epidemiological studies have found that PM_{2.5} can significantly damage human health by exacerbating respiratory and cardiovascular diseases (Bartell et al., 2013; Brauer et al., 2012; Pascal et al., 2014), and also has an impact on weather and climate change (Wang et al., 2014; Smith et al., 2020; Kalisoras et al., 2023).”

3. Lines 60-61: The study by Geng et al. (2021) can be also cited here.

This is now added in the revision:

Line 59-61: “China's rapid and energy-intensive development over the past several decades has led to severe air pollution and negative public health impacts (Huang et al., 2014; Geng et al., 2021).”

4. Lines 68-69: For this statement a reference is required.

We thank reviewer to point this out. This is now added in the revision:

Line 66-70: “To address this issue, the Chinese government implemented the Air Pollution Prevention and Control Action Plan (denoted "Action Plan") from 2013 to 2017 and the Blue Sky Protection Campaign from 2018 to 2020, which effectively controlled anthropogenic emissions and reduced ambient PM_{2.5} concentrations (Zhang et al., 2019; Du et al., 2022).”

5. Line 72: Change to Wu et al. (2021) and apply accordingly where applicable in the manuscript.

We thank reviewer to point this out. This is now added in the revision:

Line 73-76: “Wu et al. used the HYSPLIT model to simulate the 24-hour backward trajectory in Zhoushan (Wu et al., 2021), and identified continental air masses that spent more than 5% of the previous 24 hours over the continent region, while the remaining air masses were identified as oceanic-influenced air masses.”

Furthermore, we have corrected the corresponding errors throughout the manuscript.

6. Lines 116-117: “meteorological data, synoptic scale, planetary boundary layer height (PBLH),” What do you mean by synoptic scale? PBLH can be considered a meteorological parameter as well. Please rephrase.

We thank reviewer to point this out. This is now added in the revision:

Line 117-119: “For instance, Grange et al. used meteorological data, synoptic scale weather patterns, and time variables to explain daily PM₁₀ concentrations in Switzerland (Grange et al., 2018).”

7. I suggest listing the selected hyperparameters for each XGBoost model to facilitate reproduction if needed.

We thank reviewer to point this out. This is now added in the revision:

Line 185-189: “The hyperparameters used in the model for local (ambient) conditions include a maximum number of boosting iterations of 6067 (13421), a learning rate of 0.1, a maximum tree depth of 7 (11), a minimum sum of instance weight needed in a child of 5 (3), a subsampling ratio of 0.8 (0.6) for training instances, and a subsampling ratio of 0.8 for columns when constructing each tree.”

8. Lines 131-132: A url and/or reference for the PM_{2.5} observations is needed here.

We have added the PM_{2.5} data to the Code and Data Availability section.

Line 368-375: “The codes used in this study are archived on Zenodo: the machine learning code at <https://doi.org/10.5281/zenodo.14677125>, the CWT code at <https://doi.org/10.5281/zenodo.13994400>, ECMWF data at <https://doi.org/10.5281/zenodo.14353871>, GDAS data at <https://doi.org/10.5281/zenodo.14347277>, HySplit Trajectory Ensemble at <https://doi.org/10.5281/zenodo.14375567>, and PySPLIT at <https://doi.org/10.5281/zenodo.14354765>. The meteorology and PM_{2.5} data used in this study can be accessed at <https://dx.doi.org/10.17632/bhftx3kz8.2>.”

9. Line 136: A reference is needed here for ERA5 data set.

We thank reviewer to point this out. This is now added in the revision:

Line 140-143: “Meteorological data, including temperature, relative humidity, pressure, precipitation, wind speed, and planetary boundary layer height (PBLH), were sourced from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 hourly reanalysis dataset (<https://cds.climate.copernicus.eu/datasets>).”

10. Lines 148-160: Equation 2 is referred first. Please either refer to equation 1 first or change the order of equations.

We thank reviewer to point this out. We have made the following modifications:

Line 156-171: “In the CWT analysis method, each grid point is assigned a weight, and the contribution of each grid point to the pollutant concentration at the target site is calculated using the air mass residence time and pollutant concentration (Hopke et al., 1993; Polissar et al., 1999; Xu and Akhtar, 2010) (equation 1). The grid point resolution was set to $0.25^\circ \times 0.25^\circ$ for this study. In equations 1, C_{ij} is the average weighted concentration at grid point (i, j) , l is the trajectory index, M represents the total number of trajectories, C_l is the $PM_{2.5}$ concentration corresponding to the target site, and τ_{ijl} is the residence time of trajectory l passing through the grid point. In calculation, the number of trajectories falling on each grid point is used instead of the residence time.

$$C_{ij} = \frac{\sum_{l=1}^M C_l \times \tau_{ijl}}{\sum_{l=1}^M \tau_{ijl}} \quad (1)$$

To reduce the effect of small values of n_{ij} , the CWT values were multiplied by an arbitrary weight function $W(n_{ij})$ to better reflect the uncertainty in the values for these grids (equation 2).

$$W(n_{ij}) = \begin{cases} 1.00, & 3n_{ave} < n_{ij} \\ 0.70, & 1.5n_{ave} < n_{ij} \leq 3n_{ave} \\ 0.4, & n_{ave} < n_{ij} \leq 1.5n_{ave} \\ 0.17, & n_{ij} \leq n_{ave} \end{cases} \quad (2)$$

where n_{ij} represents the number of trajectories that fall within the grid point, and n_{ave} represents the average number of trajectories passing through each grid point.”

