Referee's comments are in red, our reply on black, quotes in the revised manuscript in blue.

Referee#1

I have limited my review to focus on areas which the authors have changed in response to my prior comments. Line numbers refer to the marked-up manuscript.

I was very impressed by the thoroughness of the authors' response. In particular I was happy to see the transition to an observational dataset in place of model reanalysis, and the extension to investigate $PM_{2.5}$ elevates the paper substantially and brings it much closer to publication. However this also means that the manuscript must now pass the same level of scrutiny as any other investigation of future $PM_{2.5}$, which is a high bar. This is the focus of my remaining concerns. The analysis is novel and I recognize the need for an efficient approach rather than (say) an additional set of CCM simulations, but the MLR approach used by the authors does cause me some concern. I have enumerated those concerns below and hope that the authors can address them.

The most significant issue is that a regression on a limited set of variables from historical data is used to predict future conditions. This is not inherently/fundamentally flawed, but there is a large body of literature which has investigated the nuanced relationship between future changes in climate and air quality, and how they are moderated by meteorological change (e.g. Jacob and Winner 2009, Fiore et al 2015). This has been looked at specifically in the context of health in China for ozone by e.g. Westervelt et al (2019). The challenge for modelers (including, now, the authors of this study) is whether past conditions accurately reflect the changes which will occur in the future. For example, it is possible that a geoengineering scenario could modify large scale dynamics in a way which is not reflected in past conditions, and which is different again from how those dynamics will be affected by climate change (Cheng et al 2022). It is also possible that the precursors dominating $PM_{2.5}$ will change, modifying the relationship between emissions and concentrations. Such changes would affect the patterns of pollution movement and evolution in a way which a local regression would not be able to capture. With that in mind, I would recommend three significant further revisions (two focused on the above and one on framing).

Reply: We would like to thank the referee for taking the time to review our manuscript again. Thank you very much for your affirmation of our first round of modification and constructive suggestions for the rationality of MLR in projecting $PM_{2.5}$. We have responded to the following comments one by one.

First, I recommend that the authors take an existing dataset of air quality outcomes for current and future conditions and show that the MLR method is capable of providing reasonable results when past conditions are used to build a regressor which predicts future $PM_{2.5}$ with evolving emissions and climate. One possibility in this regard would be the AerChemMIP model outputs. It is fair to say that there is a lack of data to accomplish this for geoengineering output (although GeoMIP and/or GLENS output may be sufficient). If the authors can at least show that a regressor provides a reasonably accurate prediction under a significant change in climate and emissions that would significantly strengthen their findings in this paper.

Reply: We thank the referee's comments and suggestions. We found one paper which looked at the future $PM_{2.5}$ concentration in the similar region and asked for their data to assess our results. Li et al (2023) used the CMAQ model coupled WRF driven by GFDL-ESM2G and SMOKE model to explore the influence of emissions on air quality in the Beijing-Tianjin-Hebei region of China in 2050. The authors used the dynamical downscaled meteorological factors by WRF driven by GFDL-ESM2G and two air pollution emission scenarios, one is "base" based on the Beijing City Master Plan (2016-2035) and another is "EIT1" based on the emission reduction for WHO Interim Target-1 to compare the impact of different emission scenarios on $PM_{2.5}$ concentration in 2050 under RCP4.5. To assess the performance of our regression model we also downloaded the meteorological variables from GFDL-ESM2G under RCP4.5 and the "EIT1" emission data.

The statistical downscaled meteorological factors during 2008-2017 and 2050 under RCP4.5 were used as independent variables in the regression model to project $PM_{2.5}$ concentration in 2050 under RCP4.5 with the "EIT1" scenario. The spatial pattern is shown in the following figure S7. Although $PM_{2.5}$ concentration is nearly twice as high as from Li et al., $PM_{2.5}$ concentration from our regression model is also higher than the referenced data during 2008-2017, and our projections are similar to the spatial pattern of the seasonal $PM_{2.5}$ concentration from the chemical transport model, with correlation coefficient of 0.68-0.73. We also compare the spatial pattern of differences in $PM_{2.5}$ concentration between "base" and "EIT1" under RCP4.5 (Figure S8). Because of the small slope coefficient of $PM_{2.5}$ emission in our MLR we do not capture the large reduction of $PM_{2.5}$ concentration in the Beijing city center seen by Li et al (2023), (Fig. S8).

We added the following figures in the supplementary information.

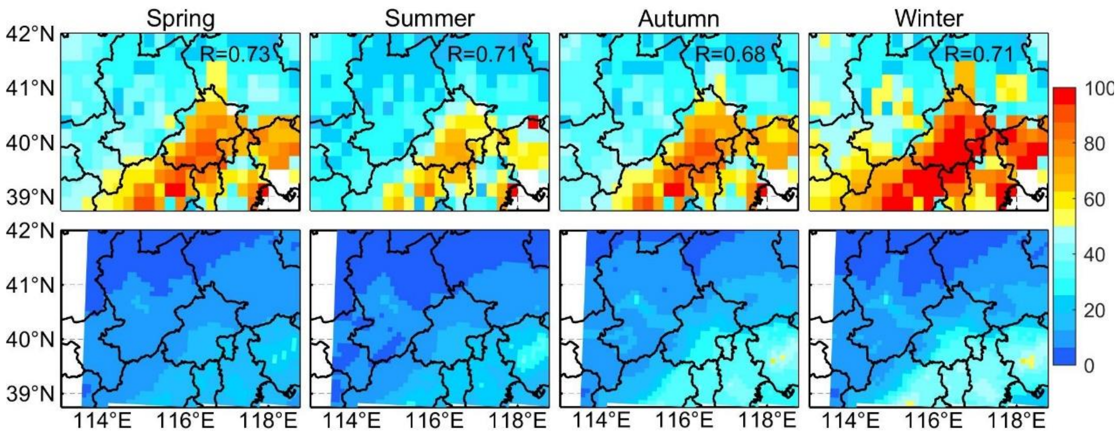

Figure S7. Comparison of our MLR model projection and Li et al. (2023) RCP4.5 simulations. Li et al (2023) use the CMAQ model coupled WRF driven by GFDL-ESM2G and SMOKE model to explore the influence of emissions on air quality in the Beijing-Tianjin-Hebei region of China in 2050. The authors used the dynamical downscaled meteorological factors by WRF driven by GFDL-ESM2G and two air pollution emission scenarios, one is "base" based on the Beijing City Master Plan (2016-2035) and another is "EIT1" based on the emission reduction for WHO Interim Target-1 to compare the impact of different emission scenarios on PM2.5 concentration in 2050 under RCP4.5. To assess the performance of our regression model we also downloaded the meteorological variables from GFDL-ESM2G under RCP4.5 and the "EIT1" emission data. The statistical downscaled meteorological factors during 2008-2017 and 2050 under RCP4.5 were used as independent variables in the regression model to project PM2.5 concentration in 2050 under RCP4.5 with the "EIT1" scenario. The top row are calculated by our regression model, and the bottom row are from Li et al. R is the correlation coefficient of $PM_{2.5}$ concentration spatial pattern between our results and Li et al.

