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).

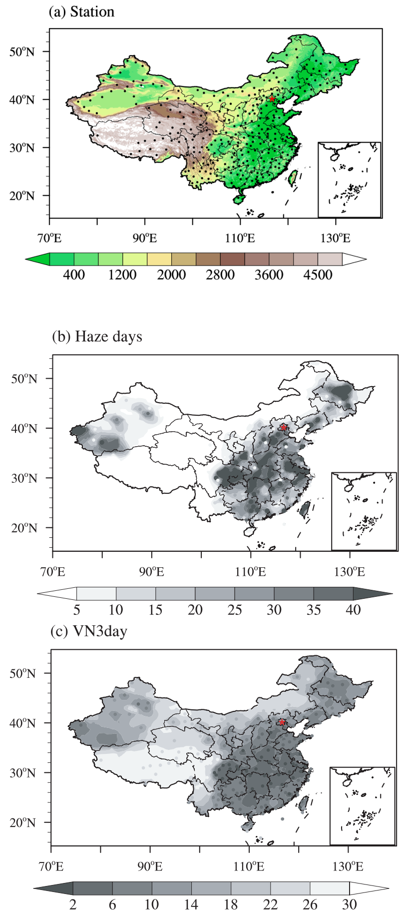

Fig.A1 (a) Distribution of the observational stations (dots) in NCDC GSOD. Shading indicates the altitude of the topography (unit: m). (b) Spatial distribution of climate winter mean number of haze days (unit: days) of 1974-2013 by gridding stations on to a 0.5×0.5° grid point. (c) Same as (b), but for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

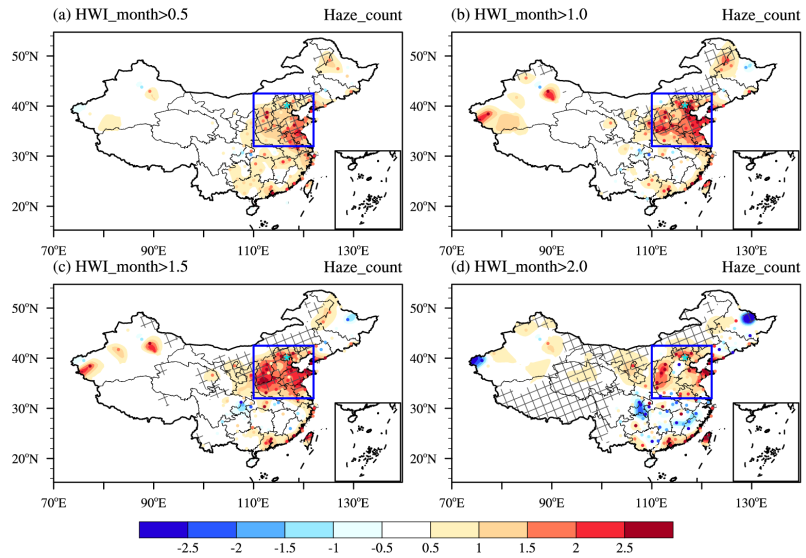

Fig.A2 Anomalies of haze occurrence (unit: days) when (a) HWI-month≥ 0.5, (b) HWI-month≥1, (c) HWI-month≥1.5 and (d) HWI-month≥2.0. The hatched areas are statistically significant at the 5% level based on a Student's t-test. The box indicates the North China region where the haze occurrence and haze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

