Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., author comment AC2
https://doi.org/10.5194/acp-2021-28-AC2, 2021

Reply on RC2

Qingyang Xiao et al.

Author comment on "Separating emission and meteorological contributions to long-term $PM_{2.5}$ trends over eastern China during 2000–2018" by Qingyang Xiao et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-28-AC2, 2021

This paper separates the contributions of emissions and meteorology to PM2.5 trends in different regions of China for 2000-2018 by reconstructing the PM2.5 record back to 2000 using satellite AOD data and a machine-learning approach including information from $PM_{2.5}$ data,WRF-CMAQ model results, and meteorological variables. This is a remarkable piece of work and the first such analysis to my knowledge that goes back to 2000, thus providing a long-term perspective on the role of meteorology and enabling a better understanding of the relation of PM2.5 trends to anthropogenic emissions. This includes better definition of the 2000s maximum. The analysis seems carefully done and the interpretation is insightful. A particularly interesting result, as the authors point out, is that interannual meteorological variability can play an important role in driving PM2.5 trends over the 3-5 year horizon of government action plans. I support publication but suggest a few editorial revisions.

Response: We thank the referee for the positive tone and the valuable suggestions to improve our manuscript.

- The writing is in general very good but there are recurring problems with the tense form of verbs. For example, the last sentence of the abstract should read 'is severe…is clustered'. At various points in the paper, 'meteorology-associate' should be 'meteorology-associated'. The authors should check throughout.

Response: We reviewed and corrected the verbs and other grammar errors.

- I found Figure 1 to be incomprehensible and Section 2.1 riddled with machine-learning jargon, and this initially discouraged me from the paper. One way to fix Figure 1 would be with a detailed caption describing the different elements of the Figure. Section 2.1.2 and other portions of the text should be edited for a readership not steeped in machine-learning packages.

Response: We added the following explanation in the caption of Figure 1,"The green process shows the two methods that separating emission and meteorology contributions to $PM_{2.5}$ in this study. The first method assesses the meteorology-associated $PM_{2.5}$ from WRF/CMAQ simulations with the fixed emissions at the 2000 level and varying meteorological inputs. The second method assesses the meteorology-associated $PM_{2.5}$ with satellite-based $PM_{2.5}$ estimations and a generalized additive model (GAM). The processing of satellite-based $PM_{2.5}$ estimation includes two stages. In stage 1 (blue), we constructed a

measurement-based high-pollution indicator and trained an extreme gradient boosting (XGB) model to predict the high-pollution indicator. In stage 2 (yellow), we trained a XGB model to predict the residuals of WRF/CMAQ simulations with high-pollution indicator as well as satellite AOD, meteorology and land use data as predictors."

We also edited section 2.1.2 to make it easy to follow: "A two-stage prediction model was developed to estimate $PM_{2.5}$ concentrations over China (Fig. 1). The first-stage model described high-pollution events that were underestimated in previous models and the second-stage model predicted residuals of CMAQ $PM_{2.5}$ simulations with the estimated high-pollution indicator from the first-stage model.

Since high-pollution events relatively rarely occur in the model training dataset, models may not appropriately characterize the associations between high $PM_{2.5}$ concentrations and predictors, leading to underestimation of high-pollution levels (Wei et al., 2020). We first defined a high-pollution indicator, describing whether the daily $PM_{2.5}$ observation was higher than the monthly average $PM_{2.5}$ concentration plus two standard deviations at each location. We noticed that only 3.9% of the daily data were assigned as high-pollution. To balance high-pollution samples and normal samples, the synthetic minority oversampling technique (SMOTE) (Torgo, 2010) that improved classifiers' performance in previous studies (Ghorbani and Ghousi, 2020;Saputra and Suharjito, 2019) was applied. The SMOTE algorithm oversampled the high-pollution data (the minority) by artificially generated new synthetic samples along the line between the high-pollution data and their selected nearest neighbors (Chawla et al., 2002;Chawla et al., 2003). This method also under-sampled the normal data (the majority) to better balance the model training dataset. After SMOTE resampling, high-pollution data accounted for 23% in the new model training dataset.

The balanced model training dataset was adopted to train the first-stage extreme gradient boosting (XGBoost) model that built the relationship between the high-pollution indicator and all the predictors, excluding CMAQ simulations. The predicted high-pollution indicator from the first-stage model was passed to the second-stage model as a predictor. We adopted the residual between the $PM_{2.5}$ measurement and the CMAQ $PM_{2.5}$ simulation as the dependent variable to train the second-stage model, thus enhances the response of predictors to $PM_{2.5}$ variations and improved the prediction accuracy."

- Lines 260-263: it would be worth citing other papers that projected the effect of climate change in BTH, particularly since they did not agree: (1) Cai, W., Li, K., Liao, H., Wang, H., and Wu, L.: Weather conditions conducive to Beijing severe haze more frequent under climate change, Nat. Clim. Change, 7, 257–262, 2017; (2) Shen, L., D.J. Jacob, L.J. Mickley, Y. Wang, and Q. Zhang, Insignificant effect of climate change on winter haze pollution in Beijing, Atmos. Chem. Phys., 18, 17489-17496, 2018. Can the current work arbitrate based on the 20-year record? Probably not but it would be worth some comment.

