Reply to Referee #1:

The authors thank to referee #1 for those comments and suggestions, which have greatly helped us to improve our manuscript. Below, we reply point-by-point, highlighting the changes we have implemented. In the following, the texts with *italic font* are your original comments, and the texts with normal font are our responses.

This paper investigated how future changes in anthropogenic aerosol emissions, determined by projected future scenarios, would impact on the frequency and intensity of haze events in Beijing. The results show that in the scenario of stringent aerosol emission reductions, the frequency of the mechanisms conducive to such haze events increases but the severity of the resulting haze events decreases. The meaningful results are obtained using model simulations of two future scenarios (current legislation emissions and maximum technically feasible aerosol reductions) from two fully coupled climate models, compared against a historical simulation. The methodology and analysis is clear; the results are comprehensive and the resulting conclusions are of real value. Although generally well written, there are numerous grammatical errors that make the manuscript, particularly in the results section, harder to understand. It would benefit from another thorough re-read to address these errors. Some of the errors/typos are highlighted in the technical comments, but not where rewording is required. After this is addressed, along with the following comments, I would recommend accepting this paper by Atmospheric Chemistry and Physics.

Response: Thank you very much for the positive comments. We have addressed all your comments according to your comments and suggestions. All of the detailed responses can be seen as following.

1) It is not clear in the explanation for how a haze event has been defined (based upon a HWI-month threshold of 1), why a higher value was not chosen initially. For example, selecting the threshold at HWI-month 2 would incorporate a greater percentage of days with HWI-daily > 0. A brief explanation of why a value greater than 1.0 was not chosen would be helpful. This would also help the reader understand the results paragraph pertaining to Figure S7.

Response: Thanks for your question. You're right, higher threshold of HWI-month can incorporate a greater percentage of days with HWI-daily>0. Because HWI is a large-scale circulation proxy of haze events, it is better to use observational haze data to check the reliability of using HWI-month as the haze-favorable circulation index. This is also one of the main concerns of Referee#2. Thus, we investigated the relationship between HWI-month and observed frequency and visibility of haze events, and expanded discussion of this in the manuscript in the revised manuscript.

We use observed daily visibility, relative humidity and wind speed from 1974 to 2013 from the National Climatic Data Center (NCDC) Global Surface Summary of the Day (GSOD) database (Fig.A1a). Haze days are defined as days with daily visibility less than 10km, relative humidity at less than 90% and surface wind speed less than 7 m s^{-1} (Chen and Wang, 2015). The observed haze frequency is the number of haze days, and observed haze intensity is the minimum 3-day consecutive visibility (VN3day). The climatological winter mean haze days and VN3day can be seen in Fig.A1b-c.

We investigated the frequency and intensity of haze events associated with different thresholds of HWI-month (Fig.A2-3). The positive values of HWI-month are associated with more haze days and less visibility over Beijing and the surrounding region (box in Fig.A2-3), indicating the reliability of using positive HWI-month to stand for haze-favorable circulation condition (box region in Fig.A2-3). Comparing the anomaly of haze days associated with different thresholds of HWI-month, no significant change is found when HWI-month increases from 1.0 to 1.5. However, the haze days when HWI-month≥2.0 are fewer than those when HWI-month≥1 and 1.5. This is partly due to the small sample size of months with HWI≥2.0. The greater decrease in VN3day with higher HWI-month is shown in Fig.A3. It indicates a good relationship between HWI-month and observed haze intensity.

In summary, we proved the reliability of using HWI-month as the haze-favorable atmospheric circulation conditions surrounding Beijing based on observed haze dataset for 1974-2013. The threshold of 1.0 shows comparable results with the threshold 1.5. Since the sample size of HWI-month≥2.0 is small, we used HWI-month≥1 as the threshold of haze events. Most importantly, the choice of threshold of HWI does not

affect our main results, as shown in the PDF and CDF distributions of HWI (Fig.4 in the revised manuscript).