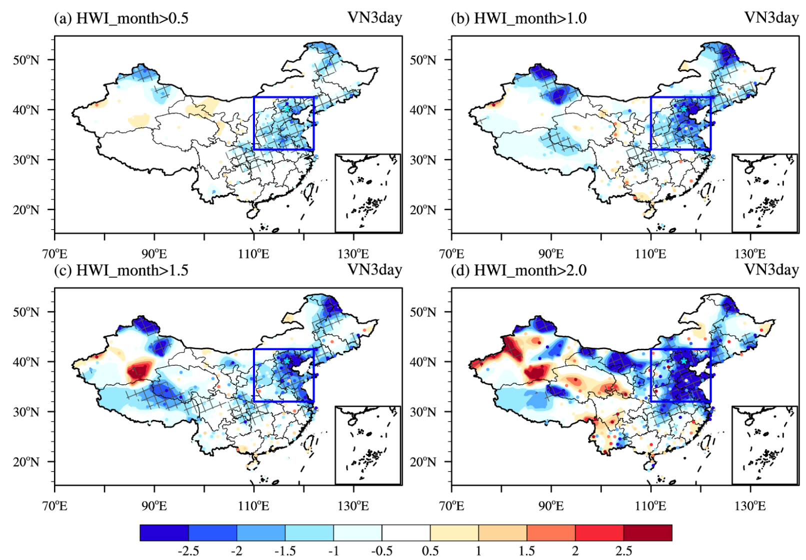

Fig.A3 Same as Fig.A2, but for VN3day anomalies (unit: km).

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)

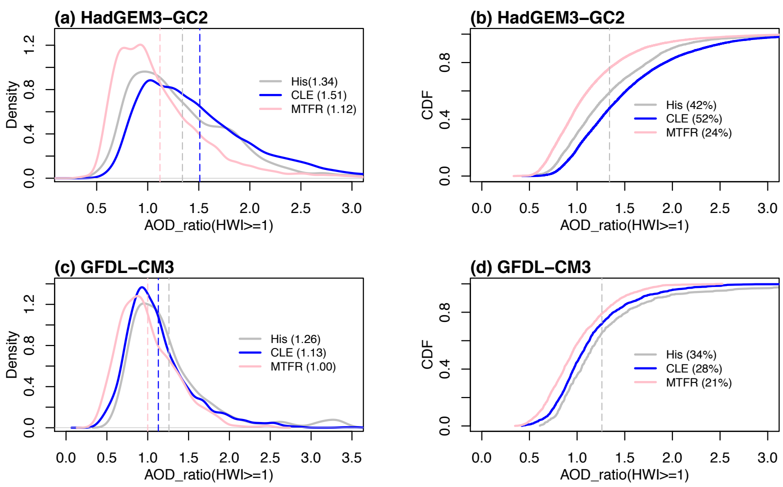

Fig.A4 (a) PDF and (b) CDF distributions of AOD_ratio(HWI≥1) over North China (33-45°N, 105-122°E, box in Fig.2) in HadGEM3-GC2. (c)-(d) are the results from GFDL-CM3. The grey, blue and pink vertical lines and numbers in (a) and (c) are the winter mean AOD_ratio(HWI≥1) of His, CLE and MTFR, respectively.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

Responses to Referee #2:

Summary: *The manuscript presents aerosol forcing sensitivity simulations from 2GCMs. It focuses on the China, and asks what effects changes in aerosols may have on future air pollution/haze events via their impact on circulation change. It is found that reductions in aerosols promote circulation patterns associated with haze events, however the intensity of such events is decreased due to the reduction of the particulate load.*

Recommendation: *Air quality in China is of major concern to public health officials, impacted citizens, and environmental scientists. The focus on the influence of aerosols on meteorology is novel and interesting. Generally speaking, projections of future air quality over China have primarily focused on the influence of GHGs on meteorological conditions. In the below I suggest greater engagement on the authors' parts with the subject of internal variability and its potential ramifications for the result presented here and the robustness thereof.*

Response: Thank you for your insightful comments and detailed instruction on how to improve the manuscript. The quality of the manuscript has been greatly improved based on your comments. In the following, the texts with *italic font* are your original comments, and the texts with normal font are our responses. Below, we reply point-by-point, highlighting the changes we have implemented.

