Responses to referee(s) comments

Dear Editor,

Thank you for handling our manuscript. We appreciate the opportunity to receive very pertinent advices from all referees. Their comments are very constructive, and now we have revised our manuscript taking into consideration of all referees' comments. Based on their helpful suggestions, we believe that now we should have appropriately and adequately addressed all the referees' issues and concerns. Please find our point-by-point responses below.

Anonymous Referee #1:

The authors have presented a manuscript that identifies the health effect of heatwaves and ozone in Beijing, China. The authors identified this effect separately and combined. Furthermore, the authors proposed the identification of the effect for urbanization and synoptic systems, with the aid of composites. In general, the article is interesting but needs improvements.

RESPONSE: Thank you for your valuable time to review this manuscript. We are grateful for your positive feedback on our work. We have carefully read all the constructive comments carefully and we changed the manuscript accordingly. Please find our point-by-point responses below.

Major issues:

1. The methodology to identify episodes of interest is clear but the explanation about the exposure functions is weak. The authors just cite Liu et al (2021) and Ying et al (2017) for the temperature and ozone parameters. Also, I do not understand why the authors use RR for Liu et al and ER for Ying (lines 120-125). Specifically, the authors state: "every 1°C increase in the daily Tmax above 31.5°C, the largest RR of mortality caused by high temperature in northern China was 1.002". Then, we see Table 3, the Tmax for Urban HW, the value is 36.1, then how the authors obtained 4.76(4,76,4.77)? I thought that It might be (36.1-31.5)*1.002 = 4.6092, but not. Please, clarify and include a similar explanation for ozone.

RESPONSE: Thanks for your valuable advice. Although we just referred to the exposure function of Liu et al (2021) and Ying et al (2017) in final calculations, we have extensively reviewed the existing literature of high temperature and O_3 exposure risk in China, which were supplemented at lines 137–142 as follows:

"Previous studies indicated that there were distinctly different magnitudes of human morbidity and mortality caused by high temperature and O_3 overexposure over various geographic regions (Huang et al., 2015; Ma et al., 2015; Wang et al., 2020; Yin et al., 2017). For instance, Huang et al. (2015) revealed that for a 1°C increase above the minimum mortality temperature, the daily mortality increased by 1.04% [95% confidence interval (CI): 0.90 to 1.18], 1.25 (95% CI: 0.71 to 1.79), 1.19 (95% CI: 0.79 to 1.58), and 1.38 (95% CI:0.54 to 2.23) in the nationwide, central China, eastern China, and south China, respectively.".

ER is the last step for the health risk for both HW and O_3 in the final calculation. As Liu et al (2021) pointed out that every 1°C increase in the daily Tmax above 31.5°C, the largest RR of mortality caused by high temperature in northern China was 1.002. During HW periods, Tmax for urban is 36.1°C, so RR=1+(36.1-31.5)*1.002%=1.046092 , ER= (RR-1)*100%=(1.046092-1)*100%=4.6092%. This is a little bit deviation with the 4.76% in Table 2, which is caused by the fact that we calculated the ER of each day and then averaged. It

cannot be ignored that, at the lowest mortality temperature of 31.5°C, the relative risk (RR) is 1. For O_3 exposure, a 10-μg m^{-3} increase in MDA8 O_3 was related to an increase in the total daily mortality of 0.39% (95% CI: 0.04%, 0.75%) in northern China during the warm season (Yin et al., 2017). That is, the coefficients of exposure response function (*β*) between O_3 and total mortality through nonlinear regression is 0.39%. For HW in urban areas, the value of MDA8 O_3 is 197.1, RR=e^(0.39%*197.1/10)=1.0799, ER=(RR-1)*100%=7.99%. We have corrected this paragraph in revised version. Please also see as follow:

"Here, we refer to the coefficients of exposure response function (β) for the high temperature as suggested by Liu et al. (2021), while that O_3 concentration as suggested by Yin et al. (2017) in northern China. In detail, Liu et al. (2021) investigated the mortality caused by high temperature in 84 cities in China from 2013 to 2016, and found that for every 1°C increase in the daily T_{max} above 31.5°C, the largest RR of mortality caused by high temperature in northern China was 1.002 (95% CI: 1.001, 1.004). According to Eq. (2), we can deduce that β_{Tmax}=0.997% (95% CI: 0.996%, 0.999%), note that RR equals to 1 when T_{max} =31.5°C. For O_3 exposure, a 10-μg m^{-3} increase in MDA8 O_3 was related to an increase in the total daily mortality of 0.39% (95% CI: 0.04%, 0.75%) in northern China during the warm season (Yin et al., 2017), that is, β_{Ozone}=0.39% (95% CI: 0.04%, 0.75%)".

