Response: North China Plain as a hot spot of ozone pollution exacerbated by extreme high temperatures

Pinya Wang, Yang Yang*, Huimin Li, Lei Chen, Ruijun Dang, Daokai Xue, Baojie Li,

Jianping Tang, L. Ruby Leung, Hong Liao

Dear Editor,

We would like to submit our revised manuscript entitled "**North China Plain as a hot spot of ozone pollution exacerbated by extreme high temperatures**" to ***Atmospheric Chemistry and Physics.***

On behalf of my co-authors, we thank you for handling the peer review of our manuscript. We appreciate your time and efforts as well as those of the two referees for the careful reviews and constructive comments that have helped improve the quality and readability of the manuscript. We have carefully revised our manuscript to address the comments accordingly. Below are the point-to-point responses to the review comments.

Kind regards,

Key:

Black: Reviewer's comments

Blue: Author's responses

Reviewer #1:

This study investigates the co-occurrences of extremes in surface O_3 and extreme heat based on observation datasets, GEOS-Chem model simulations and latest CMIP6 outputs. Detailed analysis on historical and future projections of the coupled extremes as well as the health impact is discussed. The results represent the advances in understanding the interactions between extreme weather events and air pollution. In general, I find the manuscript well written and I recommend it for publication after addressing the following comments:

Reply: We thank the reviewer for the constructive comments and suggestions, which are very helpful for improving the clarity and reliability of the manuscript. Please see our point-by-point responses to your comments below.

Major Comments:

1. The section of model evaluation: I feel the discussions can be more elaborated (Supporting information), and a bit more detailed information such as mean bias, or fractional bias, etc., is useful to indicate more confidence in interpreting the simulated results.

Reply: Thanks for your constructive and helpful comments and suggestions. To improve the model evaluation part, we've added three more statistical metrics, including mean bias (MB), mean fractional bias (MFB) and root mean square error (RMSE) to quantitively evaluate the performance of GEOS-Chem model and CMIP6 simulations, based on the equations listed in the appendix of Zhang et al. (2018). The metrics have been shown in the updated Fig.S4 and Fig.S5 (shown as below).

Accordingly, we have revised Text S1 and Text S2 by adding more interpretations：
Text S1 (Line 29-36): "The spatial correlations between the simulated and observed OPCs and CF values are all higher than 0.5 and are statistically significant at 95% confidence level, accompanied by small mean bias (MB) and root mean square error (RMSE) values. For example, the MB between the simulated and observed OPCs and CF values over China are as low as 2.34 days and -0.23%, respectively. Moreover, the mean fractional bias (MFB) for CF values is well within the limit of MFB for O_3 evaluation (15%) recommended by EPA (2007). The statistical metrics suggest that the model can reasonably reproduce the observed spatial patterns and magnitudes of OPCs and CF over NCP during 2014-2017."

Text S2 (Line 50-54): "Similarly, the MFB and RMSE for both simulated OPCs and CF values under SSP3-7.0 are the lowest among the four scenarios. The relatively higher MB and RMSE under SSP2-4.5 come from the overestimation of OPCs and CF values over the whole China, likely related to the inaccurate of SSPs emissions in China during this time period (Chen et al., 2021; Wang et al., 2021)."

Cheng, J., Tong, D., Liu, Y., Yu, S., Yan, L., Zheng, B., et al. (2021). Comparison of current and future PM2.5 air quality in China under CMIP6 and DPEC emission scenarios. Geophysical Research Letters, 48, e2021GL093197. https://doi.org/10.1029/2021GL093197

Wang, Z., Lin, L., Xu, Y., Che, H., Zhang, X., Dong, W., Wang, C., Gui, K., and Xie, B.: Incorrect Asian aerosols affecting the attribution and projection of regional climate change in CMIP6 models, npj Clim. Atmos. Sci., 4, 2, https://doi.org/10.1038/s41612-020-00159-2, 2021.

Zhang, J., Y. Gao, K. Luo, L. R. Leung, Y. Zhang, K. Wang, and J. Fan (2018), Impacts of compound extreme weather events on ozone in the present and future, Atmospheric Chemistry and Physics, 18(13), 9861-9877.

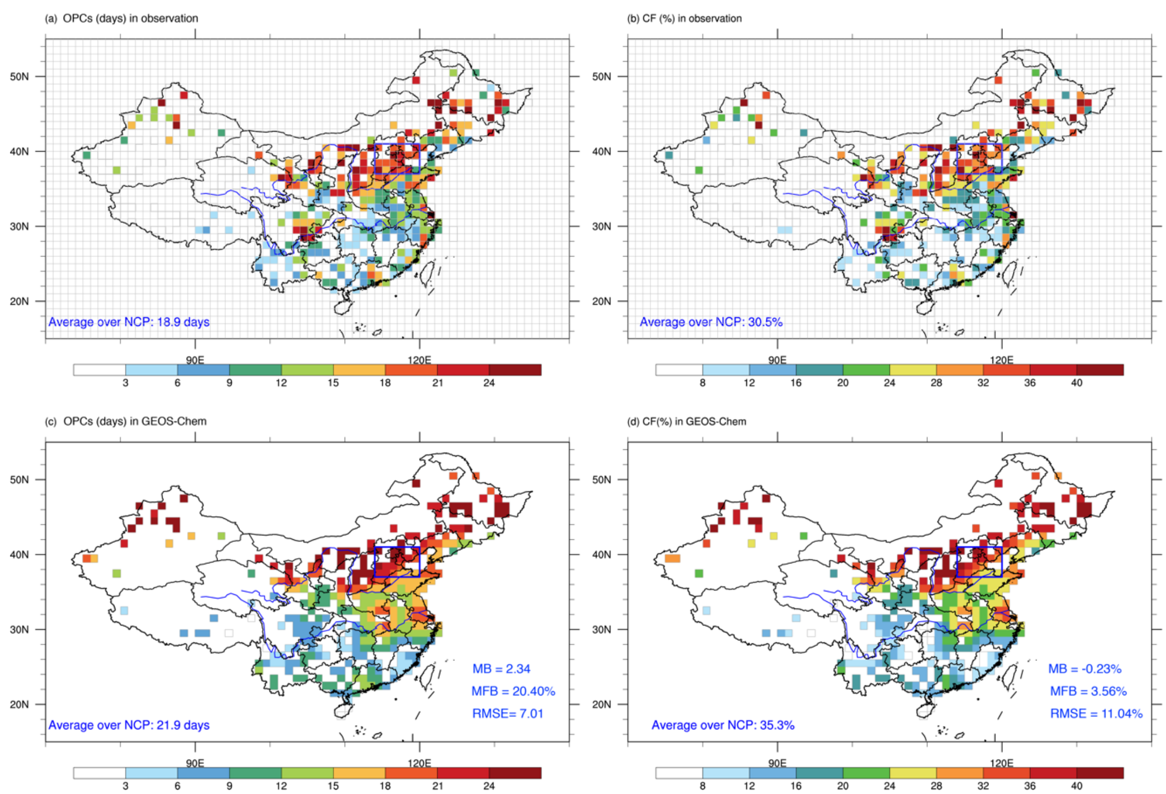


Figure S4. Spatial patterns of observed (a) OPCs (days) and (b) CF values (%) during May-September of 2014-2017. (c) and (d) are same as (a) and (b) but for the GEOS-Chem simulation. Observed and simulated values of OPCs(days) and CF averaged over NCP (37-41°N; 114-120°E) are indicated at the bottom left corner of each panel. Statistical metrics including MB, MFB, and RMSE are noted at the bottom right of panels (c) & (d). Note that the three metrics are obtained over the whole China, with equations listed in the appendix of Zhang et al. (2018).