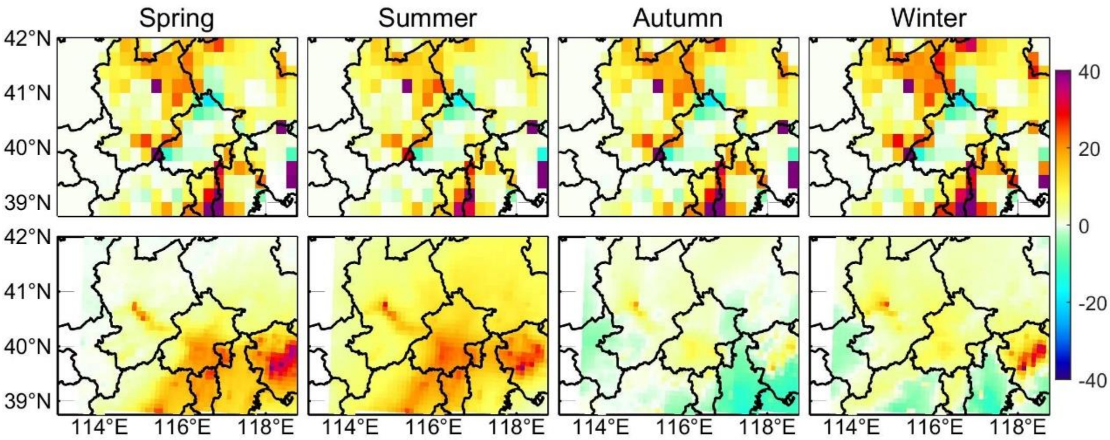

Figure S8. Spatial pattern of differences in PM2.5 concentration under RCP4.5 between "base" and "EIT1" emission scenarios in Li et al (2023). The top row are calculated by our regression model, and the bottom row are from Li et al.

We added the following sentences in line 299.

We also tested the accuracy of our MLR model projection against simulations (Li et al., 2023) with the Community Multiscale Air Quality (CMAQ) model developed by the United States Environmental Protection Agency and which can simulate particulate matter on local scales (Foley et al., 2010; Yang et al., 2019) when coupled to WRF. We used the same meteorological forcing as Li with the "EIT1" $PM_{2.5}$ emissions scenario in 2050 under RCP4.5 (Fig. S7).

The spatial patterns are well correlated in all seasons (0.68-0.73), but $PM_{2.5}$ concentrations are about twice as high in our MLR model as from Li et al., (2023). $PM_{2.5}$ concentrations from our regression model are also higher than the referenced data during 2008-2017. While the difference in absolute PM2.5 concentrations are significant, we mainly consider differences of $PM_{2.5}$ concentration between G4 and RCP4.5/RCP8.5 in our study which we cannot compare these anomalies with the single

RCP4.5 scenario simulated by Li et al. (2023). We do compare the spatial pattern of differences in $PM_{2.5}$ concentration between "base" and "EIT1" under RCP4.5. Because of the small slope coefficient of $PM_{2.5}$ emission in our MLR, we do not capture the large reduction of $PM_{2.5}$ concentration in the Beijing city center seen by Li et al (2023), (Fig. S8).

References

Li, D., Wu, Q., Feng, J., Wang, Y., Wang, L., Xu, Q., Sun, Y., Cao, K., and Cheng, H.: The influence of anthropogenic emissions on air quality in Beijing-Tianjin-Hebei of China around 2050 under the future climate scenario, J. Cleaner Prod., 388, 135927, https://doi.org/10.1016/j.jclepro.2023.135927, 2023.

Foley, K. M., Roselle, S. J., Appel, K. W., Bhave, P. V., Pleim, J. E., Otte, T. L., Mathur, R., Sarwar, G., Young, J. O., Gilliam, R. C., Nolte, C. G., Kelly, J. T., Gilliland, A. B., and Bash, J. O.: Incremental testing of the Community Multiscale Air Quality (CMAQ) modeling system version 4.7, Geosci. Model Dev., 3, 205–226, https://doi.org/10.5194/gmd-3-205-2010, 2010.

Yang, X., Wu, Q., Zhao, R., Cheng, H., He, H., Ma, Q., Wang, L., and Luo, H.: New method for evaluating winter air quality: PM2.5 assessment using Community Multiscale Air Quality Modeling (CMAQ) in Xi'an, Atmos. Environ., 211, 18-28, https:// doi.org/10.1016/j.atmosenv.2019.04.019, 2019.

One of the most significant concerns I have in this respect is actually the nature of the regression. If I understand Sections 2.2 and 2.5 correctly, the authors are relating local $PM_{2.5}$ concentrations to local $PM_{2.5}$ emissions and local meteorology. However, $PM_{2.5}$ is will known to be influence by both upwind (i.e. regional) emissions of $PM_{2.5}$ and by emissions of $PM_{2.5}$ precursors such as SO2, NOx, and ammonia. The importance of taking these factors is elevated when looking at higher resolution data. Based on my interpretation of lines 285-293, these factors are not included in the MLR which would be concerning. If my interpretation is incorrect then I recommend that the authors clarify this in the relevant sections and specify clearly a) what precursors are considered, b) how spatial relationships between emissions (or other factors) and concentrations are captured, and c) how their model will be able to capture a shift in the chemical regime. Concerns a and c are only significant if secondary $PM_{2.5}$ is considered, so if instead only primary $PM_{2.5}$ is considered then I strongly recommend this be made very clear in the paper and the conclusions and abstract caveated appropriately. However, in either case the question regarding concern b remains.