Specific Comments:

- On Line 140 the authors indicate 2 models are used to assess the robustness of results. However, some modeling groups have made available single forcing large ensembles, including simulations in which the single forcing is aerosols. Both CESM and CanESM have made available these data sets. This is a big ask, but the claims of the paper are substantial and require rigorous testing. I would have much more confidence in the claims presented here, if the authors were to test their hypothesis using these data:

– Relevant CESM publication: Deser et al 2020 https://journals.ametsoc.org/jcli/article/33/18/7835/353234/Isolating-the-Evolving-

Contributions-of – Relevant CanESM publications: —Swart et al 2019 https://gmd.copernicus.org/articles/12/4823/2019/ —Santer et al 2019 https://www.pnas.org/content/116/40/19821

Responses: Thanks for recommending those individual anthropogenic forcing large ensemble simulations from CESM, CanESM2 and CanESM5. This study aims to estimate changes in Beijing haze events under different anthropogenic aerosol emission scenarios in the future. Thus, we used two climate models forced by different anthropogenic aerosol forcings but the same greenhouse gas emissions following the RCP4.5 scenario. However, the experiment designs for single forcing large ensemble simulations are different from this study. The relative role of individual anthropogenic forcing can be estimated from the single forcing simulations, while the impact of different anthropogenic forcing scenarios can not be estimated. In addition, the greenhouse gases scenario in CESM and CanESM2 are under the RCP8.5 scenario, different from this study as well. Thus, the single forcing large ensemble simulations cannot be used here to answer the scientific questions of this study. So, we didn't use those single forcing experiments in the revised manuscript. We also agree that the single forcing experiments provide good database to test the relative role GHG, AA and internal variability in HWI changes both in historical and future projection. This is discussed in the end of the revised manuscript as follows (P23 L484-487).

"But internal variability may not be fully sampled because of the limited number of realizations and models used in this study. In the future, single forcing experiments and large ensembles simulations are useful ways to confirm the relative role of greenhouse gases and anthropogenic aerosol forcing on haze event."

- I am likewise concerned with the authors' lack of consideration/discussion of model simulated internal variability. Beijing's haze events have received a lot of attention, to include work done by researchers that have articulated the role of internal variability on past (Zhang et al 2020 https://acp.copernicus.org/articles/20/12211/2020/; Callahan et al 2019 https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029738) and future

(Callahan & Mankin 2020 https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020GL088548) air quality-relevant circulation changes. I would like to see the authors contextualize their results in light of the findings presented in these papers. Do the sample sizes studied in this manuscript approach those required to account for signals rising above noise?

Response: Thanks for this insightful comment. To answer this questions, we used two methods to estimate internal variability, then tested the significance of the changes of mean HWI and the frequency with HWI≥1 under the two scenarios. Both methods support the robustness of our previous results.

Method 1: A bootstrapping approach. We estimated internal variability by performing bootstrapped samples. This resampling-based procedure involves three steps. First, we randomly select 75-month (135-month), i.e. 25-yr (45-yr) winters, from His (projections), and calculate the mean HWI change of the 75-month relative to His or frequency of month with HWI≥1 of the 75-month. The 75-month and 135-month are selected to mimic any 25-yr in the period 1980-2004 and 45-yr in 2016–2050, respectively; We repeat the first step 2000 times, and the 2000 bootstrapped samples can be viewed as internal variability of His or future projections. We then compare the results of ensemble mean of each model with those of the 2000 bootstrapped samples. If it falls outside the top 5% of the distribution, we then claim that the projected changes in mean HWI or frequency of month with HWI≥1 are statistically significant at the 5% level and beyond the variability of internal variability. (P10-11 L210-222 in the revised manuscript).

The difference in mean HWI between CLE vs His, MTFR vs His, and CLE vs MTFR, are also statistically significant at the 5% level in both models (Fig.A1c-d here, Fig.4a-b in the revised manuscript). The frequency with HWI⩾1 in CLE and MTFR are both statistically different from that in His in the two models, while only in HadGEM3-GC2 simulations is the frequency in MTFR statistically significant from that in CLE at the 5% level (Fig.A1c-d here, Fig.4c-d in the revised manuscript). (P14 L291-296 in the revised manuscript)

