Manuscript acp-2018-1191

Title: Effects of air pollution control policies on $PM_{2.5}$ pollution improvement in China from 2005 to 2017: a satellite based perspective

Responses to RC2

General comment:

The paper provides a useful overview of recent air quality control policies in China, while using an independent source of data to assess their efficacy. A statistical method is used to correlate satellite retrievals of Aerosol Optical Depth (AOD) to ground level PM2.5 in China, by correlating AOD with meteorological data, fire spots and forest cover. It uses the large network of Chinese measurement stations to verify the model. The 2013 model, which was developed in another paper (Ma et al 2016) is used to project the concentration of PM2.5 backwards to 2005, while a separate model is developed each year for 2014 - 2017. This gives a 13-year PM2.5 dataset with complete spatial and temporal coverage for 2005 – 2017, which is then used to assess the success of China's air quality control policy that underwent significant changes during this period. Linear trends are calculated for the periods corresponding to specific policies (e.g. Five Year Plans). Calculated PM2.5 concentrations are also compared with official government data, to verify that targets were met. While this retrospective analysis of the success of China's control of PM2.5 pollution is very useful, the authors need to ensure that they acknowledge the role that inter-annual variation in meteorology may play in these trends.

Response: We would like to thank the reviewer for his valuable comments. We have revised the manuscript according to the comments, please see the following responses. For the impact of meteorological conditions, we have discussed this in Lines 11-14, 20-21, P19.

(1) Abstract The majority of the abstract summarizes the discussion section. A brief description of the two stage statistical model, including its predictors could be added.

Response: A brief description of the two stage statistical model and its predictors have been added in abstract. See Lines 8-11, P2.

(2) Intro P3, L23: It may be worth adding a sentence that briefly explains what the ‘scaling method’ is. There is a citation to Liu 2014 to back up the statement that, “Compared to the scaling method, statistical models have greater prediction accuracy but require large amount ground-measured PM2.5 data to develop the statistical models (Liu, 2014)”. However, there is not a reference that corresponds to the “Liu, 2014” citation. Since the justification of method choice relies on this reference, it should be added before the paper is reviewed again.
Response: Done. See Line 25, P3~Line 1, P4.

(3) Overview of air pollution control policies in China from 2005 to 2017 This section is a very broad summary of the actions within Five Year Plans and other major government directives that are relevant to air pollution control. The specific policies (e.g. ‘Implement desulphurization and denitration facilities for coal-fired power sector and major industrial sectors’) are summarised in Table 1, along with the metrics by which the policies’ success will be judged. It may be useful to, where possible, cite government press releases/reports or literature that assess the success of these policies. However, the text in this section does not make any mention of the policies listen in Table 1. It would be useful for the reader for some information from Table 1 to be synthesised into this section, along with citations to previous studies that have attempted to assess the success of these policies (e.g. Schreifels et al, 2012)
Reference: Schreifels, Jeremy J., Yale Fu, and Elizabeth J. Wilson. "Sulfur dioxide control in China: policy evolution during the 10th and 11th Five-year Plans and lessons for the future." Energy Policy48 (2012): 779-789.
Response: This comment is helpful. After careful consideration, we added major air pollution control measures, corresponding achievements, and how these policies were associated with PM2.5 pollutions in the main text and cited relevant references, including reference of Schreifels et al, 2012. See Lines 4-13, P13; Lines 22-27, P13; Lines 13-21, P14; Line 23, P14~Line 3, P15; Line 23, P15~Line 22, P16.

(4) P5, L13. It may be worth defining what China's 'new air quality standard' here, where it is first mentioned. It may be useful to provide the old air quality standard, and the name of the standard (GB 3095-2012). Currently the actual threshold number of China's air quality standard is first referenced of P13, L10 in the conclusion.

Response: Done. We briefly described the new air quality standard in Lines 4-11, P15.

(5) Data and Method P6, L19: Paper uses PM2.5 data from the CNEMC. Other papers, (e.g. Rohde and Muller (2015); Liu et al (2016)) have noted quality issues with this data. Were any quality control procedures applied to this data?

References: Liu, Jianzheng, Weifeng Li, and Jie Li. "Quality screening for air quality monitoring data in China." Environmental pollution216 (2016): 720-723.

Rohde, Robert A., and Richard A. Muller. "Air pollution in China: mapping of concentrations and sources." PloS one10.8 (2015): e0135749.

Response: Yes, we performed the data screening procedure before model fitting. Abnormal values (extreme high or extreme low values for a site compared with its neighboring sites, repeated values for continuous hours, etc.) were deleted before model fitting. We required at least 20 hourly records to calculate the daily average $PM_{2.5}$ concentrations. Please see Lines 17-21, P7.

