

We thank the reviewers for their valuable comments. We have made efforts to improve the manuscript accordingly. This document is organized as follows: the referees' comments are in **Bold black**, our responses are in plain black text, and the revisions in the manuscript are shown in blue. The line numbers in this document refer to the updated manuscript.

Anonymous Referee #2

**In this study Zhai et al. use a combination of PM<sub>2.5</sub> observations and multiple linear regression modelling to analyse the trend in PM<sub>2.5</sub> concentration across mainland China during 2013-2018 and to quantify the meteorological contribution to this trend. Overall the paper is well thought-out and written, and figures are well presented. The topic of the study is interesting and well within the scope of ACP. I recommend publication once the comments below (mostly regarding the processing of the data) have been addressed.**

**1. Abstract, L25: I suggest specifying that the contribution is “statistically” significant, otherwise the sentence reads a bit odd.**

Thanks. To make it clear, we have rephrased this part to:

The meteorology-corrected PM<sub>2.5</sub> trends after removal of the MLR meteorological contribution can be viewed as driven by trends in anthropogenic emissions. The mean PM<sub>2.5</sub> decrease across China is -4.6 ug m<sup>-3</sup> a<sup>-1</sup> in the meteorology-corrected data, 12% weaker than in the original data. The trends in the meteorology-corrected data for the five megacity clusters are: ...

**2. Abstract, L26: I think the statement “reduces the uncertainty on the emission-driven trends” needs more explanation in the abstract. Earlier in the abstract you refer to the difficulty of trend attribution because of the meteorologically driven interannual variability in PM<sub>2.5</sub> concentrations. However, it is not immediately clear what you mean by “uncertainty on the emission-driven trends”. (It is worded more clearly in the conclusions section).**

To make it clear, we deleted ‘reduces the uncertainty on the emission-driven trends’, and reworded this part in the abstract as:

The meteorology-corrected PM<sub>2.5</sub> trends after removal of the MLR meteorological contribution can be viewed as driven by trends in anthropogenic emissions. The mean PM<sub>2.5</sub> decrease across China is -4.6 μg m<sup>-3</sup> a<sup>-1</sup> in the meteorology-corrected data, 12% weaker than in the original data. The trends in the meteorology-corrected data for the five megacity clusters are: ...

**3. Introduction, L52: Please add an explanation to why the PM<sub>2.5</sub> concentration is correlated to V850, particularly in the NCP.**

Explanation is added in lines 93-94: V850 in particular is a strong predictor of PM<sub>2.5</sub> wintertime pollution events in the North China Plain, because northerly winds (negative V850) ventilate the region with clean dry air (Cai et al., 2017; Pendergrass et al., 2019).

Added reference:

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, 257-263, 2017.

**4. Introduction, final paragraph: References to a few papers that have identified/quantified recent trends in PM<sub>2.5</sub> concentrations across China seem to be missing from the introduction (Ma et al., 2019; Silver et al., 2018; Liang et al., 2016).**

Added. Thanks.

**5. Section 2.1, L88: Can you give any example references here for these previous studies?**

Example references (Wang et al., 2014; Cai et al., 2017; Shen et al., 2017; Leung et al., 2018; Song et al., 2019; Zou et al., 2017) are added.

**6. Section 2.1, L90: Why was 70% chosen? It seems quite low to me. Please add some justification. Did you do any sensitivity tests changing the threshold to a higher percentage?**

This threshold is aimed to include sites that have continuous observations since early 2013 (mainly sites located in the 74 major cities). We have tried to use data from the 74 major cities and obtained identical trend results. I then improved the threshold to 80% and 90% and find that although the number of valid sites in each target region decreased a little bit, the pollutants trends have negligible differences compared with the trends when the '70% threshold' was used.

Added justification in Lines 96-97: [We did sensitivity tests with data coverage thresholds changing from 70% to 90% and obtained similar pollutants trends. To make the most use of available data, 70% is chosen.](#)

**7. Section 2.1, L90-91: As above, can you add some justification for the coarse grid chosen? Is this recommended by Tai et al. (2012)? Did you test any other grid resolutions?**

It is recommended by Tai et al. (2012) and Shen et al. (2017). I have tried to use  $0.5^{\circ} \times 0.625^{\circ}$  grid resolution. However, finer resolution will result in too few valid grids (grids that have both PM<sub>2.5</sub> and meteorology observations).

We have reworded the text in the manuscript as: For the MLR model, we further average all data on a  $2^{\circ} \times 2.5^{\circ}$  grid to increase statistical robustness [following Tai et al. \(2012\) and Shen et al. \(2017\)](#).

**8. Section 2.1: I see that you removed severe outliers from the observation dataset but what did you do about repeating consecutive values in the dataset (e.g. identified in Rohde and Muller (2015)) and day-to-day repeating sequences of values (e.g. identified by Silver et al., 2018)? If these were not removed, please at least acknowledge that data issues are likely remain in the dataset.**

Reply: Thank you for pointing this out. I then checked the impacts of those consecutive repeats on this study. It turned out that these consecutive repeating values have negligible impacts on results in this study. Nevertheless, consecutive repeats identified by Rohde and Muller (2015) and Silver et al. (2018) are unlikely 'realistic' values, and are then removed throughout this study. We removed values from the hourly time series when there are >24 consecutive repeats. The threshold of '>24 consecutive repeats' were chosen by applying a series of thresholds from '>4' to '>24', and we find that different thresholds lead to negligible changes in this study results.