Reply: Thank you very much for this comment. a) The reviewer is right, we only considered the primary $PM_{2.5}$ emissions and did not consider the precursor gases for secondary $PM_{2.5}$. Although secondary $PM_{2.5}$ emission is not included, $PM_{2.5}$ concentration includes both primary and secondary $PM_{2.5}$ in our model. b) The referee is correct that $PM_{2.5}$ concentration is not only related to local meteorological conditions and emissions. Limited by our model being a statistical model rather than a chemical transport model, we expect that by having meridional and latitudinal winds as variables in our model that these $PM_{2.5}$ advections can be accounted for. c) We note that the future precursor mix will change in ways that are rather speculative as they depend on technological innovation and policies that are inherently unpredictable.

We have added the following sentences in line 275.

Here, we use $PM_{2.5}$ concentration including both primary and secondary $PM_{2.5}$ as the dependent variable and primary $PM_{2.5}$ emission and meteorological factors as independent variables in the MLR. Future $PM_{2.5}$ emissions will change in ways that are rather speculative as they depend on technological innovation and policies that are inherently unpredictable. The MLR assumes that the past emissions mix and secondary aerosols remain unchanged in the future, but meteorological factors will also indirectly impact secondary $PM_{2.5}$ to some extent.

We have added the following sentences in line 810.

Our study did not consider the impacts of socio-economic pathways on $PM_{2.5}$ future emissions, instead we explore the meteorological differences between the SAI G4 scenario and the greenhouse gas RCP4.5/RCP8.5 on $PM_{2.5}$ concentrations. $PM_{2.5}$ emissions were defined by the uncontrolled ("baseline") and a scenario where technological intervention ("mitigation") reduces emissions. There are some limitations in our study. Firstly, the HTAP_V3 dataset only includes anthropogenic $PM_{2.5}$ emission, not natural $PM_{2.5}$ emission. Natural $PM_{2.5}$ will also change in the future under changing climate. The sources of natural $PM_{2.5}$ include the sandstorms that sometimes occur in spring as extreme winds mobilize dry unvegetated soils. These relatively extreme conditions are difficult to simulate in ESM and subject to land use policy e.g., the numerous ecosystem service measures undertaken by China over the last five decades (Miao et al.,2015). Secondly, although $PM_{2.5}$ concentration includes both primary and secondary $PM_{2.5}$ during model training, we do not consider the precursor gases for secondary $PM_{2.5}$ directly. The sensitivity of MLR may diminish at the high $PM_{2.5}$ values when secondary $PM_{2.5}$ dominates the variability of total $PM_{2.5}$ (Upadhyay et al., 2018). Thirdly, we only consider the effect of dominant near-surface meteorological variables on the $PM_{2.5}$. However, the vertical transport of pollutants related to vertical atmospheric stability should not be ignored (Lo et al., 2006; Wu et al., 2005), and this may contribute to the differences in RCP4.5 scenario from our MLR model and more sophisticated simulations (Fig. S7). Finally, although it is insignificant for the Beijing and Tianjin provinces, the MLR model suffers collinearity problems in some areas. These factors play smaller roles as we are mainly considering changes in $PM_{2.5}$ concentration between different climate scenarios. Nevertheless, projection for changes in $PM_{2.5}$ between SAI scenarios and per greenhouse gas scenarios would be valuable for global air quality impacts from geoengineering.

References

Lo, J., Lau, A., Fung, J., and Chen, F.: Investigation of enhanced cross-city transport and trapping of air pollutants by coastal and urban land-sea breeze circulations, J. Geophys. Res.-Atmos., 111(D14), https://doi.org/10.1029/2005JD006837, 2006.

Wu, D., Tie, X., Li, C., Ying, Z., Kai-Hon Lau, A., Huang, J., Deng, X., and Bi, X.: An extremely low visibility event over the Guangzhou region: a case study, Atmos. Environ., 39, 6568-6577, https://doi.org/10.1016/j.atmosenv.2005.07.061, 2005.

Miao, L., Moore, J. C., Zeng, F., Lei, J., Ding, J., He, B., and Cui, X.: Footprint of research in desertification management in China, Land Degrad. Dev., 26, 450-457, https://doi.org/10.1002/ldr.2399, 2015.

Second, I recommend that the authors compare their findings against existing projections of the change in surface $PM_{2.5}$ in the target region over the next 40 years. There are several studies looking at how surface air quality in China might evolve under different scenarios (see e.g. Hong et al 2019). Showing that the regression-based approach can recover the majority of the climate change-induced signal would be valuable not only from the perspective of this paper, but from the perspective of the field more broadly.

Reply: Thank you, we added some sentences in the discussion in line 765.

Xu et al. (2021) projected 2030 $PM_{2.5}$ concentrations will decrease by 8.8% and 5.5% under RCP4.5 and RCP8.5 respectively relative to 2015. Wang et al. (2021) also projected decreasing trends in China under RCP4.5 and RCP8.5 during 2030-2050. There were seasonal changes in $PM_{2.5}$ concentration differences between RCP4.5/8.5 scenarios and the historical scenario near the Bohai Sea (Dou et al., 2021). However, there are also some simulations where $PM_{2.5}$ concentrations increase in warmer climates. Hong et al. (2019) suggest that annual mean $PM_{2.5}$ concentrations will increase 1-8 $\mu g/m^3$ in an area including Beijing and Tianjin under RCP4.5 during 2046-2050, compared with 2006-2010. These inconsistent responses are mainly caused by the differences in the selection of ESMs, chemical transport models and climate/emission scenarios. Different RCP scenarios not only correspond to different future climate states, but also have different anthropogenic emissions of air pollutants. In our study, we do not consider the $PM_{2.5}$ emission differences between RCP4.5 and RCP8.5, and instead applied the ECLIPSE $PM_{2.5}$ emission scenarios in our MLR projection.

References

Xu, J., Yao, M., Wu, W., Qiao, X., Zhang, H., Wang, P., Yang, X., Zhao, X., and Zhang, J.: Estimation of ambient $PM_{2.5}$-related mortality burden in China by 2030 under climate and population change scenarios: A modeling study, Environ, Int., 156,106733, https://doi.org/10.1016/j.envint.2021.106733, 2021.