(6) Since the ground monitoring stations are typically within urban areas, could this bias the statistical model so that the PM2.5 predictions for non-urban areas is inaccurate? Why use the updated data to create separate statistical models for 2014, 2015, 2016 and 2017, yet only use the 2013 model to project back the PM2.5? Why should the 2013 model be more appropriate than the other years? Why not combine all the years where measurements are available? How is it justified to fit the model separately to the data in each province? Isn't using province boundaries somewhat arbitrary?

Response: Yes, we acknowledge this is a problem in statistical modeling of satellite PM2.5. We have discussed this in Lines 6-12, P11.

There are two reasons that we only use the 2013 model to project back the PM2.5. First, the historical data were derived from our previous study, which only used the 2013 model. This dataset has been shown high accuracy and has been widely used in environmental

epidemiological (Liu et al., 2016;Wang et al., 2018a), health impact (Liu et al., 2017;Wang et al., 2018b), and social economic impact (Chen and Jin, 2019;Yang and Zhang, 2018) studies in China. Second, a recent study has shown that the historical hindcast ability of the annual model decreased when hindcast year was long before the model year (Xiao et al., 2018). Therefore, we did not use the models of 2014 to 2017 to estimate the hindcast $PM_{2.5}$.
For provincial models, we added the description how we fit the provincial model in Line 13-16, P8. We added the provincial results in Table S2-S4 (Supplementary Materials). And described the results in Line 23, P10~Line 5, P11. Results showed that the performance of first-stage LME model would greatly decreased if we fit the model for entire China.

(7) Many other studies of trends in atmospheric concentrations use a non-parametric trend estimator such as the Thiel-Sen slope estimator. The authors should justify their choice of the least squares regression to estimate the slope of the trend.
Response: In fact, the method we used in this study has been successfully applied to trend analyses of monthly mean $PM_{2.5}$ and AOD anomaly time-series data (Hsu et al., 2012;Boys et al., 2014;Zhang and Reid, 2010;Xue et al., 2019). Therefore, we thought that the method we used is appropriate. See Lines 22-24, P9. Besides, we added a description of the method. Please see Lines 14-22, P9.

(8) In the results section, and Figures 6 & 7, a p threshold of 0.1 is mentioned, but you do not mention in the methods which statistical test you used to check the significance of your trends.
Response: The method of *t* test was used to obtain the statistical significance of the trends. See Lines 21-22, P9.

(9) Some of these questions about the methodology can be answered by reading the author's previous Ma et al 2016 paper, which is published in Environmental Health Perspectives. I recommend the authors reduce their reliance on referring to this previous paper, so that the methods section in the current paper can be understood without referring to another paper which the reader will not necessarily have access to.
Response: We added details about the equations of the two-stage model, please see P8-P9.

(10) P5, L26: Is it useful to the reader to list 9 studies that have referenced your previous paper? This list includes studies that this paper's co-authors are also co-authors on.

Response: These papers were the follow up studies using the $PM_{2.5}$ dataset from 2004 to 2013 we developed in our previous study. Although some of them are the follow up studies by co-authors of this study, the publications of these studies show that the $PM_{2.5}$ dataset has been widely recognized and used in academic field. And these references can support the rationality that we directly use this $PM_{2.5}$ dataset from 2004 to 2013 in current study.

According to this comment, we have removed 3 references here (see Lines 13-15, P6) to simplify this paragraph.

(11) Results and Discussions Is it really useful to compare the PM2.5 trend with the corresponding FYP policies? This suggests that policies have immediate effects, and that they are the main contributor to the trends in PM2.5. There are other important confounding factors such as interannual variation in meteorology, China's economic output etc. May be best to avoid statements on the effectiveness of certain policies, or mention the above caveats in the conclusion.

Response: We added discussions about the impacts of meteorology and economic. See Lines 11-21, P19.

(12) I suggest the authors add a comparison of their results with other research that quantifies the trend in PM2.5 derived AOD in China, such as Lin et al., 2017. It may be interesting to perform a non-linear trend analysis on this dataset in certain key regions (e.g. Jing-Jin-Ji or PRD).

Reference: Lin, C. Q., Liu, G., Lau, A. K. H., Li, Y., Li, C. C., Fung, J. C. H., & Lao, X. Q. (2018). High-resolution satellite remote sensing of provincial PM2. 5 trends in China from 2001 to 2015. Atmospheric Environment, 180, 110-116.

Response: The revision has been made. We compared our results with two recent studies. See Lines 3-13, P18.