Changes in the manuscript (Line80-83):

At the end of Section 2.1: [There are also occasional consecutive repeats of data that may be caused by faulty instruments or reporting \(Rohde and Muller, 2015; Silver et al., 2018\). Here we removed values](#)

from the hourly time series when there are >24 consecutive repeats. This in whole removed 7.4%, 7.0%, 6.4%, and 6.7% of the PM<sub>2.5</sub>, SO<sub>2</sub>, NO<sub>2</sub>, and CO data respectively.

Added reference:

Rohde, R. A., and Muller, R. A.: Air pollution in China: mapping of concentrations and sources, PloS one, 10, e0135749, 2015.

**9. Section 2.1: If each year of data and each station were considered separately when applying the 70% data threshold, what does the introduction of more stations/data towards the end of the sampling time period (as more stations come online) do to the trends? I realize the data is averaged over large grid cells, but introduction of many stations later in the time-series (that are not consistent) may impact the trends calculated. Please add some explanation. Did you attempt to calculate the trends based only on data from stations that were online in 2013?**

Thanks. This is a good point and we have already considered this in this study. For trend analysis, we only retained data from sites with at least 70% of data coverage for each year from 2013 to 2018. That is, the selected sites must have at least 70% data coverage for each of the 6 years from 2013 to 2018 simultaneously. In this way, we are using consistent sites throughout 2013-2018 for trend analysis. To make it clear, we have made modifications in Lines 96, 106 & 449: [for each of the 6 years from 2013 to 2018.](#)

**10. Section 2.2 and 2.3: I am slightly confused by the explanation of the deseasonalizing and detrending process. Perhaps the explanation could be reworded slightly? I have understood it as: the data is de-seasonalized and detrended by taking the 50-day moving average from the 10-day means; whereas the anomalies are deseasonalized (but not detrended) by taking the 6-year mean 50-day moving average from the 10-day means; is this correct?**

Yes, this is correct. Explanations in the manuscript are reworded as follows:

P4, Line 112-113: [The deseasonalized and detrended time series are obtained by removing the 50-day moving averages from the 10-day mean time series.](#)

P5, Line 129-130: [We thus apply equation \(1\) to the meteorological anomalies  \$X\_{a,i,k}\$ , obtained by removing the 6-year means of the 50-day moving averages from the 10-day mean time series.](#)

**11. Section 3.1: I think this section is really nice and gives some good explanations (with references) for the drivers of the changes in pollutants and/or emissions. However, there is no comparison with previous studies that have calculated trends in PM<sub>2.5</sub> concentrations over similar time periods (e.g. Ma et al., 2019; Silver et al., 2018; there may be others). Are the calculated trends consistent between studies, despite differences in the data or data processing? I understand that the trends are all calculated over slightly different time periods, but at least a qualitative comparison should be added to the text.**

Thanks, the following line is added in P9, Line 168-169: [Trends in China PM<sub>2.5</sub>, SO<sub>2</sub>, and NO<sub>2</sub> presented here are consistent with previous studies \(Silver et al., 2018; Ma et al., 2019\) that cover a shorter time period than 2013-2018.](#)

**12. Conclusions: this section is a nice summary of the main points of the paper. However, it would make the results even clearer if the percentage difference from the original trend was**

**quoted here again as in the abstract and it also might be worth explaining again here what is meant by meteorologically corrected data.**

P9, Lines 241-245 are revised as:

We refer to the data series after removal of meteorological variability as the meteorology-corrected data. Thus the 2013-2018 PM<sub>2.5</sub> decrease for Beijing-Tianjin-Hebei is  $-9.3 \pm 1.8 \mu\text{g m}^{-3} \text{ a}^{-1}$  in the original data and is 14% weaker in the meteorology-corrected data ( $-8.0 \pm 1.1 \mu\text{g m}^{-3} \text{ a}^{-1}$ ). For the Sichuan Basin where the meteorological correction is particularly large, the PM<sub>2.5</sub> decrease is  $-6.7 \pm 1.3 \mu\text{g m}^{-3} \text{ a}^{-1}$  in the original data and is reduced by 27% to  $-4.9 \pm 0.9 \mu\text{g m}^{-3} \text{ a}^{-1}$  in the meteorology-corrected data.

Explanation of the meteorology-corrected data are also added in the Lines 24-26 and in Lines 137-138.

Additional changes:

1) References updated or newly added:

Cheng, J., Su, J., Cui, T., Li, X., Dong, X., Sun, F., Yang, Y., Tong, D., Zheng, Y., Li, Y., Li, J., Zhang, Q., and He, K.: Dominant role of emission reduction in PM<sub>2.5</sub> air quality improvement in Beijing during 2013-2017: a model-based decomposition analysis, *Atmos. Chem. Phys.*, 19, 6125-6146, 10.5194/acp-19-6125-2019, 2019.

Song, S., Gao, M., Xu, W., Sun, Y., Worsnop, D. R., Jayne, J. T., Zhang, Y., Zhu, L., Li, M., Zhou, Z., Cheng, C., Lv, Y., Wang, Y., Peng, W., Xu, X., Lin, N., Wang, Y., Wang, S., Munger, J. W., Jacob, D. J., and McElroy, M. B.: Possible heterogeneous chemistry of hydroxymethanesulfonate (HMS) in northern China winter haze, *Atmos. Chem. Phys.*, 19, 1357-1371, 10.5194/acp-19-1357-2019, 2019.

Zou, Y., Wang, Y., Zhang, Y., and Koo, J.-H.: Arctic sea ice, Eurasia snow, and extreme winter haze in China, *Science Advances*, 3, e1602751, 10.1126/sciadv.1602751, 2017.

2) Emission trends in Figure 2 for Pearl River Delta and Sichuan Basin are correctly reversed.