11. Lines 154-156: I am a bit confused here. First you say “ τ_{ijl} is the residence time of trajectory l passing through the grid point” and then “In calculation, the number of trajectories falling on each grid point is used instead of the residence time”. The residence time is calculated for each trajectory, but then how can a residence time for a trajectory be calculated from a number of trajectories? I may miss something here, please clarify.

The calculation process only uses the number of times a trajectory passes through a specific grid point instead of counting the trajectory's residence time within that grid point.

12. Lines 162-166: Which is the rationale behind the definition of the regions? I think a small sentence is needed.

We thank reviewer to point this out. This is now added in the revision:

Line 172-173: "The potential source contribution to PM_{2.5} at the target site was investigated by categorizing the backward air masses into five different source regions centered around Beijing."

13. Lines 168-170: So, if the contribution from the north sector is 41% and the one from local is 40% then is classified as regional from the north sector?

We thank reviewer to point this out. This study considers the dominant air mass to be the main incoming air mass. This method has been applied in previous studies (Streets et al., 2007; Tian et al., 2019; Liu et al., 2020). However, these studies only consider westward air masses as regional transport processes. To explore the pollution and cleaning effects of regional transport, we have improved the classification method by using the dominant air mass. We have made the following modifications to the corresponding parts of the article:

Line 179-182: "The region with the highest contribution is used to determine the dominant source of air masses in Beijing at each time, classifying the overall air mass sources into local emissions (Fig. 1g) and regional transport (Fig. 1h). It is important to note that local emission periods were also influenced by persistent regional transport, and vice versa."

14. Figure 5: Just for clarification, summing the individual histograms over the years will result in 100%? In some cases, the pie charts sum is not 100%. I assume this is related to the rounding of percentages.

We have verified the summation results. For sums that deviate from 100%, the error values are consistently 1% due to rounding.

References

Akritidis, Dimitris, et al. "Implications of COVID-19 restriction measures in urban air quality of Thessaloniki, Greece: A machine learning approach." *Atmosphere* 12.11 (2021): 1500

Geng, Guannan, et al. "Drivers of PM_{2.5} air pollution deaths in China 2002–2017." *Nature Geoscience* 14.9 (2021): 645-650.

Kalisoras, Alkiviadis, et al. "Decomposing the effective radiative forcing of anthropogenic aerosols based on CMIP6 Earth system models." *Atmospheric Chemistry and Physics* 24.13 (2024): 7837-7872.

Smith, Christopher J., et al. "Effective radiative forcing and adjustments in CMIP6 models." *Atmospheric Chemistry and Physics* 20.16 (2020): 9591-9618

Zhang, Qiang, et al. "A PM2. 5 concentration prediction model based on multi-task deep learning for intensive air quality monitoring stations." *Journal of Cleaner Production* 275 (2020): 122722