Wang, Y., Hu, J., Zhu, J., Li, J., Qin, M., Liao, H., Chen, K., and Wang, M.: Health Burden and economic impacts attributed to $PM_{2.5}$ and O_3 in China from 2010 to 2050 under different representative concentration pathway scenarios, Resour. Conserv. Recy., 173, 105731, https://doi.org/10.1016/j.resconrec.2021.105731, 2021.

Dou, C., Ji, Z., Xiao, Y., Zhu, X., and Dong, W.: Projections of air pollution in northern China in the two RCPs scenarios, Remote Sens., 13, 3064, https://doi.org/10.3390/rs13163064, 2021.

Hong, C., Zhang, Q., Zhang, Y., Davis, S., Tong, D., Zheng, Y., Liu, Z., Guan, D., He, K., and Schellnhuber, H. J.: Impacts of climate change on future air quality and human health in China, PNAS, 116, 17193-17200, https://doi.org/10.1073/pnas.1812881116, 2019.

Finally, I am surprised that the abstract and conclusions still do not provide any quantitative data regarding how the different downscaling methods affect the outcomes inspected here. By including the extension to $PM_{2.5}$ I think the authors have done a good job of addressing my prior major concern (of this manuscript having no novelty when considered next to their existing work), but it would be helpful to include some high level conclusions regarding the degree to which model- (WRF) or statistics-based (ISIMIP) downscaling results in different or similar outcomes for health risks under different scenarios.

Reply: Thanks for your suggestions. we add some sentences in line 781 in the discussion.

There are some differences in projecting $PM_{2.5}$ concentration between WRF and ISIMIP methods. Compared to the 2010s reference, $PM_{2.5}$ concentration in ISIMIP are projected to decrease more than using WRF in G4 under the "mitigation" scenario during the 2060s over the Tianjin province (Fig. 11a, e). However, the spatial patterns of changes in $PM_{2.5}$ concentration between G4 and RCP4.5/8.5 under the "mitigation" scenario during 2060s are similar (Fig. 11c-d, g-h). This means that the effects of different downscaled methods on projecting $PM_{2.5}$ are small if we only consider the climate change alone without considering emissions changes. Due to the larger regression coefficient of emissions in the MLR under the ISIMIP method (Fig. S25, S26), the negative changes in $PM_{2.5}$ concentration are larger between "mitigation" and baseline under G4 during 2060s than that under the WRF method. Correspondingly, the ISIMIP method has a greater reduction in $PM_{2.5}$ related RR than WRF under three future climate scenarios during the 2060s.

We add the following sentences in line 29 in the abstract.

Compared with the 2010s, $PM_{2.5}$ concentration is projected to decrease 5.4 μg/m^3 in the Beijing-Tianjin province under the G4 scenario during the 2060s from the WRF downscaling, but decrease by 7.6 μg/m^3 using ISIMIP. The relative risk of 5 diseases decreases by 1.1%-6.7% in G4/RCP4.5/RCP8.5 using ISIMIP, but have smaller decrease (0.7%-5.2%) using WRF.

Minor comments

While I understand the authors' statement that health impacts only matter when people are affected, I still believe that line 270 ("Since health impacts are more important where there are more people") is likely to cause misunderstanding. I would recommend wording instead along the lines of "Since health impacts scale with the number of people affected". As written, it sounds like a single person's exposure is more important if they live in an urban rather than rural environment, when the intended meaning is instead (presumably) that an increase in concentration causes more health impact when a large number of people are exposed.

Reply: Done. We have changed the original sentence with that you suggested.

Since health impacts scale with the number of people affected,

Upon review, it appears that the Eastham et al. (2018) study does include limited meteorological effects (line 120). It would perhaps be more accurate to state that the study included only a first-order estimate of temperature and precipitation change.

Reply: Done. We rewrote the sentence in line 114.

However, this study included only a first-order estimate of temperature and precipitation change on $PM_{2.5}$ concentration under geoengineering, and also did not consider the situation in a highly polluted urban environment such as included in our domain, and which is typical of much of the developing world.

There remain some minor grammar and spelling errors (e.g. "statistically approach" on line 21, "gird" on line 334, "includes" should be "include" on line 326). Similarly, there is some confusing wording (e.g. "the ~1-2% wetter humidity has ~10% negative effect on decrease of $PM_{2.5}$" – the multiple negatives here make it difficult to understand whether increasing humidity is causing an increase or decrease in $PM_{2.5}$, "2.5" in $PM_{2.5}$ is not subscripted in line 944). These are rare but I would suggest the authors take another pass through the manuscript to clean up these few issues.

Reply: Done. We apologize for our errors, and we have rewritten the sentence in line 623 and make it clear. We also corrected all the subscripts of $PM_{2.5}$ in the manuscript.

The ~1-2% increase of humidity leads to ~10% increase of $PM_{2.5}$ concentration in the south of Beijing (Fig. 12g), and 0.2-0.3 m/s deceases of U-wind leads to 0-10% increase of $PM_{2.5}$ concentration in Zhangjiakou (Fig. 12h).

References

Cheng, W., MacMartin, D. G., Kravitz, B., Visioni, D., Bednarz, E. M., Xu, Y., Luo, Y., Huang, L., Hu, Y., Staten, P. W., Hitchcock, P., Moore, J. C., Guo, A., and Deng, X.: Changes in Hadley circulation and intertropical convergence zone under strategic stratospheric aerosol geoengineering, npj Climate and Atmospheric Science, 5, 1–11, 2022.

Fiore, A. M., Naik, V., and Leibensperger, E. M.: Air quality and climate connections, J. Air Waste Manag. Assoc., 65, 645–685, 2015.

Hong, C., Zhang, Q., Zhang, Y., and Schellnhuber, H. J.: Impacts of climate change on future air quality and human health in China, PNAS, 2019.

Jacob, D. J. and Winner, D. a.: Effect of climate change on air quality, Atmos. Environ., 43, 51–63, 2009.