(13) As you break down the trend into multiple overlapping periods of different lengths, it is difficult to get an overall impression of the rises and falls in the trend in different regions. Alternatively, a figure could be added with the yearly or monthly deseasonalised PM2.5 (averaged by different regions)

Response: We have added a new figure (Figure 6, P27) according to the comment. And we moved a table from supplementary materials to the main manuscript (see Table 2), which corresponds to Figure 6.

(14) I suggest the authors also mention the possibility of contribution of natural sources of aerosol to the trends. At P10, L16, the authors mention that the majority of the trend in PM2.5 during 2010-2013 are driven by decreases in Xinjiang and Central Inner Mongolia, which are both desert regions where the PM2.5 likely has a high dust component. This can be seen in your results. For example in panel (e) of Figure 7, where the western half of the Taklamakan desert has a strong positive trend, despite it being unlikely that there are large changes in emissions in this unpopulated region.

Response: The possible impact of dust in this region has been added. See Lines 6-9, P14.

(15) P3, L8: “However, the Chinese government did not realize the PM2.5 issues until 2012.” This sentence seems disingenuous and qualitative so should be removed or rephrased.

Response: We have changed “realize” to “focus on”. See Line 8, P3.

(16) P4, L6: Remove or replace the word ‘preliminary’.

Response: We changed it to “preliminarily”. Line 11, P4.

(17) P5 L14. “These policies indicated that the air pollution control in China began to focus on air quality improvement.” This sentence could be rephrased, as it is currently seems tautological.

Response: We changed it to “These policies indicated that the focus of air pollution control in China began to focus on $PM_{2.5}$ concentrations reductions”. See Lines 19-20, P5.

(18) P10, L22: The sentence “As the further development of social economic, the ECER policy had shown its bottleneck for PM2.5 reductions.” does not make sense. Bottleneck may be the wrong word to describe this.

Response: We have rephrase “bottleneck” to “limitation”. See Line 13, P14.

(19) P12, L25. Change ‘to addressed’ to “to address.”

Response: Done. See Line 3, P19.

(20) P13, L6. ‘Temporal’ is not the right word here. Should be temporary?

Response: Done. See Line 13, P16.

References:

Boys, B., Martin, R., van Donkelaar, A., MacDonell, R., Hsu, C., Cooper, M., Yantosca, R., Lu, Z., Streets, D. G., Zhang, Q., and Wang, S.: Fifteen-year global time series of satellite-derived fine particulate matter, Environ. Sci. Technol., 48, 11109-11118, 2014.

Chen, S., and Jin, H.: Pricing for the clean air: Evidence from Chinese housing market, J. Clean. Prod., 206, 297-306, 2019.

Hsu, N. C., Gautam, R., Sayer, A. M., Bettenhausen, C., Li, C., Jeong, M. J., Tsay, S. C., and Holben, B. N.: Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010, Atmos. Chem. Phys., 12, 8037-8053, 2012.

Liu, C., Yang, C., Zhao, Y., Ma, Z., Bi, J., Liu, Y., Meng, X., Wang, Y., Cai, J., and Kan, H.: Associations between long-term exposure to ambient particulate air pollution and type 2 diabetes prevalence, blood glucose and glycosylated hemoglobin levels in China, Environ. Int., 92, 416-421, 2016.

Liu, M., Huang, Y., Ma, Z., Jin, Z., Liu, X., Wang, H., Liu, Y., Wang, J., Jantunen, M., Bi, J., and Kinney, P. L.: Spatial and temporal trends in the mortality burden of air pollution in China: 2004–2012, Environ. Int., 98, 75-81, 2017.

Wang, C., Xu, J., Yang, L., Xu, Y., Zhang, X., Bai, C., Kang, J., Ran, P., Shen, H., and Wen, F.: Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study, The Lancet, 391, 1706-1717, 2018a.

Wang, Q., Wang, J., He, M. Z., Kinney, P. L., and Li, T.: A county-level estimate of PM2. 5 related chronic mortality risk in China based on multi-model exposure data, Environ. Int., 110, 105-112, 2018b.

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, Environ. Sci. Technol., 52, 13260-13269, 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, Environ. Int., 123, 345-357, 2019.

Yang, J., and Zhang, B.: Air pollution and healthcare expenditure: Implication for the benefit of air pollution control in China, Environ. Int., 120, 443-455, 2018.

Zhang, J., and Reid, J. S.: A decadal regional and global trend analysis of the aerosol optical depth using a data-assimilation grade over-water MODIS and Level 2 MISR aerosol products, Atmos. Chem. Phys., 10, 10949-10963, 2010.