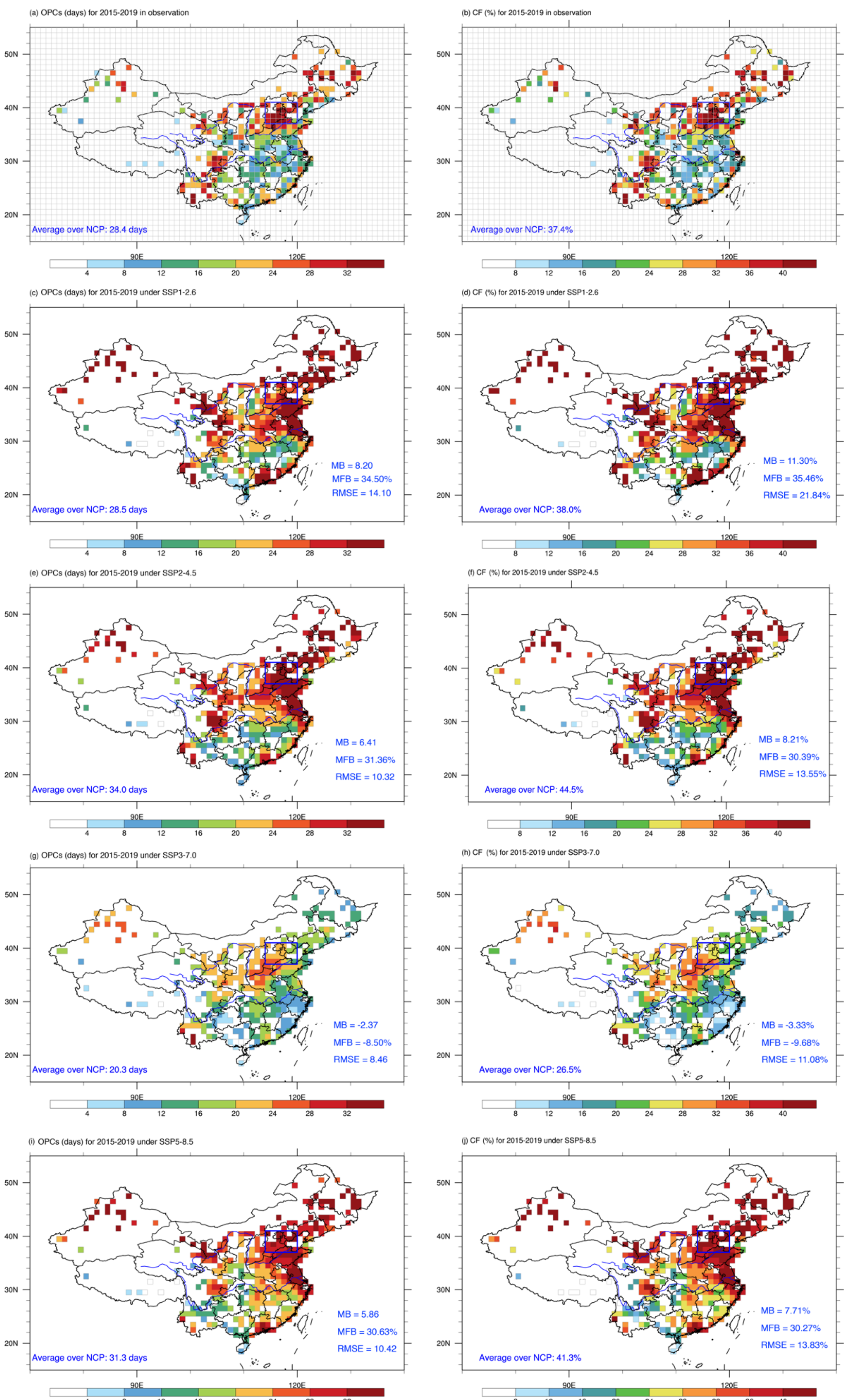

(a) OPCs (days) for 2015-2019 in observation
Average over NCP: 28.4 days
(b) CF (%) for 2015-2019 in observation
Average over NCP: 37.4%
(c) OPCs (days) for 2015-2019 under SSP1-2.6
MB = 8.20
MFB = 34.50%
RMSE = 14.10
Average over NCP: 28.5 days
(d) CF (%) for 2015-2019 under SSP1-2.6
MB = 11.30%
MFB = 35.46%
RMSE = 21.84%
Average over NCP: 38.0%
(e) OPCs (days) for 2015-2019 under SSP2-4.5
MB = 6.41
MFB = 31.36%
RMSE = 10.32
Average over NCP: 34.0 days
(f) CF (%) for 2015-2019 under SSP2-4.5
MB = 8.21%
MFB = 30.39%
RMSE = 13.55%
Average over NCP: 44.5%
(g) OPCs (days) for 2015-2019 under SSP3-7.0
MB = -2.37
MFB = -8.50%
RMSE = 8.46
Average over NCP: 20.3 days
(h) CF (%) for 2015-2019 under SSP3-7.0
MB = -3.33%
MFB = -9.68%
RMSE = 11.08%
Average over NCP: 26.5%
(i) OPCs (days) for 2015-2019 under SSP5-8.5
MB = 5.86
MFB = 30.63%
RMSE = 10.42
Average over NCP: 31.3 days
(j) CF (%) for 2015-2019 under SSP5-8.5
MB = 7.71%
MFB = 30.27%
RMSE = 13.83%
Average over NCP: 41.3%
50N
40N
30N
20N
90E
120E
4 8 12 16 20 24 28 32
8 12 16 20 24 28 32 36 40

Figure S5. Spatial patterns of (a) OPCs (days) and (b) CF values (%) during May-September of 2015-2019 in observation; (c)~(d), (e)~(f), (g)~(h), and (g)~(h) are same as (a) and (b) but for CMIP6 simulations under SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5, respectively. OPCs (days) and CF averaged over NCP (37-41°N; 114-120°E) are indicated at the bottom left corner of each panel. Statistical metrics including MB, MFB, and RMSE are noted at the bottom right of panels (c)~(j). Note that the three metrics are obtained over the whole China.