Westervelt, D. M., Ma, C. T., He, M. Z., Fiore, A. M., Kinney, P. L., Kioumourtzoglou, M.-A., Wang, S., Xing, J., Ding, D., and Correa, G.: Mid-21st century ozone air quality and health burden in China under emissions scenarios and climate change, Environ. Res. Lett., 14, 074030, 2019.

Referee#2

This manuscript seeks to comprehend the changes in regional apparent temperature and $PM_{2.5}$ concentrations under the conditions of global warming and sulfate aerosol injection. This understanding is achieved through the utilization of data from multiple Earth System Model simulations, two downscaling methods, and two statistic linear regression functions. The topic is both significant and innovative. Nevertheless, several substantial concerns persist:

Reply: We would like to thank the referee for taking the time to review our manuscript again. Thanks for your positive response and constructive comments for our new manuscript. We have responded to your comments one by one.

The methodology employed to calculate $PM_{2.5}$ concentration only considers factors such as temperature, humidity, wind speed, and anthropogenic emissions. However, two critical elements have been overlooked: precipitation, and natural aerosol emission. Precipitation has a crucial role in 'cleansing' air pollutants, including $PM_{2.5}$, and future alterations in precipitation patterns could considerably influence regional $PM_{2.5}$ concentrations. Furthermore, natural aerosol emissions, such as dust and sea salt, constitute more than half of the average global $PM_{2.5}$. In regions like Beijing, "dust storms" are a significant air pollution phenomenon in the spring, contributing substantially to $PM_{2.5}$ levels. The absence of these two factors from the calculation or discussion makes the projected future changes in $PM_{2.5}$ unreliable.

Reply: We agree with the referee's concern of variables in our regression model. Actually, multiple meteorological factors are contributed to $PM_{2.5}$ concentration, such as temperature (You et al., 2017), humidity (Cheng et al., 2017), wind (Yin et al., 2017), precipitation (Guo et al., 2016), atmospheric pressure (Zhang et al., 2015), radiation (Chen et al., 2017) and planetary boundary layer height (Zheng et al., 2017) etc. Crudely, the dominant meteorological factors vary with areas. In our analysis, we did not apply all possible variables in our regression model, and we only considered the main meteorological factors in our domain. Chen et al (2020) pointed out that humidity and wind speed are the two dominant meteorological factors in the Jing-Jin-Ji region (which contains Beijing and Tianjin). Based on their study, we included temperature, humidity, as well as meridional and latitudinal winds into our regression model. Natural emission, such as "dust storms", also contributed to $PM_{2.5}$ concentration, but the composition of dust is complex, generally in Beijing bringing in coarser PM_{10} and not so much $PM_{2.5.}$ Furthermore, the sources of natural $PM_{2.5}$ include the sandstorms that sometimes occur usually in spring as extreme winds mobilize dry unvegetated soils. These extreme conditions are difficult to simulate in ESM and subject to land use policy e.g. the numerous ecosystem service measures undertaken by China over the last five decades (Miao et al.,2015). On the other hand, both HTAP_V3 and ECLIPSE V6b dataset do not offer the natural aerosol emission. So, there are some limitations in our study. In regard to precipitation, the ESM estimates of anomalies relative to historical are 24.0, 45.3, and 63.2 mm/year under G4, RCP4.5 and RCP8.5, respectively (Table S2). Among the four ESMs, no ESM shows significant changes, although differences are significant for the ensemble mean between RCP8.5 and 2010s.

We add some sentences in our manuscript.

We add the following sentences in line 260.

Many meteorological factors, such as temperature (You et al., 2017), precipitation (Guo et al., 2016), wind speed (Yin et al., 2017), radiation (Chen et al., 2017), planetary boundary layer height (Zheng et al., 2017) etc., can affect the $PM_{2.5}$ concentration. Their relative importance differs regionally. But here we consider only differences that are produced by the three scenarios, so for example we do not include precipitation in our analysis because none of the ESM simulate significant changes in our domain (Table S2).

We add the following sentences in line 810.

Our study did not consider the impacts of socio-economic pathways on $PM_{2.5}$ future emissions, instead we explore the meteorological differences between the SAI G4 scenario and the greenhouse gas RCP4.5/RCP8.5 on $PM_{2.5}$ concentrations. $PM_{2.5}$ emissions were defined by the uncontrolled ("baseline") and a scenario where technological intervention ("mitigation") reduces emissions. There are some limitations in our study. Firstly, the HTAP_V3 dataset only includes anthropogenic $PM_{2.5}$ emission, not natural $PM_{2.5}$ emission. Natural $PM_{2.5}$ will also change in the future under changing climate. The sources of natural $PM_{2.5}$ include the sandstorms that sometimes occur in spring as extreme winds mobilize dry unvegetated soils. These relatively extreme conditions are difficult to simulate in ESM and subject to land use policy e.g., the numerous ecosystem service measures undertaken by China over the last five decades (Miao et al.,2015). Secondly, although $PM_{2.5}$ concentration includes both primary and secondary $PM_{2.5}$ during model training, we do not consider the precursor gases for secondary $PM_{2.5}$ directly. The sensitivity of MLR may diminish at the high $PM_{2.5}$ values when secondary $PM_{2.5}$ dominates the variability of total $PM_{2.5}$ (Upadhyay et al., 2018). Thirdly, we only consider the effect of dominant near-surface meteorological variables on the $PM_{2.5}$. However, the vertical transport of pollutants related to vertical atmospheric stability should not be ignored (Lo et al., 2006; Wu et al., 2005), and this may contribute to the differences in RCP4.5 scenario from our MLR model and more sophisticated simulations (Fig. S7). Finally, although it is insignificant for the Beijing and Tianjin provinces, the MLR model suffers collinearity problems in some areas. These factors play smaller roles as we are mainly considering changes in $PM_{2.5}$ concentration between different climate scenarios. Nevertheless, projection for changes in $PM_{2.5}$ between SAI scenarios and per greenhouse gas scenarios would be valuable for global air quality impacts from geoengineering.

We add the following table in the supplementary information.

Table S2. The changes in annual mean precipitation (mm/year) between G4/RCP4.5/RCP8.5 during 2060s and references during 2010s over the domain. Bold indicates that differences are significant.