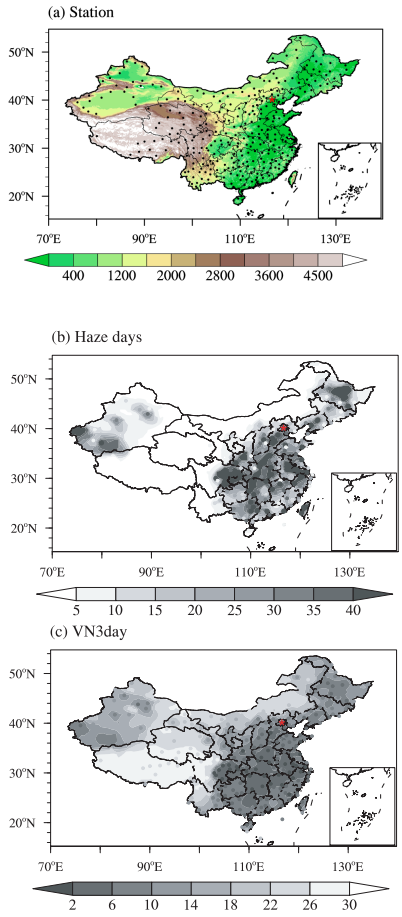

Fig.A1 (a) Distribution of the observational stations (dots) in NCDC GSOD. Shading indicates the altitude of the topography (unit: m). (b) Spatial distribution of climate winter mean number of haze days (unit: days) of 1974-2013 by gridding stations on to a 0.5×0.5° grid point. (c) Same as (b), but for VN3day (unit: km). The red dot denotes the location of Beijing.

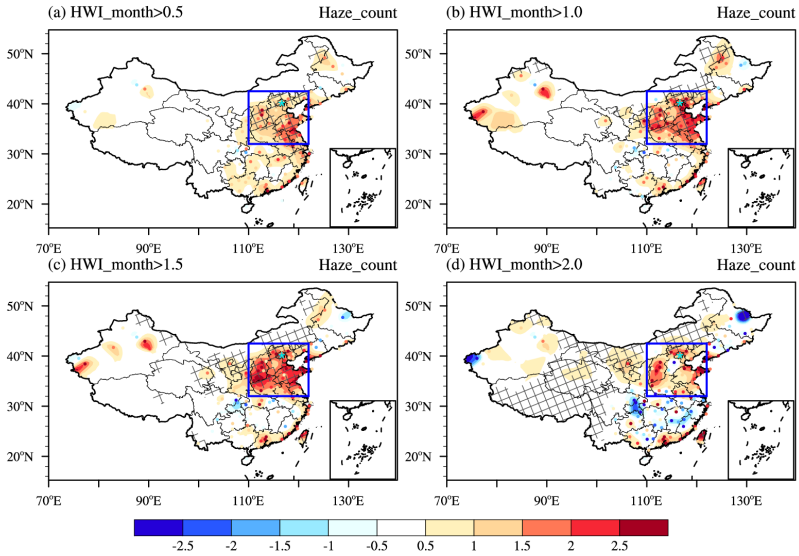

Fig.A2 Anomalies of haze occurrence (unit: days) when (a) HWI-month≥ 0.5, (b) HWI-month≥1, (c) HWI-month≥1.5 and (d) HWI-month≥2.0. The hatched areas are statistically significant at the 5% level based on a Student's t-test. The box indicates the North China region where the haze occurrence and haze intensity are highly correlated with HWI-month.

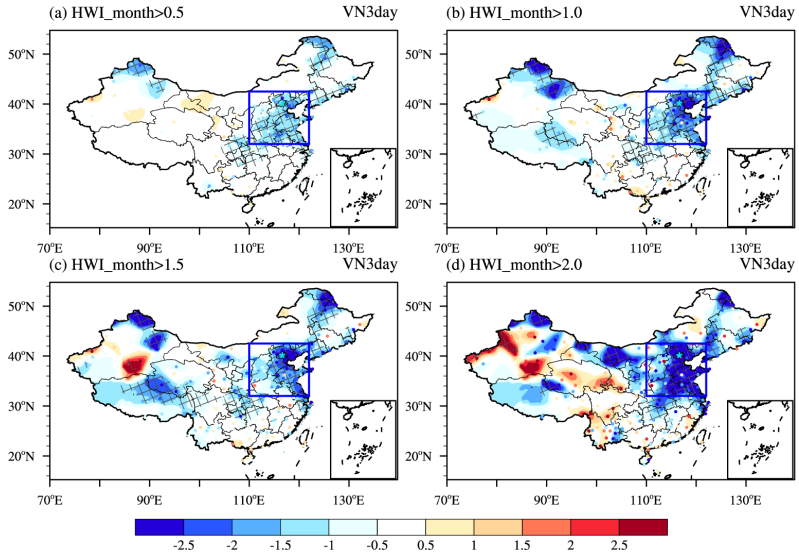

Fig.A3 Same as Fig.A2, but for VN3day anomalies (unit: km).