2. In terms of the emissions: the authors only discussed anthropogenic emission inventory. How about biogenic emissions? Considering that biogenic emissions are quite important for ozone formation, particularly of the synergic effect of biogenic and anthropogenic emissions on ozone formation, it is useful to indicate how the biogenic emissions were treated in this study.

Reply: Thanks for your constructive and helpful comments and suggestions. We have added explanation on how the biogenic emissions in the updated manuscript:

Line 138-142: "Biogenic volatile organic compound (BVOC) emissions also play vital roles in modulating the formation of ozone and secondary organic aerosols (Ma et al., 2021; Y. Gao et al., 2021). For biogenic emissions in GEOS-Chem, the Model of Emissions of Gases and Aerosols from Nature (MEGAN) v2.1 biogenic emissions are applied with updates from Guenther et al. (2012)."

Line 305-309: "In addition, Fu et al. (2015) have indicated that the enhanced biogenic emissions and the accelerated photochemical reaction rates both increased surface ozone over the US during 1988–2011. Thus, the increasing trend of biogenic emissions due to vegetation biomass variability over China (J. Gao et al., 2021) may also have potential impacts on the variations of OPCs."

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5, 1471–1492, https://doi.org/10.5194/gmd-5-1471-2012, 2012.

Ma, M., Gao, Y., Ding, A., Su, H., Liao, H., Wang, S., ... & Gao, H. (2021). Development and Assessment of a High-Resolution Biogenic Emission Inventory from Urban Green Spaces in China. Environmental science & technology.

Gao, Y., F. Yan, M. Ma, A. Ding, H. Liao, S. Wang, X. Wang, B. Zhao, W. Cai, H. Su, X. Yao and H. Gao (2021), Unveiling the dipole synergic effect of biogenic and anthropogenic emissions on ozone concentrations, Sci. Total Environ., 151722.

Cao, J., Situ, S., Hao, Y., Xie, S., & Li, L. (2021). Enhanced summertime ozone and SOA from biogenic volatile organic compound (BVOC) emissions due to vegetation biomass variability during 1981–2018 in China. Atmospheric Chemistry and Physics Discussions, 1-21.

3. About the impact of extreme events on ozone: the compound extreme events have recently been raised as a substantial concern to ozone formation. At least adding a few sentences or references to discuss the compound extremes (i.e., multiple extremes occur simultaneously) and the associated impact on ozone formation is useful.

Reply: Thanks for your constructive and helpful comments and suggestions. We have added more discussions in the Discussion and Conclusion part: "Recently, the compound extreme events (e.g., co-occurrence of two extreme weather events simultaneously) are raised as a substantial concern to O_3 formation. For example, the co-occurrences of heat wave and air stagnation promote higher O_3 concentration compared to the single extreme events of heat wave or stagnation in the U.S. in the future relative to the present (Zhang et al., 2018; Y Gao et al., 2020)."

Zhang, J., Gao, Y., Luo, K., Leung, L. R., Zhang, Y., Wang, K., & Fan, J. (2018). Impacts of compound extreme weather events on ozone in the present and future. Atmospheric Chemistry and Physics, 18(13), 9861-9877.

Gao, Y., J. Zhang, F. Yan, L. R. Leung, K. Luo, Y. Zhang and M. L. Bell, Nonlinear effect of compound extreme weather events on ozone formation over the United States (2020), Weather and Climate Extremes, 30, 100285.

Minor Comments:

1. Lines 80, 187, change "O3" to "O_3" and check throughout the entire text.
   Changed.
2. Line 208, change "MDA O3" to "MDA8 O_3".
   Changed.
3. Missing subtitle (b) in figure 2.
   Added.
4. Line 264, please be careful that the enhanced chemical production and weakened mixing and dry deposition contribute to the increase O_3 level during OPCs.
   Thanks. Modified.
5. Please use a larger font size in Figure 4 as the subtitle in each panel is hard to read. The same applies for Figure 5.
   Thanks. Both Figure 4 and Figure 5 are updated with a larger font size.
6. In terms of the health impacts of OPCs, have you considered the possible impacts of temperatures on surface ozone related health risk, i.e., higher temperatures may worsen the health impacts of surface ozone.
   As claimed in the manuscript (Line 359-362), previous studies have pointed out that O_3-related mortality may change with different air temperature levels, and yet the conclusions can be contrasting or inconsistent for different regions. Thus, this work does not consider the possible amplification/inhibition effect of combining O_3 and air temperature in affecting human health.
7. Line 212. Repeated definitions of abbreviation. An abbreviation is only needed with it

appears for the first time. Please double check the entire texts.

Thanks. Deleted.

8. As the author stated that GEOS-Chem simulations cover only the period of 2014-2017, does this mean that the definitions of OPCs and OPIs are applied to 2014-2017 for both observation and simulations? How about future?

Thanks for your question. Yes, as addressed in the Text S1&S2, the GEOS-Chem simulations are conducted for 2014-2017. And the model simulations are evaluated based on observations during 2014-2017. Thus, both observed and simulated OPCs and OPIs are applied to 2014-2017. For the future projections, future OPCs during the mid-century (2046-2050) and end-century (2096-2100) are compared with OPCs during 2015-2019 for a consistency in time length.

9. The caption of Figure S3: downward solar radiation flux Does this mean downward surface solar radiation?

Yes. We have made it clear in Sec. 2.1 of the updated manuscript (Line 120): "downward solar radiation flux (DSR) and sensible heat flux (SH) at surface."

10. Figure S8 includes some important information, and it is good to move it to the main manuscript.

Thanks. Figure S8 has been put in the main manuscript and renamed as Figure 7 in the updated version.

Reviewer #2:

In this manuscript, the authors present an interesting study that can be a valuable contribution to the existing understanding of ozone pollution in China and its connections to health and climate. The study is carefully thought-out, conducted through a series of original analyses, and arrives at significant findings. In addition, the manuscript clearly describes the work conducted. Several major comments and some additional minor points that should be addressed by the authors prior to publication are included below.

Reply: Thanks for the constructive comments and suggestions, which are very helpful for improving the clarity and reliability of the manuscript. Please see our point-by-point responses to your comments below.

1. The authors rely on the 90th percentile at each grid cell to define O_3 and temperature extremes. Given that the exceedance of air quality standards and health impacts are dependent on O_3 concentrations rather than a percentile score, would a consistent threshold across all grid cells not be a more relevant metric for extreme air pollution? By relying on the 90th percentile for each individual cell, high O_3 pollution cells are not consistently defined across the domain and thus a location with a large number of O_3 pollution days as defined in the study may be experiencing lower total O_3 pollution than one with a smaller number of O_3 pollution days. A clearer description of this "local-specific" threshold, including justification and implications of its selection, is needed.