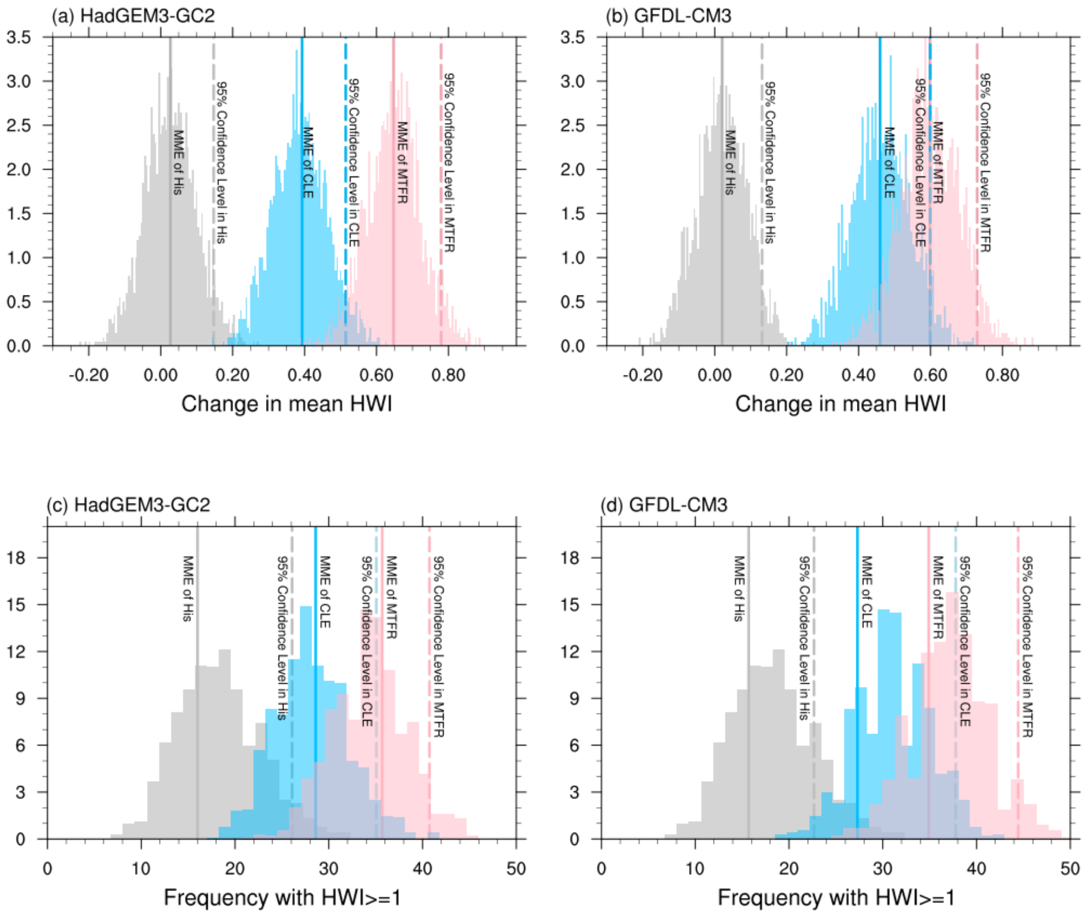

Fig.A1 Histogram plots of (a) changes in winter mean HWI, and (c) frequency of month with HWI≥1 in HadGEM3-GC2 for the 2000 bootstrapped samples, and (b), (d) similarly for GFDL-CM3. The grey, blue and pink shadings are the results estimated from His, CLE and MTFR respectively. Solid (dashed) grey, blue and pink lines are the results of multi-member mean (95% confidence level) in His, CLE and MTFR, respectively. (Fig.4 in the revised manuscript)

Method2: We used historical simulations of multi-model output from CMIP5 to estimate internal variability of His. 108 realizations of CMIP5 for the period 1961-2005 are used in total (model list is shown in Table A1). Internal variability of changes in mean HWI is estimated as ranges of the difference between any running 25-year mean HWI and the baseline (1985-2004) of each model. Internal variability of frequency of month with HWI≥1 is the range of frequency of month with HWI≥1 in any running 25-year in historical simulation. The results are shown in Fig.A2. The range of internal variability estimated from method 2 is larger than that from method 1, as it also includes a degree of structural uncertainty between the CMIP5 models. As for the winter mean