References

- Akritidis, D., Zanis, P., Georgoulas, A. K., Papakosta, E., Tzoumaka, P., and Kelessis, A.: Implications of COVID-19 restriction measures in urban air quality of Thessaloniki, Greece: A machine learning approach, *Atmosphere*, 12, 1500, 2021.
- Bartell, S. M., Longhurst, J., Tjoa, T., Sioutas, C., and Delfino, R. J.: Particulate air pollution, ambulatory heart rate variability, and cardiac arrhythmia in retirement community residents with coronary artery disease, *Environmental health perspectives*, 121, 1135-1141, 2013.
- Brauer, M., Amann, M., Burnett, R. T., Cohen, A., Dentener, F., Ezzati, M., Henderson, S. B., Krzyzanowski, M., Martin, R. V., and Van Dingenen, R.: Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution, *Environmental science & technology*, 46, 652-660, 2012.
- Du, H., Li, J., Wang, Z., Chen, X., Yang, W., Sun, Y., Xin, J., Pan, X., Wang, W., and Ye, Q.: Assessment of the effect of meteorological and emission variations on winter PM_{2.5} over the North China Plain in the three-year action plan against air pollution in 2018–2020, *Atmospheric Research*, 280, 106395, 2022.
- Geng, G., Xiao, Q., Liu, S., Liu, X., Cheng, J., Zheng, Y., Xue, T., Tong, D., Zheng, B., and Peng, Y.: Tracking air pollution in China: near real-time PM_{2.5} retrievals from multisource data fusion, *Environmental Science & Technology*, 55, 12106-12115, 2021.
- Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., and Hueglin, C.: Random forest meteorological normalisation models for Swiss PM₁₀ trend analysis, *Atmospheric Chemistry and Physics*, 18, 6223-6239, 2018.
- Hopke, P. K., Gao, N., and Cheng, M.-D.: Combining chemical and meteorological data to infer source areas of airborne pollutants, *Chemometrics and Intelligent Laboratory Systems*, 19, 187-199, 1993.
- Huang, R.-J., Zhang, Y., Bozzetti, C., Ho, K.-F., Cao, J.-J., Han, Y., Daellenbach, K. R., Slowik, J. G., Platt, S. M., and Canonaco, F.: High secondary aerosol contribution to particulate pollution during haze events in China, *Nature*, 514, 218-222, 2014.
- Kalisoras, A., Georgoulas, A. K., Akritidis, D., Allen, R. J., Naik, V., Kuo, C., Szopa, S., Nabat, P., Olivie, D., and Van Noije, T.: Decomposing the Effective Radiative Forcing of anthropogenic aerosols based on CMIP6 Earth System Models, *Atmospheric Chemistry & Physics Discussions*, 2023.
- Liu, D., Hu, K., Zhao, D., Ding, S., Wu, Y., Zhou, C., Yu, C., Tian, P., Liu, Q., and Bi, K.: Efficient vertical transport of black carbon in the planetary boundary layer, *Geophysical Research Letters*, 47, e2020GL088858, 2020.
- Pascal, M., Falq, G., Wagner, V., Chatignoux, E., Corso, M., Blanchard, M., Host, S., Pascal, L., and Larrieu, S.: Short-term impacts of particulate matter (PM₁₀, PM_{10-2.5}, PM_{2.5}) on mortality in nine French cities, *Atmospheric Environment*, 95, 175-184, 2014.
- Polissar, A., Hopke, P., Paatero, P., Kaufmann, Y., Hall, D., Bodhaine, B., Dutton, E., and Harris, J.: The aerosol at Barrow, Alaska: long-term trends and source locations, *Atmospheric Environment*, 33, 2441-2458, 1999.
- Smith, C. J., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., Boucher, O., Dufresne, J.-L., Nabat, P., and Michou, M.: Effective radiative forcing and adjustments in CMIP6 models, *Atmospheric Chemistry and Physics*, 20, 9591-9618, 2020.

Streets, D. G., Fu, J. S., Jang, C. J., Hao, J., He, K., Tang, X., Zhang, Y., Wang, Z., Li, Z., and Zhang, Q.: Air quality during the 2008 Beijing Olympic games, *Atmospheric environment*, 41, 480-492, 2007.

Tian, P., Liu, D., Huang, M., Liu, Q., Zhao, D., Ran, L., Deng, Z., Wu, Y., Fu, S., and Bi, K.: The evolution of an aerosol event observed from aircraft in Beijing: An insight into regional pollution transport, *Atmospheric Environment*, 206, 11-20, 2019.

Wang, Y., Wang, M., Zhang, R., Ghan, S. J., Lin, Y., Hu, J., Pan, B., Levy, M., Jiang, J. H., and Molina, M. J.: Assessing the effects of anthropogenic aerosols on Pacific storm track using a multiscale global climate model, *Proceedings of the National Academy of Sciences*, 111, 6894-6899, 2014.

Wu, Y., Liu, D., Wang, X., Li, S., Zhang, J., Qiu, H., Ding, S., Hu, K., Li, W., and Tian, P.: Ambient marine shipping emissions determined by vessel operation mode along the East China Sea, *Science of The Total Environment*, 769, 144713, 2021.

Xu, R., Ye, T., Yue, X., Yang, Z., Yu, W., Zhang, Y., Bell, M. L., Morawska, L., Yu, P., and Zhang, Y.: Global population exposure to landscape fire air pollution from 2000 to 2019, *Nature*, 621, 521-529, 2023.

Xu, X. and Akhtar, U.: Identification of potential regional sources of atmospheric total gaseous mercury in Windsor, Ontario, Canada using hybrid receptor modeling, *Atmospheric Chemistry and Physics*, 10, 7073-7083, 2010.

Zhang, Q., Wu, S., Wang, X., Sun, B., and Liu, H.: A PM_{2.5} concentration prediction model based on multi-task deep learning for intensive air quality monitoring stations, *Journal of Cleaner Production*, 275, 122722, 2020.

Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., and Liu, W.: Drivers of improved PM_{2.5} air quality in China from 2013 to 2017, *Proceedings of the National Academy of Sciences*, 116, 24463-24469, 2019.