Reply: Thanks for the kind suggestion. It should be noted that, the 90th percentile of O_3 over typical mega-city cluster NCP, Yangtze River Delta, Sichuan Basin and Pearl River Delta are 97.7 ppb, 84.4 ppb, 73.7 ppb, 76.8 ppb, respectively, close to/above China's Grade II air quality standard (160 ug/m^3) for MDA8 O_3 (around 80 ppb under standard atmospheric conditions). Thus, the local-specific threshold of 90th percentile is a reasonable metric to identify O_3 extremes in China. And the local-specific thresholds have been widely used in recent works of ozone pollution (Schnell& Prather, 2017; Lin et al., 2019; Qin et al., 2021). In fact, with the local 90th percentiles as the extreme thresholds, the O_3 and temperature extremes over all grid cells are equivalent to 100 days (6 years × 153 days/year × 0.1) during the warm season (May 1-Septemper 30) of 2014-2019. Besides, a location experiencing the higher O_3/temperature levels has a higher value of the extreme threshold. Though the O_3 and temperature exceedance days are spatially equal, their co-occurrences exhibit spatial variance, with the highest co-occurrence over North China Plain (NCP), associated with the distinctive relationship between O_3 and temperature.

Accordingly, we have added more explanations on the "local-specific threshold" in the revised manuscript (Line 174-178): "The local-specific thresholds have been widely used in recent studies of ozone pollution (e.g., Schnell and Prather, 2017; Lin et al., 2019; Qin et al., 2021). Note that the 90th percentile of MDA8 O_3 over NCP, Yangtze River Delta Sichuan Basin and Pearl River Delta are 97.7 ppb, 84.4 ppb, 73.7 ppb, 76.8 ppb, respectively, close to China's Grade II air quality standard for MDA8 O_3 (around 80 ppb under standard atmospheric conditions)."

Qin, Y., Li, J., Gong, K., Wu, Z., Chen, M., Qin, M., ... & Hu, J. (2021). Double high pollution events in the Yangtze River Delta from 2015 to 2019: Characteristics, trends, and meteorological situations. Science of The Total Environment, 148349.

Lin, X., Yuan, Z., Yang, L., Luo, H., & Li, W. (2019). Impact of extreme meteorological events on ozone in the Pearl River Delta, China. Aerosol and Air Quality Research, 19(6), 1307-1324.

Schnell, J. L., and M. J. Prather (2017), Co-occurrence of extremes in surface ozone, particulate matter, and temperature over eastern North America, Proc Natl Acad Sci U S A, 114(11), 2854-2859.

2. For the study's estimates of health impacts, a deeper discussion of the epidemiological studies and β coefficients selected is needed, including:

• How do the β coefficients compare to others reported by different epidemiological studies, including those most commonly used internationally? How extensively have the coefficients used here been applied, and why would they need to be specifically derived from data in China?

Reply: Thanks. Currently, the commonly used β coefficients for mortality risks associated with the short-term exposures to ozone pollution and air temperatures are evaluated based on datasets from developed countries (e.g., Bell et al. 2004, 2005; Gryparis et al. 2004; Ng et al. 2013). Because China has higher air pollution levels and may also differ in terms of age structure, population sensitivity to air pollution/heat exposures, and components of air pollution mixture compared to developed countries (K Chen et al, 2018), we use China-specific concentration and temperature response functions in the present study, as indicated in the recent nationwide studies (Yin et al., 2017; Huang et al., 2015). Accordingly, we have added the explanation on the adopted β coefficients in Set 2.5.

Chen, K., Fiore, A. M., Chen, R., Jiang, L., Jones, B., Schneider, A., ... & Kinney, P. L. (2018). Future ozone-related acute excess mortality under climate and population change scenarios in China: A modeling study. PLoS medicine, 15(7), e1002598.

• The assumption of no lower threshold for O_3 mortality (C_0=0) is not applied in other analyses. Rather, an assumption that a threshold for ozone effects is likely near the lower limit of ambient ozone concentrations in countries like the US is often considered. Given that all results presented here are based on the ratio of relative risks, it would appear that defining a C_0 (and T_0) threshold is not necessary and can be avoided.

Reply: Thanks for the suggestion. The definitions of C_0 and T_0 has been omitted.

• While in Discussion and conclusions, the study acknowledges the uncertainty associated with combining relative risks of O_3 and temperature without considering coupled effects, this limitation should be mentioned earlier when describing the methods. To what extent do each of the studies of O_3 and T mortality from which the β coefficients are taken control for the other variable?

Reply: Thanks for the kind suggestion. We have moved the discussions on the limitation to the method part. Also, we have added a sentence in the updated manuscript as "Previous studies have claimed that O_3-related mortality increases with higher temperatures, although several studies presented contrasting results or inconsistent relationships for different regions (R Chen et al., 2014; Jhun et al., 2014; Ren et al., 2008). By analyzing the total mortality rates associated with short-term O_3 exposure over East Asia among four seasons, R Chen et al (2014) found that the higher temperatures in summer significantly increased the O_3-related mortality rates."

3. While temperature is a key driver of O_3 formation, emissions of O_3 precursors also play a major role in O3 pollution. The discussion of emissions in the paper is minimal. How do emissions, including anthropogenic and biogenic precursors, vary temporally and spatially? Beyond meteorological factors, could variability in emissions be partially be driving for the frequency and geographic differences in the co-occurrence of high O3 and temperature?

Reply: Thank you for the suggestion. The anthropogenic emissions are obtained from the MEIC emission inventory (http://meicmodel.org/), and have been widely used and illustrated in previous studies (e.g., Li et al., 2019). For biogenic emissions in GEOS-Chem, the Model of Emissions of Gases and Aerosols from Nature (MEGAN) v2.1 biogenic emissions are applied with updates from Guenther et al. (2012). We have also added more discussions on the spatiotemporal variations of air pollutant emission over China, including anthropogenic and biogenic emissions (Line 297-309):

"Besides meteorological effects, the O_3 precursor emissions should partially contribute to the spatiotemporal variations of OPCs over China. It's reported that surface O_3 pollution levels are strongly correlated with daytime surface temperatures, especially in highly polluted regions, with strong precursor emissions (Poter and Heald, 2019). NCP has the highest anthropogenic emissions compared to the other regions in China, which should benefit the higher correlations between surface O_3 and air temperatures, and thus the higher OPCs therein. Moreover, the increasing trend of OPCs over NCP in recent years may be associated with the continued anthropogenic increases in O_3, as well as the unmitigated emissions of VOCs (Li et al., 2019), emphasizing the need for controlling anthropogenic emissions of VOCs. In addition, Fu et al. (2015) have indicated that the enhanced biogenic emissions and the accelerated photochemical reaction rates both act to increase surface ozone over the US during 1988–2011. Thus, the increasing trend of biogenic emissions due to vegetation biomass variability over China (Gao et al., 2021) may also have potential impacts on the variations of OPCs".