HWI changes and frequency of month with ≥1 in CLE and MTFR projected by HadGEM3-GC2 and GFDL-CM3, they are also statistically significant at the 5% level using this method, consistent with *Method 1*. A shortcoming of *Method 2* is that it cannot estimate internal variability of future projection. Thus, we used *Method1* in the revised manuscript rather than assuming that internal variability remains the same in future.

Table A1. List of CMIP5 models used to estimate internal variability

Model	Institute, country	Atmosphere Resolution (lat×lon)	realizations
ACCESS1-0	CSIRO-BOM, Australia	1.2°×1.9°	1
bcc-csm1-1-m	BCC, China	2.8°×2.8°	3
bcc-csm1-1	BCC, China	1.1°×1.1°	3
BNU-ESM	BNU, China	2.8°×2.8°	1
CanESM2	CCCma, Canada	2.8°×2.8°	5
CanCM4	CCCma, Canada	2.8°×2.8°	10
CCSM4	NCAR, USA	0.9°×1.2°	8
CESM1-CAM5	NSF-DOE-NCAR, USA	0.9°×1.2°	3
CMCC-CN	CMCC, Italy	3.7°×3.7°	1
CNRM-CM5	CNRM, France	1.4°×1.4°	10
EC-EARTH	EC-EARTH, Europe	1.1°×1.1°	6
FIO-ESM	FIO, SOA, China	2.8°×2.8°	3
GFDL-CM3	NOAA/GFDL. USA	2.0°×2.5°	5
HadCM3	MOHC, UK	2.5°×3.7°	10
HadGEM2-CC	MOHC, UK	1.2°×1.9°	3
HadGEM2-ES	MOHC, UK	1.2°×1.9°	4
inmcm4	INM, Russia	1.5°×2.0°	1
IPSL-CM5A-LR	IPSL, France	1.9°×3.7°	6
IPSL-CM5A-MR	IPSL, France	1.4°×2.5°	3
MIROC5	AORI-NIES-JAMSTEC, Japan	1.4°×1.4°	5
MIROC-ESM	AORI-NIES-JAMSTEC, Japan	2.8°×2.8°	3
MPI-ESM-MR	MPI-M, Germany	1.9°×1.9°	6
MRI-CGCM3	MRI, Japan	1.1°×1.1°	5

NorESM1-M	NCC, Norway	1.9°×2.5°	3

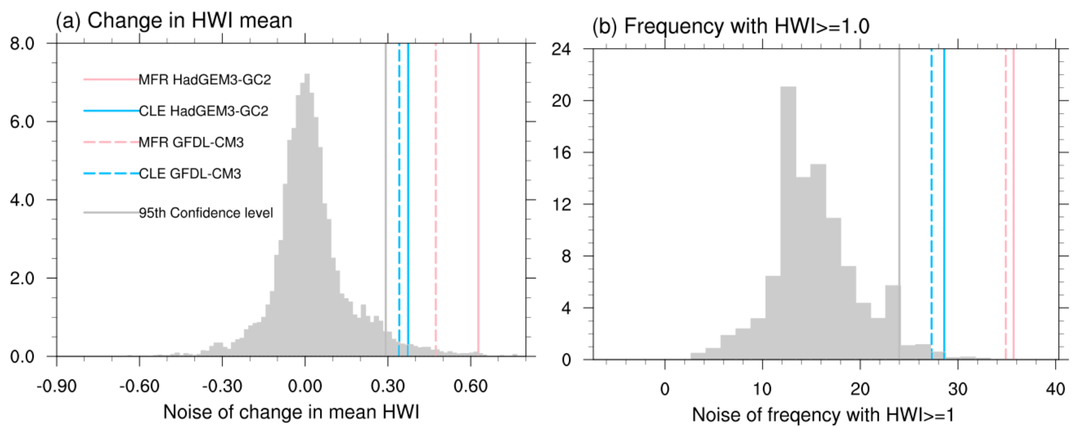

Fig.A2 Histogram plots for (a) winter mean HWI change and (b) frequency of month with HWI≥1 in His estimated from 108 realizations of CMIP5 multi-model output. The solid grey lines are the 95% confidence level in His. The solid (dashed) blue and pink lines are the results from CLE and MTFR simulated by HadGEM3-GC2 (GFDL-CM3), respectively.