2. In Figure 5, the authors show the diurnal cycle for some variables claiming that there are significant differences. Do the authors mean statistical significant in the difference, perhaps after applying a test Mann-Withney? Was this test applied to the point values shown in Figure 5? Finally, neglecting the contribution of O3 precursors to explain the difference in O3 during HW events (line 160), I think it is wrong. Even the authors state in the manuscript that during HW events there are more biogenic VOC emissions. Also, the wind speed is higher during HW, which favours the transport of pollution, from rural to urban areas for instance. Actually, a recent paper published in ACP shows the contribution of local and regional emissions to air quality (https://acp.copernicus.org/articles/21/18195/2021/).

RESPONSE: Thanks for your constructive advice. We add as supplement the Kruskal-Wallis test for the variables of Figure 5, with a significance of the p-values less than 0.001 for each variable in all three cases (Table S2). The Kruskal-Wallis test is a generalization of the Mann-Whitney test and is suitable for multiple groups of independent samples.

Table S2: The Kruskal-Wallis test for figure 5 under different weather conditions (HWs, NHWs and precipitation).

	Source	SS	df	MS	F	P
T	Group	3.45285E+06	2	1726424	145.03	3.21143E-32
	Error	9.23655E+06	531	17394.6		
	Total	1.26894E+07	533			
	Source	SS	df	MS	F	P
RH	Group	1.30882E+06	2	654407.7	73.42	1.13823E-16
	Error	6.90874E+06	459	15051.7		
	Total	8.21756E+06	461			
	Source	SS	df	MS	F	P
HI	Group	1.90775E+06	2	953874.7	107.02	5.75597E-24
	Error	6.30981E+06	459	13746.9		

	Total	8.21756E+06	461			
WS	Source	SS	df	MS	F	P
	Group	544952.7	2	272476.3	30.57	2.29868E-07
	Error	7672602.8	459	16715.9		
	Total	8217555.5	461			
O_3	Source	SS	df	MS	F	P
	Group	1.48979E+06	2	744893.9	82.15	1.4506E-18
	Error	6.94306E+06	463	14995.8		
	Total	8.43285E+06	465			

Thanks again for pointing out our mistakes. We have rephrased and added evidence and explanation of emissions at lines 177–202 as follows:

**"*In general, the difference in O_3 concentration was mainly due to meteorological conditions and the precursors emission paired with photochemical reactions in the boundary layer. We further investigated the diurnal variation for surface air temperature (T), RH, HI, BLH and WS under HW, NHW and precipitation conditions (Figure 5), and these five variables also showed significant differences (passed the Kruskal-Wallis test of 0.001) in the three periods. For HW days, HI raised more by increased air temperature, and although the RH was relative lower, people still suffered from higher apparent temperature than actual air temperature. Under HW conditions, solar radiation reaching the ground heats the atmosphere increasing the near-surface temperature. Warmer air convection promotes atmospheric instability, with increased WS and higher BLH. It is clear that the meteorological variables at daytime were significantly different during HW periods with respect to NHW periods. Similarly, hourly O_3 concentrations also showed significantly difference under different meteorological conditions, and reached the peaks in the afternoon on HW days (Figure 5f). In addition, the contribution of local and regional emissions (transport of pollution between urban and rural areas) to air quality at a city scale should be focused (Thunis et al., 2021), which can also induce urban-rural differences. We assumed that the intraseasonal differences in precursor emissions can be ignored, and further compared the diurnal variation differences in NO_2 and CO and O_3 between different stations (Figure 6). CO and NO_2 levels were higher at traffic stations than urban and suburban stations due to enhanced emission from vehicles, and the lowest CO and NO_2 levels appeared at rural stations. Generally speaking, high precursor levels are supposed to correspond to high resultant levels, but the lowest O_3 levels were found at traffic stations, followed by rural stations, then urban and suburban stations. Since automobile exhaust in the traffic and urban stations also caused heavily NO emission (Colvile et al., 2001), ambient O_3 can be titrated by NO via the reaction $NO + O_3 \rightarrow NO_2 + O_2$ (Gao et al., 2020; Murphy et al., 2007; Sillman, 1999), this process in turn led to higher NO_2 levels and the loss of O_3 in traffic and urban areas. As for rural stations, low pollutant emissions may be the primary reason for the lower O_3 levels. Note that although the CO and NO_2 emissions were significantly higher at urban stations than suburban stations, there was less difference in O_3 concentrations between these two-type stations, which may be due to O_3 consumption induced by titration at urban stations, or more biogenic VOCs at suburban stations. This is because that the difference in O_3 concentrations between the rural and the suburban stations were the largest in the afternoon, while the difference in CO and NO_2 levels were the smallest, indicating that anthropogenic emissions have less impact in suburban areas, coupled*

with more than half of suburban stations are covered by vegetations leading to more bio-VOCs emissions”