Porter, W. C., & Heald, C. L. (2019). The mechanisms and meteorological drivers of the summertime ozone–temperature relationship. Atmospheric Chemistry and Physics, 19(21), 13367-13381.

Fu, T. M., Zheng, Y., Paulot, F., Mao, J., & Yantosca, R. M. (2015). Positive but variable sensitivity of August surface ozone to large-scale warming in the southeast United States. Nature Climate Change, 5(5), 454-458.

Cao, J., Situ, S., Hao, Y., Xie, S., & Li, L. (2021). Enhanced summertime ozone and SOA from biogenic volatile organic compound (BVOC) emissions due to vegetation biomass variability during 1981–2018 in China. Atmospheric Chemistry and Physics Discussions, 1-21.

4. The authors acknowledge the role of interannual variability in the analysis of the 2014-2019 period, noting that trends reflect interannual variability rather than a long-term warming trend. Recent work has shown the significant influence that internal variability can have on long-term projections of both temperature and O_3 concentration. Here, 5-year periods are used to characterize future temperature and air quality at midcentury and the end of the century. The length of these periods is insufficient to confidently distinguish a forced signal in temperature and O_3 from the noise imposed by natural variability. How do the climate simulations used account for internal variability and to what extent may internal variability be affecting the climate-related findings of this study? At a minimum, the authors must acknowledge the large uncertainty imposed by natural variability on the projected coupled extremes.

Reply: Thanks for the kind suggestion. In the CMIP6 analysis, we use five model results to minimize the influence of nature variability, but it may not be sufficient. We have added discussions to highlight the potential impacts of internal variability on the projected extremes in the updated manuscript (Line 362-367): "Note that for the future changes of OPCs, the influences of natural variability are less considered, whereas previous studies have emphasized the significant role of natural variability on altering the robustness of climate projections and their impacts on air quality (e.g, Garcia-Menendez et al., 2017). The detection of the anthropogenic-forced signal demands a lager model ensemble and a longer simulation length that deserves further explorations."

Garcia-Menendez, F., Monier, E., & Selin, N. E. (2017). The role of natural variability in projections of climate change impacts on US ozone pollution. Geophysical Research Letters, 44(6), 2911-2921.

5. Evaluation of model results against observations, for GEOS-Chem and CIMP6

simulations, should be expanded beyond a visual comparison of the spatial pattern and country total number of OPCs. Established model performance statistics (e.g., normalized mean bias, normalized mean error (NME), and correlation coefficient) can more definitely determine if the models indeed "reasonably capture" observed values and meet accepted performance standards. For the GCM simulations specifically, the models are known to often have high biases in modeled O_3. Would bias-correcting the projected concentrations alter the findings?

Reply: Thanks for your suggestion. We have we've added more statistical metrics, including correlation coefficient, mean bias (MB), mean fractional bias (MFB) and root mean square error (RMSE) to quantitively evaluate the performance of GEOS-Chem model and CMIP6 simulations, based on the equations listed in the appendix of Zhang et al. (2018). The metrics have been shown in the updated Fig.S4 and Fig.S5. And accordingly, we have revised Text S1 and Text S2 by adding more interpretations:

Text S1 (Line 29-36): "The spatial correlations between the simulated and observed OPCs and CF values are all higher than 0.5 and are statistically significant at 95% confidence level, accompanied by small mean bias (MB) and root mean square error (RMSE) values. For example, the MB between the simulated and observed OPCs and CF values over China are as low as 2.34 days and -0.23%, respectively. Moreover, the mean fractional bias (MFB) for CF values is well within the limit of MFB for O_3 evaluation (15%) recommended by EPA (2007). The statistical metrics suggest that the model can reasonably reproduce the observed spatial patterns and magnitudes of OPCs and CF over NCP during 2014-2017."

Text S2 (Line 49-52): "Similarly, the MFB and RMSE for both simulated OPCs and CF values under SSP3-7.0 are the lowest among the four scenarios. The relatively higher MB and RMSE under SSP2-4.5 come from the overestimation of OPCs and CF values over the whole China, likely related to the inaccurate of SSPs emissions in China during this time period (Chen et al., 2021; Wang et al., 2021)."