- The presentation of the results gets muddled beginning on Line 210. Up to this point we've been discussing HWI greater than 0, as defined by Cai, but now we've switched to greater than 1.0. We are also now talking about reanalysis data, but we only know that from the figure caption.

Response: Thanks for pointing this out. It should be HWI-daily greater than 0.0. We've corrected it in the revised manuscript. To make it clearer, we revised the title of section 3 into "Favorable climatic conditions for Beijing haze events in reanalysis", and moved the model evaluation to section 4.1.

- JRA55 renanlysis is used to assess the ability of the models to simulate key synoptic features relevant to air quality/haze events. However, I am curious if JRA55 is able to simulate historical poor quality conditions over China. Haze data in China goes back a few decades and there are some notable examples or extremely poor air quality. Does the JRA55 capture these events? Can they be identified on Figure 3?

Response: Thanks for this good question. To answer your question, we tested the consistency between JRA-55 and NCEP-NCAR, and examined the relationship between HWI-month derived from JRA-55 and observed haze occurrence and haze visibility. The observed haze is calculated from the station observations from the National Climatic Data Center (NCDC) Global Surface of the Day (GSOD) database for 1974-2013. The main results are as following:

(1) Consistency between JRA-55 and NCEP-NCAR reanalysis.

HWI-daily is defined based on the synoptic circulations associated with observed PM2.5 in Beijing by using NCEP-NCAR reanalysis (Cai et al. 2017), which has proved that HWI-daily can well represent the poor air-quality conditions in Beijing. Here we used JRA-55 because of its higher resolution and better quality over East Asia. We compared the HWI derived from JRA-55 and NCEP-NCAR reanalysis from daily time scale and long-term changes (Fig.A3). We found that the two reanalysis datasets are almost identical with each other in depicting the daily variation of HWI.

A short discussion is added in the revised manuscript (P7 L140-142) as: "The variation of haze index derived from JRA-55 are highly consistent with those from NCEP-NCAR reanalysis (not shown). We only use JRA-55 in this study."

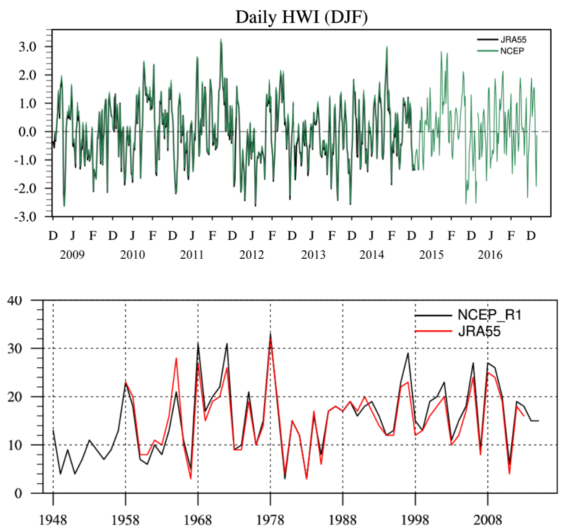

Fig.A3 (a) daily changes of HWI from 1st December 2009 to 31st December 216 in JRA-55 (black) and NCEP-NCAR (green). This period is same as Figure 2 in Cai et al. (2017). (b) Number of days (unit: days) with daily HWI>0 in the winter (DJF) from 1948 to 2013 in JRA-55 (red) and NCEP-NCAR (black).