	G4-2010s	RCP4.5-2010s	RCP8.5-2010s
MIROC-ESM	73.1	50.5	51.8
MIROC-ESM-CHEM	-4.9	43.2	47.1
HadGEM2-ES	69.1	114.1	147.6
BNU-ESM	-41.6	-26.6	6.4
Ensemble	24.0	45.3	**63.2**

References

You, T., Wu, R., Huang, G., Fan, G.: Regional meteorological patterns for heavy pollution events in Beijing, J. Meteorolog. Res., 31, 597-611, https://doi.org/10.1007/s13351-017-6143-1, 2017.

Cheng, L., F, M., Chen, L., Jiang, T., Su, L.: Effects on the haze pollution from autumn crop residue burning over the Jing-Jin-Ji Region, China Environ. Sci., 37, 2801-2812, 2017.

Yin, Z., Wang, H., and Chen, H.: Understanding severe winter haze events in the North China Plain in 2014: roles of climate anomalies, Atmos. Chem. Phys., 17, 1641–1651, https://doi.org/10.5194/acp-17-1641-2017, 2017.

Guo, L., Zhang, Y., Lin, H., Zeng, W., Liu, T., Xiao, J., Rutherford, S., You, J., Ma, W.: The washout effects of rainfall on atmospheric particulate pollution in two Chinese cities, Environ. Pollut., 215, 195-202, https://doi.org/10.1016/j.envpol.2016.05.003, 2016.

Zhang, Y., Cao, F.: Fine particulate matter ($PM_{2.5}$) in china at a city level, Sci. Rep., 5, 14884, https://doi.org/10.1038/srep14884, 2015.

Chen, Z., Cai, J., Gao, B., Xu, B., Dai, S., He, B., Xie, X.: Detecting the causality influence of individual meteorological factors on local $PM_{2.5}$ concentrations in the Jing-Jin-Ji region, Sci. Rep., 7, 40735, https://doi.org/10.1038/srep40735, 2017.

Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., Chen, T., Wu, F., and Qiu, Y.: Analysis of influential factors for the relationship between $PM_{2.5}$ and AOD in Beijing, Atmos. Chem. Phys., 17, 13473–13489, https://doi.org/10.5194/acp-17-13473-2017, 2017.

Both apparent temperature and $PM_{2.5}$ calculations use a simple linear regression.

However, there exists a high correlation between the climate variables used, such as temperature and water vapor pressure/humidity. The uncertainties arising from this calculation method need to be addressed.

Reply: Yes, collinearity of variables is inevitable in our domain. The domination of the seasonal winter and summer monsoonal weather patterns mean that temperatures, precipitation and wind direction are all highly seasonal and correlated. In winter, precipitation is minimal and northerly winds predominate, in summer the opposite is true. However, the three fields are important in their own right since emission sources are essentially absent from the north, while temperature and humidity dominate aerosol microphysics. Furthermore, we used a widely used empirical formula to calculate the apparent temperature (Steadman 1984), that combines various meteorological fields.

We use the variance inflation factor (VIF) to test if there is excessive collinearity in our MLR model. Generally, if VIF value is greater than 10, there is collinearity problem between variables. As shown in figure S3 below, there are indeed collinearity problems in some areas. The problem doesn't occur in Beijing-Tianjin province, so there is no impact on the results for Beijing-Tianjin urban areas. To further explore the impact of collinearity on the results in high VIF grid cells, we compared the differences of $PM_{2.5}$ concentration in the future between removing factors with VIF greater than 10 and the full variables model (figure S4 and figure S5). Using ISIMIP downscaling, we only removed the temperature, while we removed the temperature and U-wind in the WRF method. In figure S4, we can see that $PM_{2.5}$ concentration show an increase of ~1 ug/m^2 in all ESMs under G4 with the "baseline" scenario (except HadGEM2-ES under ISIMIP method) after dealing the collinearity problem. In figure S5, $PM_{2.5}$ concentration has nearly 5-15 ug/m^2 decrease in all ESMs under G4 with "mitigation" scenario after dealing the collinearity problem. This means that $PM_{2.5}$ concentration has more sensitivity to the $PM_{2.5}$ emission after dealing the collinearity problem. The difference in $PM_{2.5}$ concentration between different scenarios with the removal of collinearity variables is shown in the following Figure 11, and that without removal of collinearity is shown in the Figure S18. The reductions in $PM_{2.5}$ between G4 and 2010s are a little higher in the area where there are collinearity problems after dealing the collinearity problem.

Although the absolute $PM_{2.5}$ concentration is different whether we consider collinearity or not, there are little differences in the changes of $PM_{2.5}$ concentration between G4 and 2010s/RCP4.5/RCP8.5. We also acknowledge that there are large uncertainties in $PM_{2.5}$ concentration in the future with considering collinearity or not. But in our study, we pay attention to the differences of $PM_{2.5}$ concentration between G4 and RCP4.5/8.5. So considering collinearity is not so important, and as shown there is no collinearity problem in Beijing-Tian province.
We rewrote the sentences in line 223.

We used a widely used empirical formula to calculate the apparent temperature

(Steadman 1984), that combines various meteorological fields, which also has been widely used to study heat waves, heat stress and temperature-related mortality.

We add the following sentences in line 281.

Collinearity of variables is inevitable in our domain. The domination of the seasonal winter and summer monsoonal weather patterns mean that temperatures, precipitation and wind direction are all highly seasonal and correlated. In winter, precipitation is minimal and northerly winds predominate, in summer the opposite is true. These three meteorological fields are important also important for emissions, since sources are essentially absent from the north, while temperature and humidity dominate aerosol microphysics.

We use the variance inflation factor (VIF) to test if there is excessive collinearity in our MLR models. Generally, if VIF value is greater than 10, there is collinearity problem between variables. Figure S3 shows that there are indeed collinearity problems in some areas, but not in Beijing-Tianjin province, so there is no impact on the results for the urban areas. We explored the impact of collinearity on the results in high VIF grid cells by removing factors with VIF greater than 10 and the full variables model (Fig. S4 and Fig. S5). Using ISIMIP downscaling, we only removed the temperature, while we removed the temperature and U-wind in the WRF method. $PM_{2.5}$ concentrations increased by ~1 μg/m^2 in all ESMs under G4 with the "baseline" scenario (Fig. S4), in contrast, $PM_{2.5}$ concentrations decreased by 5-15 μg/m^2 with the "mitigation" scenario (Fig. S5) after dealing the collinearity problem. This means that $PM_{2.5}$ concentration has more sensitivity to the $PM_{2.5}$ emission after accounting for collinearity. Although the absolute $PM_{2.5}$ concentrations are different accounting for collinearity, there are no significant differences in the changes of $PM_{2.5}$ concentration between G4 and the 2010s/RCP4.5/RCP8.5 (Fig.11, Fig. S18).