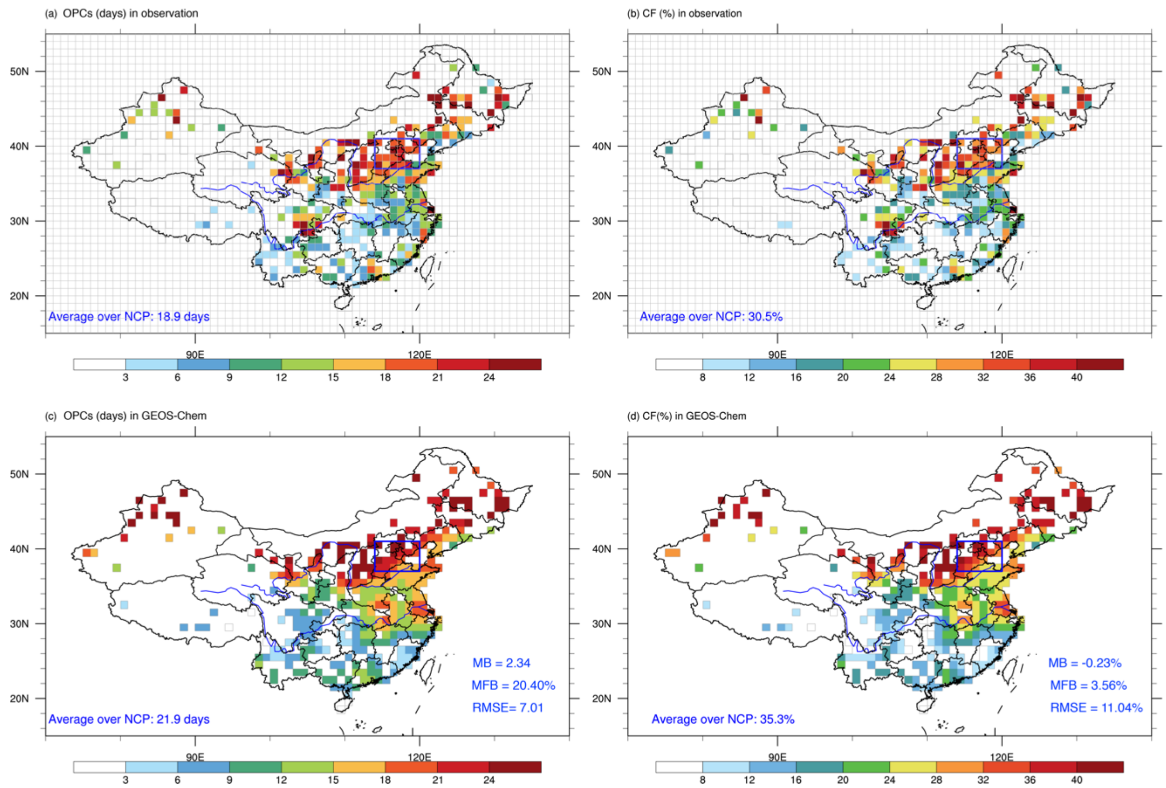


Figure S4. Spatial patterns of observed (a) OPCs (days) and (b) CF values (%) during May-September of 2014-2017. (c) and (d) are same as (a) and (b) but for the GEOS-Chem simulation. Observed and simulated values of OPCs(days) and CF averaged over NCP (37-41°N; 114-120°E) are indicated at the bottom left corner of each panel. Statistical metrics including MB, MFB, and RMSE are noted at the bottom right of panels (c) & (d). Note that the three metrics are obtained over the whole China, with equations listed in the appendix of Zhang et al. (2018).

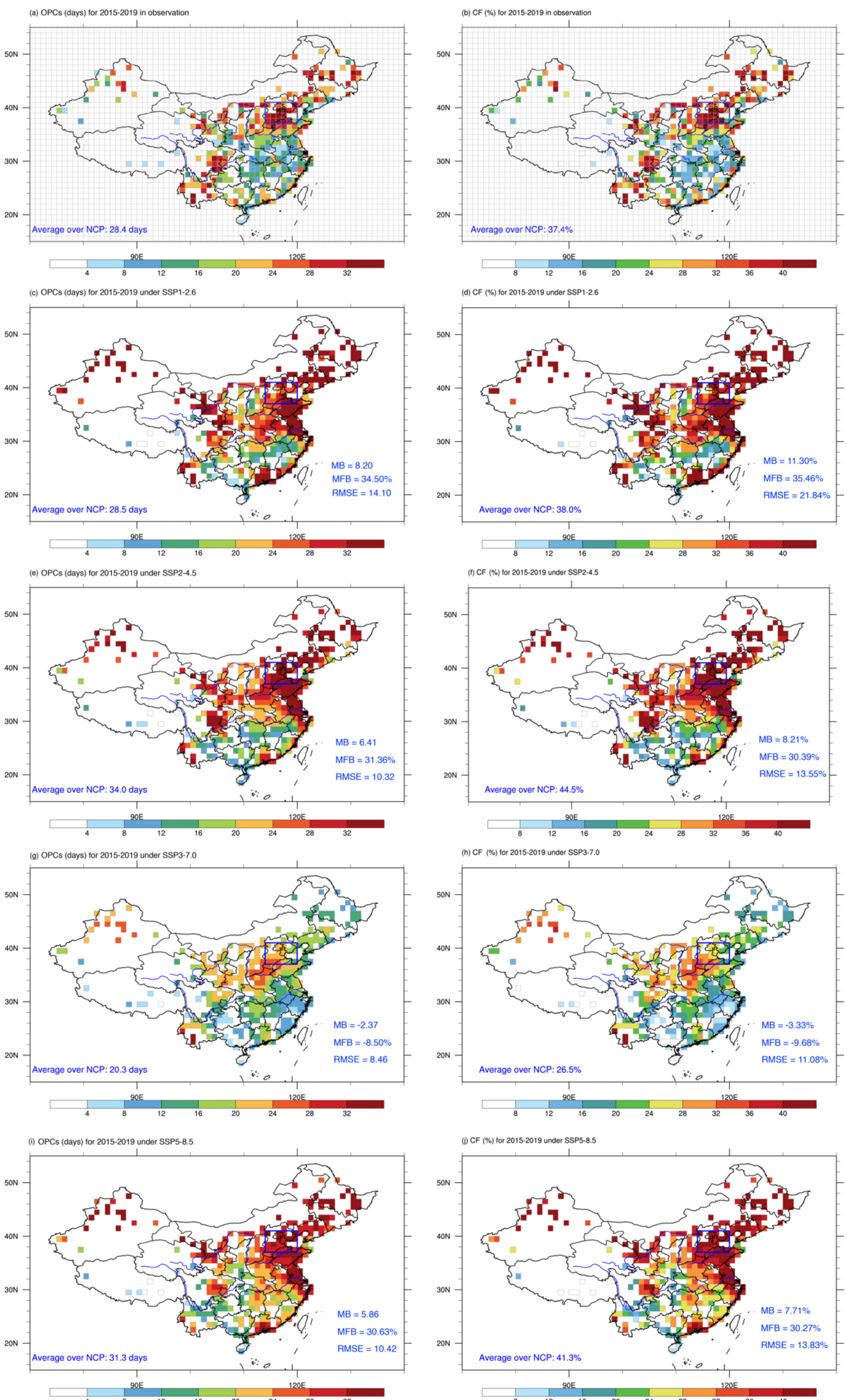

(a) OPCs (days) for 2015-2019 in observation
Average over NCP: 28.4 days
(b) CF (%) for 2015-2019 in observation
Average over NCP: 37.4%
(c) OPCs (days) for 2015-2019 under SSP1-2.6
MB = 8.20
MFB = 34.50%
RMSE = 14.10
Average over NCP: 28.5 days
(d) CF (%) for 2015-2019 under SSP1-2.6
MB = 11.30%
MFB = 35.46%
RMSE = 21.84%
Average over NCP: 38.0%
(e) OPCs (days) for 2015-2019 under SSP2-4.5
MB = 6.41
MFB = 31.36%
RMSE = 10.32
Average over NCP: 34.0 days
(f) CF (%) for 2015-2019 under SSP2-4.5
MB = 8.21%
MFB = 30.39%
RMSE = 13.55%
Average over NCP: 44.5%
(g) OPCs (days) for 2015-2019 under SSP3-7.0
MB = -2.37
MFB = -8.50%
RMSE = 8.46
Average over NCP: 20.3 days
(h) CF (%) for 2015-2019 under SSP3-7.0
MB = -3.33%
MFB = -9.68%
RMSE = 11.08%
Average over NCP: 26.5%
(i) OPCs (days) for 2015-2019 under SSP5-8.5
MB = 5.86
MFB = 30.63%
RMSE = 10.42
Average over NCP: 31.3 days
(j) CF (%) for 2015-2019 under SSP5-8.5
MB = 7.71%
MFB = 30.27%
RMSE = 13.83%
Average over NCP: 41.3%
50N
40N
30N
20N
90E
120E
4 8 12 16 20 24 28 32
8 12 16 20 24 28 32 36 40

Figure S5. Spatial patterns of (a) OPCs (days) and (b) CF values (%) during May-September of 2015-2019 in observation; (c)~(d), (e)~(f), (g)~(h), and (g)~(h) are same as (a) and (b) but for CMIP6 simulations under SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5, respectively. OPCs (days) and CF averaged over NCP (37-41^{o}N; 114-120^{o}E) are indicated at the bottom left corner of each panel. Statistical metrics including MB, MFB, and RMSE are noted at the bottom right of panels (c)~(j). Note that the three metrics are obtained over the whole China.