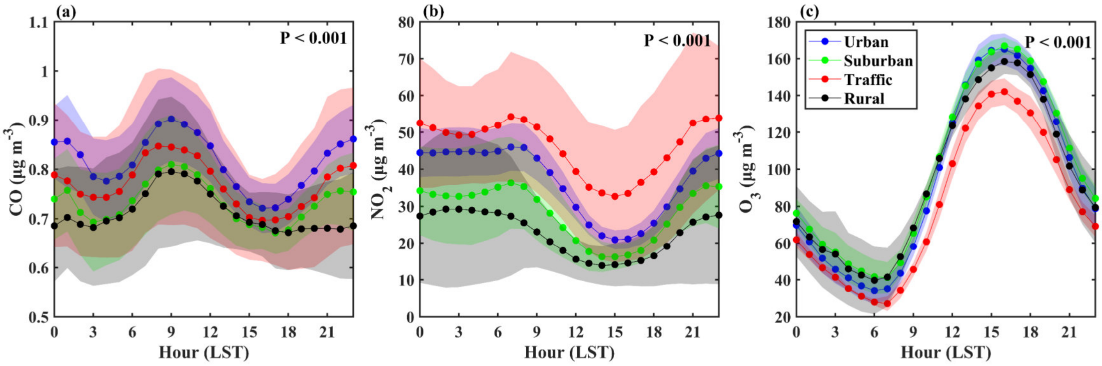

Figure 6: The diurnal variation of (a) CO, (b) NO, (c) O_3, under different stations (shading indicates standard deviation, P < 0.001 means pass the significance test).

Minor issues:

1. Line 77-80, one paragraph of just one sentence. Each paragraph should have at least three sentences, intro, body and conclusion.

RESPONSE: Thanks for your constructive suggestion. We have rephrased it as follows: “*Ground-level O_3 observation data during summertime (June–August) of 2014–2019 were retrieved from Beijing Municipal Ecological and Environmental Monitoring Center. After quality control, and excluding stations with a missing-values rate for the O_3 hourly concentration of more than 10%, 31 air quality stations [AQSs; including 11 urban stations, 11 suburban stations, three traffic stations (road monitoring stations for traffic air quality), and six rural stations] are ultimately used in this study. In order to better assess the relationship between O_3 pollution and the meteorological variables, we selected 29 automatic weather stations (AWSs) closest to the environmental monitoring stations from the high-density AWS network. Specific geographic location information can be found in Figure 1 and Table 1. Hourly 2-m air temperature, relative humidity (RH), the daily maximum temperature (T_{max}), and 10-m wind speed (WS) of these 29 AWSs were obtained from the National Meteorological Information Center of the China Meteorological Administration, and then heat index (HI) was retrieved as shown in Rothfusz (1990) as Eq. (1):*”.

2. Line 115, then again, which beta did you use?

RESPONSE: Thanks for pointing this out. For high temperature, β_{Tmax}=0.997% (95% CI: 0.996%, 0.999%), for O_3, β_{Ozone}=0.39% (95% CI: 0.04%, 0.75%).

3. Line 123-124, why do the authors use RR for temperature and then ER for ozone?

RESPONSE: Thanks for your question. We have elaborated on this question in your major issue 1, and have revised.

4. Figure 2 is not good. Provide a better figure.

RESPONSE: Thanks for your constructive suggestion. We have replotted Figure 2, and revised a lot. Please see also as follow:

"Figure 2 shows the time series of the HW, NHW, O_3 pollution, and precipitation days, and the interannual and intraseasonal variations of HW and O_3 pollution days. For interannual variation, the total days of O_3 pollution in summer was relative stable, while the total days of HW increased slightly. For intraseasonal variation, O_3 pollution was the most serious in June, while the most frequently HW events in July. Obviously, showing that higher O_3 pollution levels (>160 μg m^{-3}) were always accompanied by most HW periods (approximately 79.2% of HW days) in Beijing (Figures 2a and 3b), which were mainly in the middle of summer."