The corresponding revision is as follows:

"We use observed daily visibility, relative humidity and wind speed from 1974 to 2013 from the National Climatic Data Center (NCDC) Global Surface Summary of the Day (GSOD) database (Fig.S1a). Haze days are defined as days with daily visibility less than 10km, relative humidity less than 90% and surface wind speed less than 7m s-1 (Chen and Wang, 2015). The observed haze occurrence is the number of haze days, and observed haze intensity is defined as the minimum 3-day consecutive visibility (VN3day). Spatial distributions of winter mean haze occurrence and VN3day are shown in Fig.S1b-c." (Lines 131-138 in P6-7 of the revised manuscript)

"We also checked the observed winter haze occurrence and intensity (VN3day) anomalies when HWI-month ≥1. More haze occurrence and reduced visibility are observed over North China, indicating the reliability of using HWI-month≥ 1 as a proxy of the favorable climatic conditions for the haze events in Beijing and the surrounding region. The selection of a higher threshold of HWI-month (e.g. 1.5) does not make a great difference to our results (not shown)." Please see (P12 L247-252 in the revised manuscript).

2) The description of the results shown in Figures 12 is delivered in a rather confusing way, and not consistently between the main body and the figure caption. A more thorough description on what is being shown is needed.

Response: Thanks for your comments. To better illustrate the main result, we deleted this figure in the revised manuscript, and added the PDF and CDF distributions of AOD at 550nm surrounding the Beijing in the months with HWI≥1 (Fig.A4 here, also Fig.11 of the revised manuscript). The corresponding revision is as follows:

"To account for model differences in historical AOD, we used the ratio of AOD at 550nm (hereafter AOD_ratio) relative to a baseline winter mean to represent the air-pollution severity. When AOD_ratio is greater than 1.0, the air-pollution intensity is higher than baseline climate mean." (P19 L385-389 in the revised manuscript)

"We calculated the PDF distributions of AOD_ratio surrounding the Beijing region (box region in Fig.2) in the months with HWI≥1 in His, CLE and MTFR (Fig.11). In His, the area-averaged AOD_ratio around the Beijing region when HWI≥1 is elevated to 1.34 (1.26) times of the baseline climate mean in HadGEM-GC2 (GFDL-CM3)

(Fig11.a-b). The change in AOD_ratio with HWI≥1 under CLE relative to His is different between the two models. It increases to 1.51 in HadGEM3-GC2 but decreases to 1.13 in GFDL-GC3. As expected, the AOD_ratio with HWI≥1 in MTFR reduces in both models due to the dramatic reduction in anthropogenic aerosols. Thus, the mean air-pollution intensity with the favorable circulation conditions for haze under MTFR will be greatly relieved." (P20 L407-416 in the revised manuscript)

"To check whether extreme air pollution events would still occur, the probability of AOD_ratio when HWI≥1 in the three scenarios are examined (Fig.11b, d). In this study, the mean AOD_ratio across all months when HWI≥1 in His is regarded as the winter mean intensity of baseline haze events, i.e., the grey vertical lines in Fig.11a, c. The probability of haze event intensity exceeding this threshold is about 42% and 34% in HadGEM3-GC2 and GFDL-CM3, respectively (Fig.11b, d). Under CLE, it increases to 52% in HadGEM3-GC2 while decreases to 28% in GFDL-CM3, consistent with Fig.10a, c. In MTFR, lower probability is projected in both models, 24% in HadGEM-GC2, and 21% in GFDL-CM3. This demonstrates that severe events (i.e., higher AOD_ratio) would still happen in MTFR albeit with dramatic reduction in anthropogenic aerosol, even though the mean intensity of haze events themselves will become less dangerous if aerosol emissions are reduced." (P20-21 L422-433 in the revised manuscript)

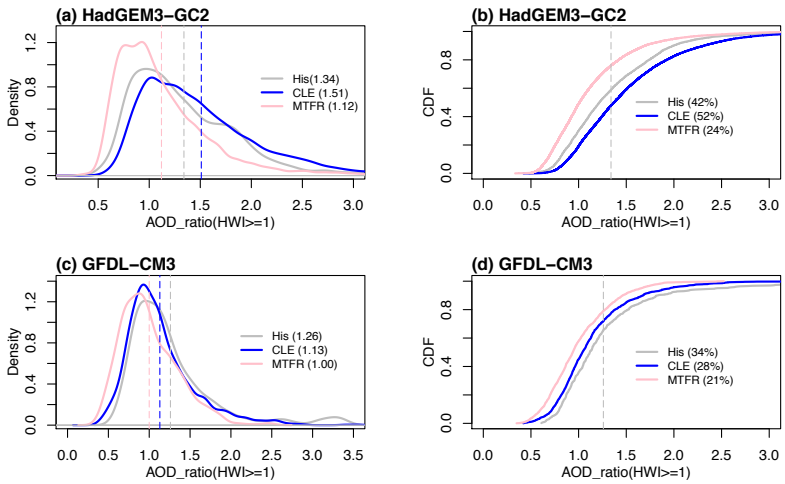

Fig.A4 (a) PDF and (b) CDF distributions of AOD_ratio(HWI≥1) over North China (33-45°N, 105-122°E, box in Fig.2) in HadGEM3-GC2. (c)-(d) are the results from GFDL-CM3. The grey, blue and pink vertical lines and numbers in (a) and (c) are the