(2) **Relationship between HWI-month derived from JRA-55 and observed poor air-quality event in China**

This question is also one of the main concerns from Referee #1. We used observed daily visibility, relative humidly and wind speed from 1974 to 2013 from the National Climatic Data Center (NCDC) Global Surface of the Day (GSOD) database (Fig.S1a in the revised manuscript). The positive values of HWI-month are associated with more haze days and less visibility over Beijing and the surrounding region (box in Fig.A4c-d). It proved the good relationship between HWI-month and observed haze occurrence and haze intensity (Fig.A4 here, and Fig.2 in the revised manuscript). Please refer our responses to the first question of Referee #1 for further details.

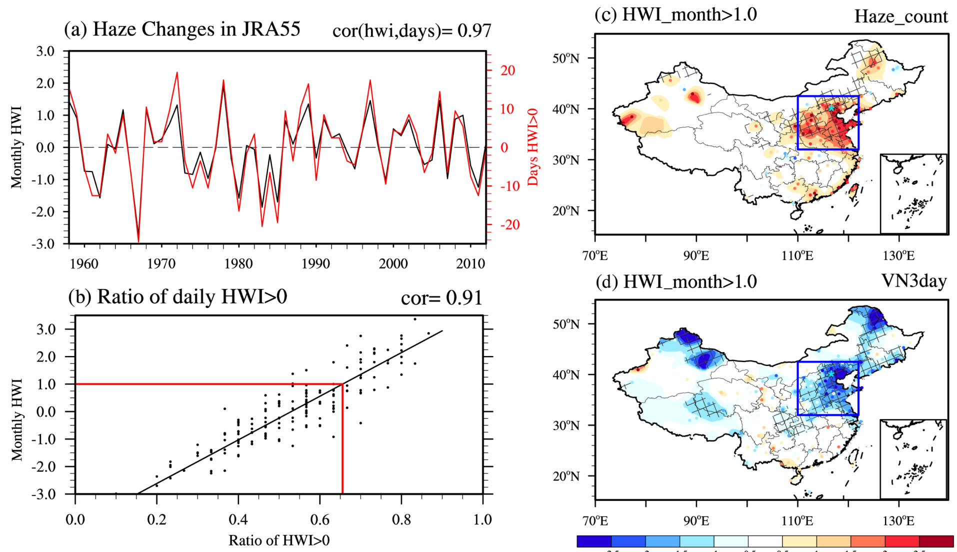

Fig. A4 Changes in winter HWI from 1958 to 2013 in JRA-55 reanalysis relative to 1958-2013 winter mean. (a) DJF mean monthly-based HWI (HWI-month, black line) and the anomalous days with daily based HWI >0 (HWI-daily, red line, unit: day), (b) scatter plot of HWI-month of December, January and February (y-axis) and the ratio of days with HWI-daily>0 (x-axis) in each winter month. HWI-month and HWI-daily are the HWI calculated from monthly data and daily data, respectively. (c)-(d) are the anomalies of haze occurrence and the VN3day when HWI≥1, where VN3day is the minimum 3-day consecutive visibility. Cross area in (c)-(d) is statistically significant at the 10% level using a Student's test. (Fig.2 in the revised manuscript)

- Figure 3 a&b: The blue CLE data obscures the MTFR data. If showing the data to your readers is important, please do so.

Response: We deleted Fig.3a-b since there is no significant trend here. Instead, we added CDF distributions of HWI in the revised manuscript (new Fig.3 and P14 L289-296).

- Figure 3 c & d: This may be personal preference, but I've always had trouble reading discontinuous box plot pdfs. Could simple linear pdfs be used here? I assume the relatively differences in chape is more important to convey than the numbers at each gradation of HWI? I would also appreciate the statistical analysis indicated on each distribution comparison, i.e., at what statistical threshold are these distributions significant?

Response: Thanks for this suggestion. I modified this plot by using linear pdfs based on non-parameter kernel method. The significant threshold is also added. (please see new plots in the revised manuscript)

- I would reiterate the above comment for all of the plots of this style in the manuscript.

Response: All plots are modified as suggested.