Response: Thank you for your suggestion. We added the following discussion of these previous studies in line 274-282 "In the context of global warming, the unfavorable meteorological conditions in the northern part of China could be worsen in the future, although previous studies on the projection of the future effects of climate change on air pollution showed inconsistent results. For example, Cai et al. (2017) projected increased frequency and persistence of haze events in Beijing in the future (2050-2099) and Shen et al. (2018) found statistically insignificant trend of haze index in the future in Beijing. In contrast,

in the southern part of China, especially in the YRD and surrounding regions, the estimated meteorological conditions were improving and were beneficial to pollution control (Chen et al., 2019). Further studies are needed to better understand the long-term trend of meteorological and climate effects on air pollution across China. Stricter clean air actions are preferred to avoid haze events in the future, considering the considerable meteorological effects on air pollution."

- Line 290: the same north-south contrast in the association of PM2.5 with RH was found by Zhai et al. (2019), previously cited but worth citing here, because they explained this contrast differently in terms of the origins of high-RH air masses and the links to aqueous chemical production and deposition.

Response: We added the following discussion and the suggested citation in line 309-312 "Zhai et al. (2019) also discussed the north-south contrast in the $PM_{2.5}$-humidity associations and indicated that the positive effects of humidity on $PM_{2.5}$ in the north were partly attributed to the favorable role of aqueous-phase aerosol chemistry in secondary $PM_{2.5}$ formation and the negative $PM_{2.5}$-humidity associations in the south were partly attributed to the precipitation related wet deposition."

- Line 325: I think 'interannual' should be 'long-term trends'

Response: We changed this word as suggested.

- Lines 361-362: I don't understand 'First, the satellite retrievals exhibited an increasing prediction error when hindcasting historical pollution levels.' …and the related discussion.

Response: We added more explanation to clarify this discussion and adjusted the sentences in line 385-391 as follows "First, as reported by previous studies (Xiao et al., 2018;Xue et al., 2019), the satellite-based $PM_{2.5}$ prediction model suffered from increasing prediction error when hindcasting historical pollution levels a long time before the model training time period. One reason could be that some unobserved parameters, e.g. $PM_{2.5}$ composition, modify the associations between $PM_{2.5}$ and predictors, leading to model overfitting. The satellite-driven $PM_{2.5}$ prediction model used in this study is a state-of-the-art prediction model with improved prediction accuracy for high-pollution events, but its hindcast prediction quality could be further improved to better describe the historical $PM_{2.5}$ spatiotemporal distribution."

References:

Cai, W., Li, K., Liao, H., Wang, H., and Wu, L.: Weather conditions conducive to Beijing severe haze more frequent under climate change, Nature Climate Change, 7, 257-262, 10.1038/nclimate3249, 2017.

Chawla, N., Bowyer, K., Hall, L., and Kegelmeyer, W.: SMOTE: Synthetic Minority Over-sampling Technique, J. Artif. Intell. Res. (JAIR), 16, 321-357, 10.1613/jair.953, 2002.

Chawla, N., Lazarevic, A., Hall, L., and Bowyer, K.: SMOTEBoost: Improving Prediction of the Minority Class in Boosting, 107-119 pp., 2003.

Chen, H., Wang, H., Sun, J., Xu, Y., and Yin, Z.: Anthropogenic fine particulate matter

pollution will be exacerbated in eastern China due to 21st century GHG warming, Atmos. Chem. Phys., 19, 233-243, 10.5194/acp-19-233-2019, 2019.

Ghorbani, R., and Ghousi, R.: Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques, IEEE Access, 8, 67899-67911, 10.1109/ACCESS.2020.2986809, 2020.

Saputra, A., and Suharjito: Fraud Detection using Machine Learning in e-Commerce, Int. J. Adv. Comput. Sci. Appl., 10, 332-339, 2019.

Shen, L., Jacob, D. J., Mickley, L. J., Wang, Y., and Zhang, Q.: Insignificant effect of climate change on winter haze pollution in Beijing, Atmos. Chem. Phys., 18, 17489-17496, 10.5194/acp-18-17489-2018, 2018.

Torgo, L.: Data Mining with R: Learning with Case Studies, Chapman & Hall/CRC, 2010.

Wei, J., Li, Z., Cribb, M., Huang, W., Xue, W., Sun, L., Guo, J., Peng, Y., Li, J., Lyapustin, A., Liu, L., Wu, H., and Song, Y.: Improved 1 km resolution PM2.5 estimates across China using enhanced space–time extremely randomized trees, Atmos. Chem. Phys., 20, 3273-3289, 10.5194/acp-20-3273-2020, 2020.

Xiao, Q., Chang, H. H., Geng, G., and Liu, Y.: An Ensemble Machine-Learning Model To Predict Historical PM2.5 Concentrations in China from Satellite Data, Environmental Science & Technology, 52, 13260-13269, 10.1021/acs.est.8b02917, 2018.

Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., and Zhang, Q.: Spatiotemporal continuous estimates of PM2.5 concentrations in China, 2000–2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations, Environment International, 123, 345-357, https://doi.org/10.1016/j.envint.2018.11.075, 2019.

Zhai, S., Jacob, D. J., Wang, X., Shen, L., Li, K., Zhang, Y., Gui, K., Zhao, T., and Liao, H.: Fine particulate matter (PM2.5) trends in China, 2013–2018: separating contributions from anthropogenic emissions and meteorology, Atmos. Chem. Phys., 19, 11031-11041, 10.5194/acp-19-11031-2019, 2019.