Cheng, J., Tong, D., Liu, Y., Yu, S., Yan, L., Zheng, B., et al. (2021). Comparison of current and future PM2.5 air quality in China under CMIP6 and DPEC emission scenarios. Geophysical Research Letters, 48, e2021GL093197. https://doi.org/10.1029/2021GL093197

Wang, Z., Lin, L., Xu, Y., Che, H., Zhang, X., Dong, W., Wang, C., Gui, K., and Xie, B.: Incorrect Asian aerosols affecting the attribution and projection of regional climate change in CMIP6 models, npj Clim. Atmos. Sci., 4, 2, https://doi.org/10.1038/s41612-020-00159-2, 2021.

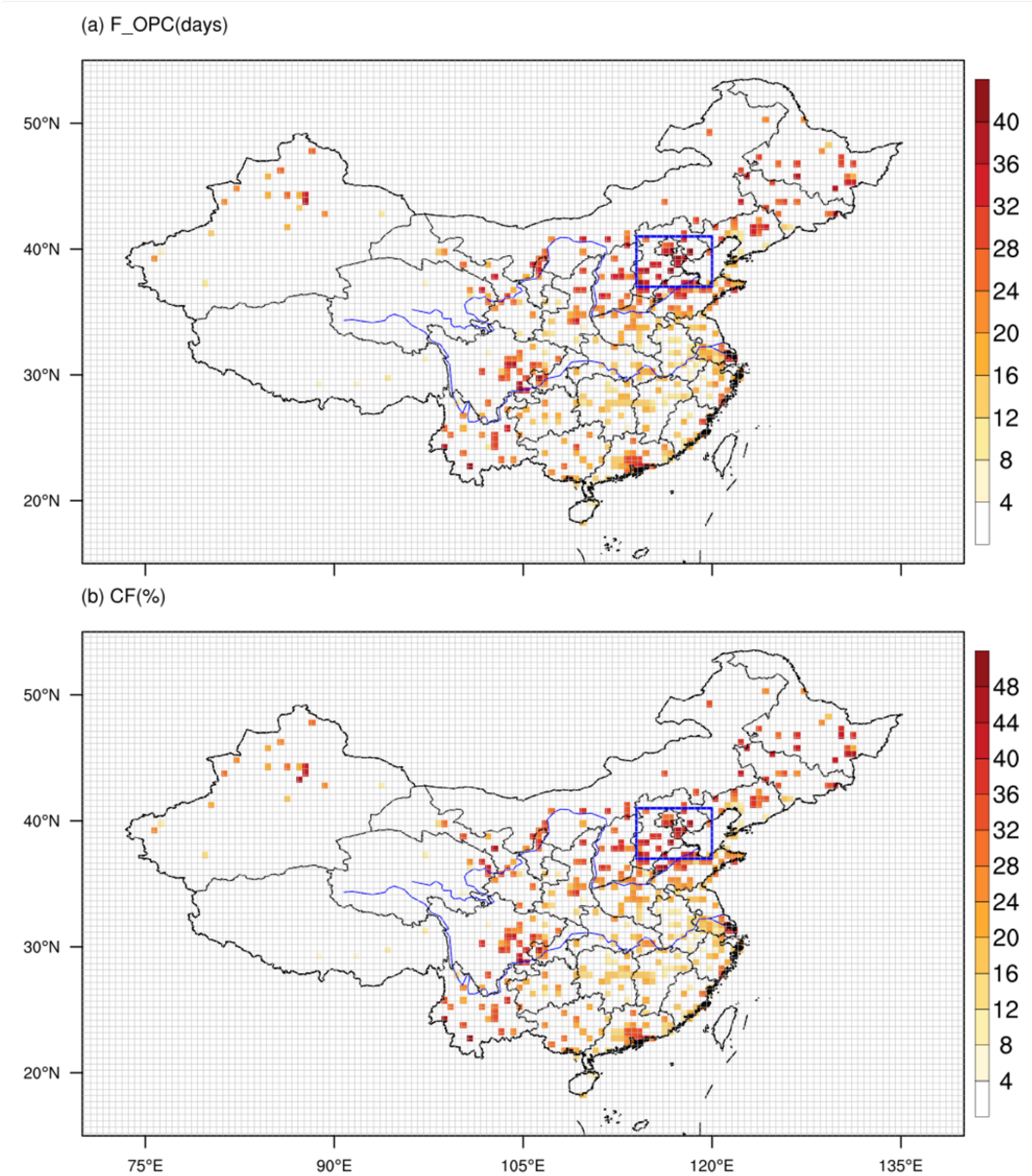
Zhang, J., Y. Gao, K. Luo, L. R. Leung, Y. Zhang, K. Wang, and J. Fan (2018), Impacts of compound extreme weather events on ozone in the present and future, Atmospheric Chemistry and Physics, 18(13), 9861-9877.

In this study, we define O_3 extreme using the 90th percentile. Although modeling could have biases in simulating absolute O_3 concentrations, it has less influence on the extreme days.

6. The spatial resolution of the analysis (1 degree) is coarse and the model simulations are even coarser. However, resolution is not discussed. To what extent may the resolution of the data fields affect the results? Further discussion of the potential limitations imposed by coarse resolution is necessary.

Thanks for the kind suggestion.

Firstly, for the resolution of the observation analysis, 1° grid is adopted in the study, and each grid value represents the averaged observations of Tmax and MDA8 O_3 within the box. The spatial patterns of OPCs and CF values with a finer resolution (0.5°) is shown below, with maximum OPCs and CF values over NCP, consistent with those with 1° grid (Figure 1 in the main text). Moreover, the magnitudes of OPCs and CF values over NCP are close the two different grid sizes. Therefore, a finer or coarser resolution are not likely to affect the results. Thus, we have added clarifications on 1° resolution in Method part of the manuscript (Line110-111): “We have also tested the grid size of 0.5° and found that the different grid resolutions have negligible influence on the results.”



Same as in Figure 1 except with 0.5° × 0.5° grid boxes.