winter mean AOD_ratio(HWI≥1) of His, CLE and MTFR, respectively. The numbers in (b) and (d) are the cumulative probability of AOD_ratio(HWI≥1) higher than the winter mean AOD_ratio(HWI≥1) of His. (refer Fig.11 in the revised manuscript)

Technical comments:

The following references are cited, but not in the reference list - please include if these are the correct references (please also check the rest of the reference list is accurate with the manuscript):

Liu, C., Zhang, F., Miao, L., Lei, Y. & Yang, Q. Future haze events in Beijing, China: When climate warms by 1.5 and 2.0C. Int. J. Climatol. 40, 3689–3700 (2019).

Liu, Z. et al. A Model Investigation of Aerosol Induced Changes in the East Asian Winter Monsoon. Geophys. Res. Lett. 46, 10186–10195 (2019).

Response: Added as suggested and all references are checked.

In some of the figures the location of Beijing is marked by a green dot. I recommend having this green dot displayed on all the figures where Beijing is shown.

Response: Added as suggested. Please see the figures in the revised manuscript.

Figures 5-7 – It would be useful to define within the captions the extent of the boxes (Fig 6 and 7) and lines (Fig 5)

Response: Boxes and lines are added into Fig.5-Fig.7 in the revised manuscript.

Figure 9 – hard to visually judge the amount of change between His, CLE and MTFR –this could do with a statistical value to complement the histograms to aid interpretation. Please add.

Response: Revised as suggested.

L58 - Citation needed

Response: Done.

L84 – An et al., 2019 – change from 2015

Response: Done.

L92 – Citation needed

Response: Citation was added is the end of this sentence.

L142 – Please clarify whether the year 2015 is correct. Table S1 states HadGEM3-GC2 historical run goes to 2014.

Response: Changed to 2014 in the revised manuscript.

L160 – include "(not shown)" after India for clarity – as India isn't shown in Fig. S1a.

Response: Done.

L161 – same as comment L160

Response: Done.

L210 – typo – should this be HWI greater than 0.0?

Response: Yes. Corrected in the revised manuscript.

L215 – Missing words

Response: Revised.

L244 – Change 500hPa to 850hPa

Response: Corrected.

Figure S4 – Denote what the green boxes show in the caption

Response: Added.

L256 – The mean for the historical period (His) is quoted over the years (1980-2014), however every other instance the baseline/His period is stated to be 1980-2004, e.g L156. Additionally, in Figure 3a, the grey data runs to 2014, not 2004. This is confusing and needs to be clarified.

Response: Revised it to 1980-2004.

L275 – For consistency should be 3 significant figures, like L273, so change to 7.1% and 7.3%.

Response: Done.

L277 – What does 'increase in atmospheric circulation patterns' mean? Please clarify

Response: This sentence is deleted in the revised manuscript.

L288 – What are the mean values? Could these be included somewhere on Figure 4?

Response: Done. We added the mean values in the PDF plots. We moved this plot to the supplementary file (Fig.S6) in the revised manuscript to shorten the paper length.

L309-310 – Sentence doesn't make sense. It refers to the difference between the 2 experiments (MTFR – CLE, not CLE – His) but this isn't stated. Also Fig.S5 c,d refer to 850hPa. Please clarify.

Response: Clarified as suggested.

L326 – should be 'MTFR v His' rather than 'MTFR vs CLE' as the anomalies over the North Pacific show less of a change between the two plots 6a and b. Please clarify.

Response: Revised.

L342 – typo

Response: Revised.

L364-365 – Siberia is SLP1 – please clarify.

Response: Clarified.

L390 – Change to Fig.11a-b

Response: Changed.

L393 – Include reference to Fig.11b-d

Response: Revised.

L665 – replace (a) with (b)

Response: Previous (b) is deleted. We have revised the figure caption.

L715 – (c)-(d) needs amending to (b)-(d)

Response: Revised.