We reploted the following figures in the manuscript.

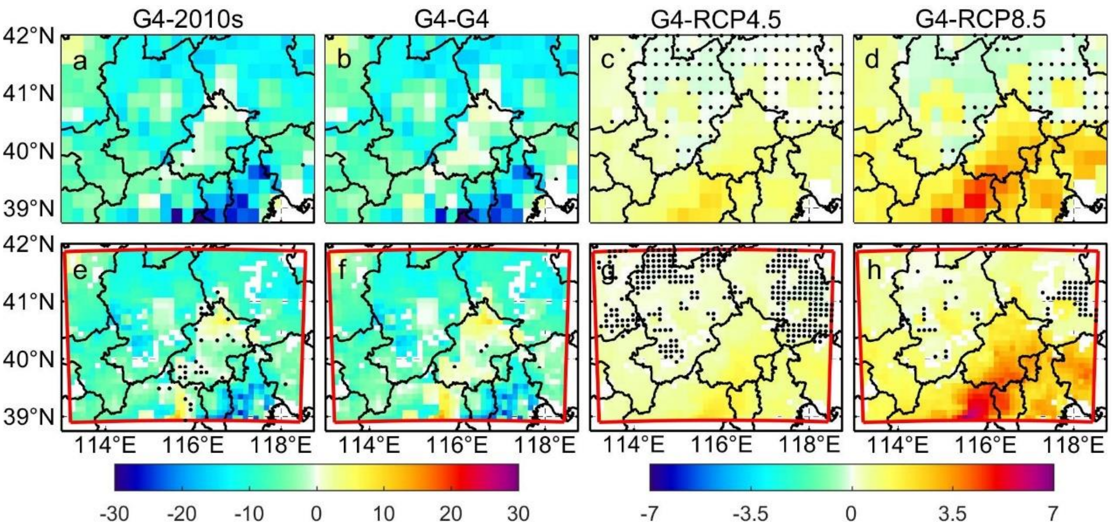

Figure 11. Spatial patterns of ensemble mean $PM_{2.5}$ concentration difference (μg/m^3) between

"mitigation" under G4 in the 2060s and reference **(a, e)**, between "mitigation" and "baseline" under G4 in the 2060s **(b, f)**, between G4 and RCP4.5 under "mitigation" scenario in the 2060s **(c, g)**, and between G4 and RCP8.5 under "mitigation" scenario in the 2060s **(d, h)** based on ISIMIP **(a-d)** and WRF **(e-h)** results. Excessive collinearity variables have been removed (Fig. S18 shows the results without this procedure). Stippling indicates grid points where differences or changes are not significant at the 5% significant level according to the Wilcoxon signed rank test.

We added the following figures in the supplementary information.

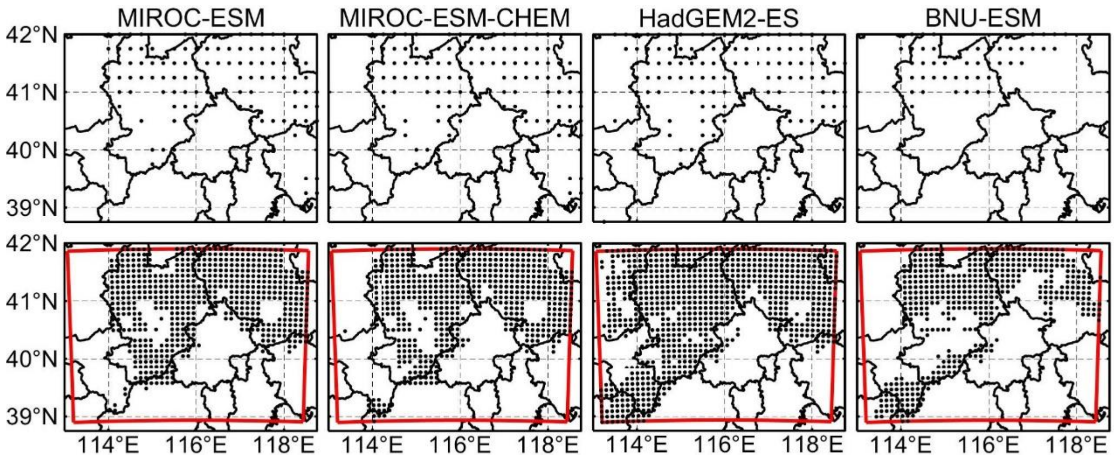

Figure S3. Variance inflation factor (VIF) test of excessive collinearity in our MLR model. VIF >10 means there is collinearity problem between variables (dotted regions).

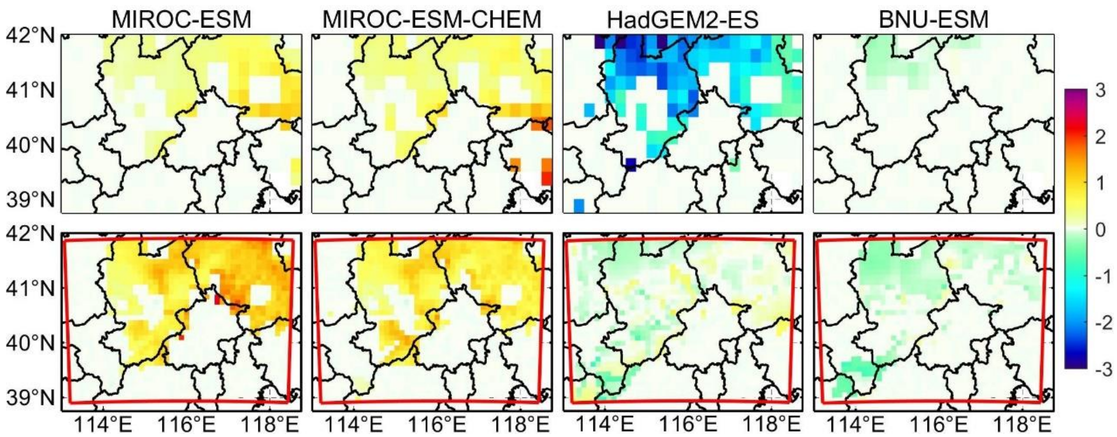

Figure S4. Difference in $PM_{2.5}$ concentration under G4 with "baseline" scenario in 2060s between removing factors with VIF greater than 10 and the full variables model.