Secondly, for the potential impacts of model resolution on the results, we have added discussions in the updated manuscript (Line 130-135): "The resolution of climate models has important effects on simulating O_3 and air temperatures (Fenech et al., 2018). By examining the simulations of surface O_3 over the U.S. with a regional climate model and the global GEOS-Chem model, Fiore et al. (2003) indicate that the ability to resolve local O_3 maxima is compromised, but the spatial correlation improves when the model resolution coarsens. The coarse-resolution global model can successfully capture the synoptic-scale processes modulating O_3 concentrations whereas a finer spatial resolution may improve the representation of processes occurring on smaller scales."

Fenech, S., Doherty, R. M., Heaviside, C., Vardoulakis, S., Macintyre, H. L., & O'Connor, F. M. (2018). The influence of model spatial resolution on simulated ozone and fine particulate matter for Europe: implications for health impact assessments. Atmospheric Chemistry and Physics, 18(8), 5765-5784.

Fiore, A. M., Jacob, D. J., Mathur, R., & Martin, R. V. (2003). Application of empirical orthogonal functions to evaluate ozone simulations with regional and global models. Journal of Geophysical Research: Atmospheres, 108(D14).

Other comments:

• Line 41: Listing $PM_{2.5}$ as an example (e.g.) of particulate matter is confusing.

Modified.

• Line 60: The 2003 example described occurred nearly 20 years ago. Is it still relevant given the significant changes that have occurred in China since?

Thanks for pointing this out. In fact, we try to emphasize the disastrous impacts of heat waves over China with an example of the typical 2003 heat wave. In the revised manuscript, we have added a more recent heat waves in summer 2018 (Line 63-66): "Mideastern China experienced an excessively long heat wave over a wide-ranging area from mid-July to mid-August 2018.The local maximum temperatures exceeded 40℃, and the spatial extent involved 18 provinces, resulting in record-breaking overloaded power grids in many areas (Li et al., 2019; Lu et al., 2020)".

Li, M., Yao, Y., Luo, D., & Zhong, L. (2019). The linkage of the large-scale circulation pattern to a long-lived heatwave over Mideastern China in 2018. Atmosphere, 10(2), 89.

Lu, C., Ye, J., Wang, S., Yang, M., Li, Q., He, W., ... & Mao, J. (2020). An unusual heat wave in North China during midsummer, 2018. Frontiers in Earth Science, 8, 238.

• It is unclear why it is necessary to "standardize" meteorological variables as described on line 116; explain the intent further.

Reply: Thanks. Following Gong et al. (2019), the standardized meteorological variables enable a direct comparison among their magnitudes during extreme O_3 and/or high temperatures. As addressed in the submission version of manuscript (Line 284-287): "Among the meteorological factors, the intensification in surface temperatures during

OPCs is the strongest among different meteorological variables with the highest magnitudes, supporting that air temperature is the most influential meteorological variable of surface O_3 over NCP (K Li et al., 2019)."

In addition, we added an explanation on the standardized meteorological variables in the Method part (Line 122-123): "The standardized meteorological variables enable a direct comparison among their magnitudes during extreme O_3 and/or high temperatures."

• Define the resolution of GCMs (line 147 and table S1) in terms of degrees rather than the number of model cells.

Reply: Thanks. Modified.

• Figure S2: Change the 'SC' label on panel (b) to YRD

Reply: Thanks. Changed.

• Figure 4: What are the units of the plots?

Reply: It's unitless as the meteorological variables have been standardized.

• Line 173 (and throughout): A better term than "mortality ratio" can be used. One that is more descriptive of what the ratio represents, enhanced O_3 mortality for coupled extreme O_3 and temperature days, should be selected.

Thanks. Changed.

Specially, we've rewritten the description on the index (Line 193-196): "we apply a ratio index to describe the combined human health impacts caused by O_3 and temperatures during OPCs, which represents the potential enhancement in mortality rates (referred as to MR hereafter) related to O_3 and temperature levels during OPC than OPIs."

• Figure 5 (and text in 3.3): Are the mortality ratios shown average values for all of China?

Reply: The MR values feature the information of NCP region. And I have made it clear in the revised manuscript: "The enhanced mortality rates for OPCs compared to OPIs ***over NCP region*** during May to September for each year of 2017-2019 are illustrated in Figure

5 and attributed to air temperature and/or O_3 concentration changes ($MR_{Temperature}$, MR_{ozone} and MR, see Sec.2).”.

• Section 3.3: Approximately to how many deaths does the increase in mortality risk due to coupled O_3 and temperature correspond?

Thanks. The health impact function is widely applied to evaluate the mortality burden attributable to short-O_3 and heat exposures.

$$\Delta Mort = BMR \times Pop \times (1 - 1/RR)$$

where ΔMort is the excess death due to O_3 exposure or heat exposure; BMR is the baseline daily mortality rate of the specific disease, and Pop is the population exposed to air pollution in different areas. RR represents the concentration-related relative risk of a specific disease caused by O_3/temperature exposures, which is calculated based on Eq. 6&7 in the main text. In this study, to obtain an estimation of daily excess death caused by increased ozone and temperature during OPCs than those during OPIs over NCP, the population and baseline mortality is collected from previous work (see Table S1 in Wang et al., 2021), using population and annual baseline mortality rate for the year of 2018 and assuming no significant changes in Pop (112.62 million) and BMR during 2014-2019. Note that the population in that study is for the Beijing–Tianjin–Hebei (BTH) region, and the baseline mortality rate are assumed evenly distributed across China as the city-level BMR is unavailable. In this study, the total excess mortality is assumed as the total excess deaths caused by increased O_3 and temperatures. Based on the equation above, around 100 daily excess deaths over NCP are attributable to the higher temperatures and ozone level during OPCs than OPIs.
We have added the Text S3 in the supplementary information on the calculation of excess death caused by OPCs. And I have added the estimation of excess deaths in the revised manuscript (Line 325-327): “Moreover, we estimate that around 100 daily excess deaths over NCP are attributable to the higher temperatures and O_3 level during OPCs than OPIs (See Text S3).”

• Line 295-296: Are the thresholds based on the historical observed or modeled values? If based on modeled values, are the thresholds unique for each model?

Thanks. The projections of OPCs in the future are obtained based on the historical threshold of model values. And the thresholds may vary among different models. In fact, the analyses on future projections of OPCs are conducted base on the multi-model ensemble mean from CMIP6 simulations, as we have addressed in the previous submission “The multi-model ensemble means can reasonably capture the observed spatial pattern of coupled extremes and their magnitudes over NCP during 2015-2019 (Fig. S5).” Yet in the updated manuscript, we have made it more clear by adding a sentence (Line 338-339): “And the analyses are based on the multi-model ensemble mean of projected OPCs for different scenarios”.

• Line 376: The value of radiative forcing of tropospheric O_3 seems irrelevant, especially for the next-to-last sentence of the article.

Thanks. Removed.