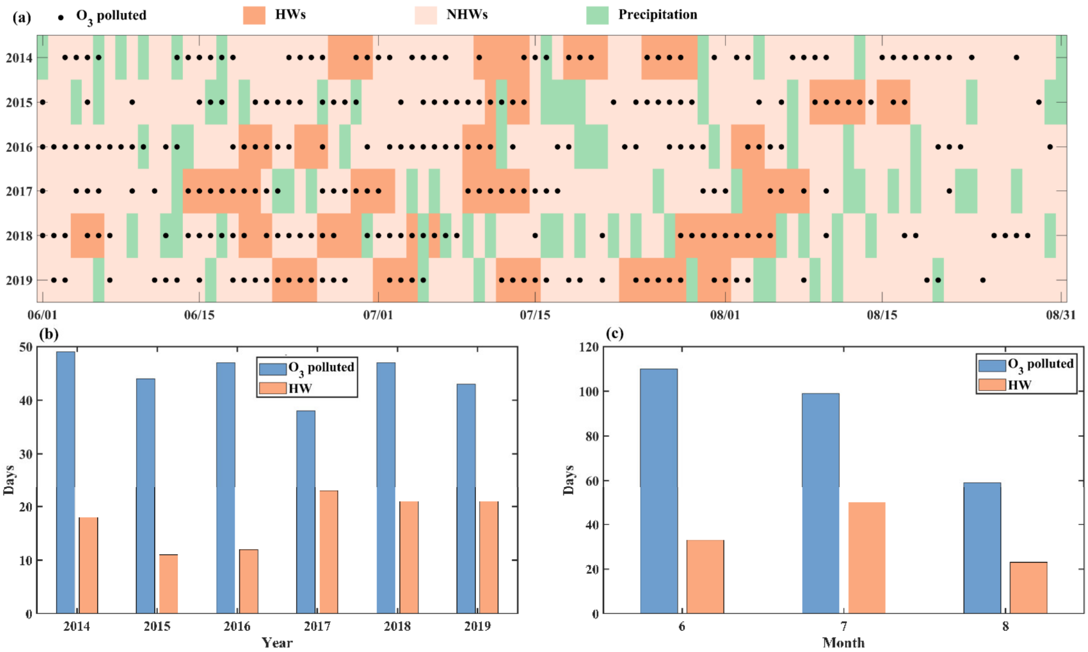

Figure 2: (a) Time series of weather types, in which the black dots indicate O_3 pollution that occurred on that day. Interannual (b) and intraseasonal (c) variations in summertime O_3 pollution and HW days.

5. Line 151, which test do the authors use?

RESPONSE: We used the Analysis of Variance test. We have stated it in revised version.

6. Lines 165-174: I understand that we might expect lower risk in traffic and urban station for ozone, but you mentioned in line 166 that ozone caused a reduction of 2.44% which means risk lower than 1. More explanation is needed.

RESPONSE: Thanks for your kind suggestion. We have revised the description in this section and explained it in more detail at lines 203–218 as follows:

"Moreover, the high temperatures on HW days not only brought a higher public risk related to high-temperature exposure, but also increased mortality related to O_3 exposure. During HW periods, high temperatures and strong solar radiation accelerate the rate of the photochemical reaction that produces O_3 (Pu et al., 2017; Sun et al., 2017), favouring the production and accumulation of O_3, thereby aggravating health risks. Regardless of the type of stations, the O_3 and high-temperature stressful conditions suffered by the human during HW days has greatly increased. Specifically, for all stations, HWs have increased the ER caused by high temperatures and O_3 by 3.867% (90% CI: 3.863%, 3.875%) and 7.9% (90%CI: 0.78%, 15.78%), respectively (Table 2). The high temperature risks were mainly manifested as followings: urban stations >

traffic stations > suburban stations > rural stations, but the health risks aroused by O_3 exposure in different underlying surface stations were more difficult to quantifying due to the complexity of O_3 photochemical reactions. As mentioned above, urbanization-enhanced NO or CO titration reduced more O_3 loss in urban areas, which was more pronounced over traffic stations. For suburban stations, the abundant biogenic VOC emitted by vegetation also contributed to O_3 generation, bio-VOC emissions enhanced more especially in hot days (Ma et al., 2019; Trainer et al., 1987; Wang et al., 2021a). As a result, O_3 exposure risks in Beijing were mainly characterized by suburban stations > urban stations > rural stations > traffic stations. Urbanization seems to have increased the ER induced by both high temperatures and O_3 exposure. In details, summertime HW, O_3 and compound ER increased by 1.67%, 0.20%, and 1.89%, respectively, compared to rural stations. Note that urbanization has alleviated O_3 pollution to a certain extent, and the health risk of O_3 at stations with developed transportation was even lower than that of rural stations."

7. Line 177, Please, do not overuse abbreviations, WPSH is not needed.

RESPONSE: Thanks for your advice. We deleted it.

8. Line 196, UHI is not defined.

RESPONSE: It have defined at line 64.