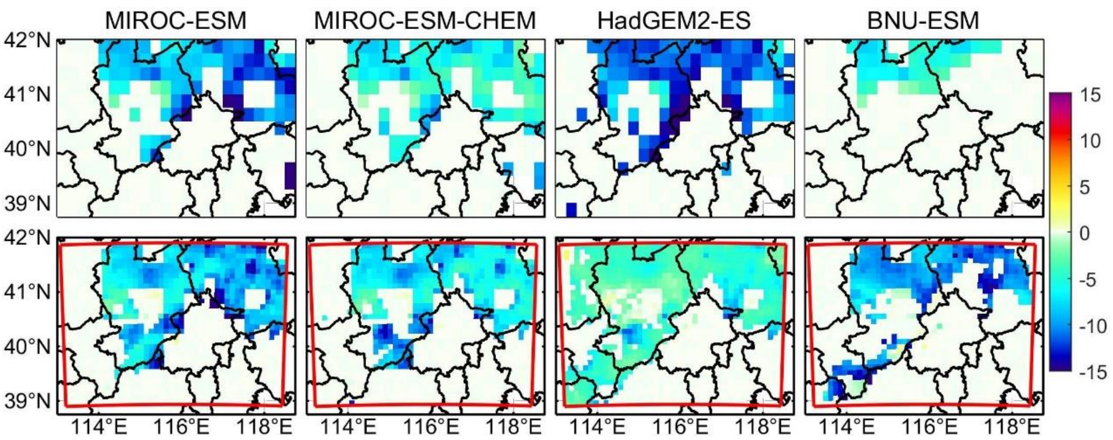

Figure S5. Difference in $PM_{2.5}$ concentration under G4 with "mitigation" scenario in 2060s between removing factors with VIF greater than 10 and the full variables model.

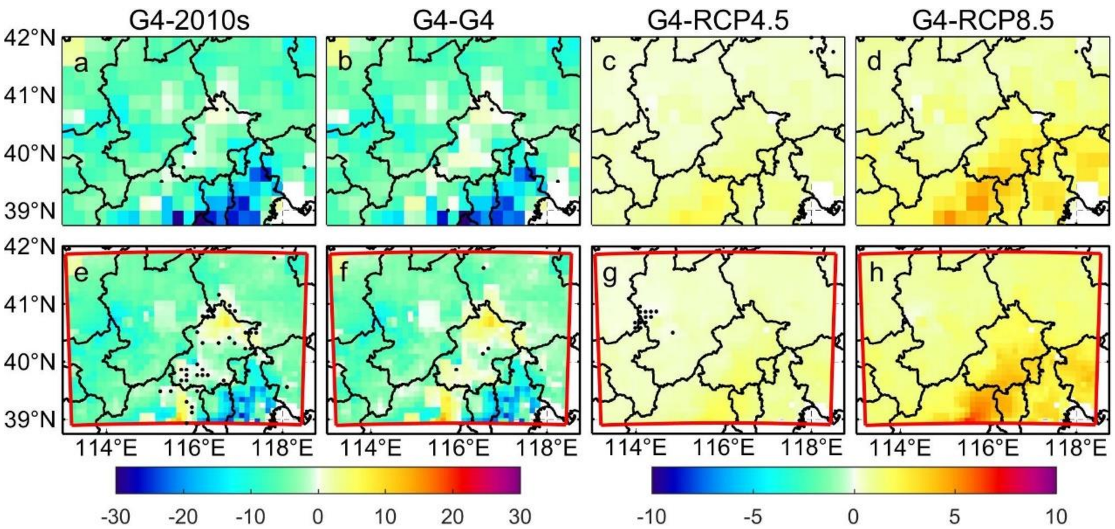

Figure S18. Same as figure 11, but the results of all variables in MLR.

We updated Fig.2, Fig.11-14, Table S2, Table S3, Fig.S3, Fig.S13-S14 and Fig. S19-S21 after removing the collinearity variables in the areas with VIF>10 in the original unrevised manuscript. We have also revised the sentences in the manuscript and the numbering of figures accordingly, and overall, the changes are not very significant.

In the discussion section, the authors declare, "If we consider the aerosol deposition under G4 scenarios, $PM_{2.5}$ concentration will be 0-1 μg/m3 higher than that without due to deposition of the SAI aerosols (Fig. S21)." This is incorrect. The injected sulfate aerosol would primarily deposit in the coarse mode and would not augment SO_4 in $PM_{2.5}$ compared to the reference case during the same period.

Reply: The referee gives no support for the assertion that the numbers we calculate are incorrect. This concerns the deposition from the SAI as $PM_{2.5}$. Eastham et al. (2018) considered this with a much more sophisticated treatment than available to us. They concluded that 1/25 of the SAI was deposited as $PM_{2.5}$. This is the ratio we use, and since it is the only study to simulate the effects, we will continue to use this number.

Lastly, the abstract lacks clarity in terms of the study's conclusions. How does $PM_{2.5}$ change under future climate conditions and sulfate aerosol injection? What is the influence of the two downscaling methods on studying the health impact of SAI?
It is better to use climate intervention instead of geoengineering.

Reply: Thanks for your comments. We changed the stratospheric aerosol injection to stratospheric aerosol intervention. We are limited in the number of words in the abstract. We add some sentences in the abstract.

Compared with the 2010s, $PM_{2.5}$ concentration is projected to decrease 5.4 μg/m^3 in the Beijing-Tianjin province under the G4 scenario during the 2060s from the WRF downscaling, but decrease by 7.6 μg/m^3 using ISIMIP. The relative risk of 5 diseases decreases by 1.1%-6.7% in G4/RCP4.5/RCP8.5 using ISIMIP, but have smaller decrease (0.7%-5.2%) using WRF.