Response to Reviewer #1

This paper presents the results of the aerosol impact on urban heat island (UHI) over Beijing. The authors first analyzed the 2016-2020 observations to link UHI intensity (UHII) with wind direction and $PM_{2.5}$ pollution, then used WRF-Chem regional model to do the perturbation simulations of a haze episode in January 2010 to substantiate the underlying mechanism of such linkage. The general conclusion was that aerosols, either locally emitted or transported and cumulated through regional circulation, reduced UHII via aerosol-radiation interactions over the study region. Though the research topic fits into the ACP scope and the paper was concisely written, more analysis and discussions are needed to reconcile the mismatch of time period and time scale between the observational and modeling analysis to make the conclusion robust under different seasons and aerosol pollution conditions. This is especially important to the regions like Beijing, who has experienced the rapid changes in landscape and pollutant emissions over the past decade. In addition, quite a large portion of the discussion were descriptive and qualitative, while more quantitative analysis should and can be done with the available observation and simulation data.

- Reply: Thanks for the careful review and the comments are valuable to enhance the quality of our manuscript.
- We have conducted additional simulations and added more discussions to reconcile the mismatch of time period and time scale between the observational and modeling analysis in Section 3.1. The additional simulations are for a light pollution event occurred in the Spring of 2018 to make the conclusion robust under different seasons and aerosol pollution conditions.
- We found that due to lower $PM_{2.5}$ concentration, ARE reduces UHII by less than 0.2 K. The lower concentration also diminishes absorption of shortwave radiation during the daytime which reduces downward longwave radiation, leading to weakened UHII in nighttime. We have also added more quantitative analysis in Section 3 (marked in the revised manuscript).

On top of the above comments, the authors are also expected to address the following specific points:

Section 2.1: Can the authors elaborate what is the criteria they chose the weather stations for UHI estimate? This information is important since the selection of urban vs. rural stations may slew the results. Only 2 urban weather stations were selected for the analysis. How representative were they for Beijing?

- Reply: We chose these weather stations because only data at these stations in Beijing are available.
- We have already clarified in Section 3.1 that our observation-based results are representative for UHI in western and northern sides of Beijing as we used stations there as rural.
- “We used rural stations located in the west and north of Beijing as rural in the calculation of UHII, and PM2.5 concentrations are usually much lower there (Fig. S2).”
- As a result, we found different results compared with previous research and we further conducted model simulations to understand the underlying mechanism.

Section 3.1 – Fig. 1 discussion: $PM_{2.5}$ data from all observation sites, urban or rural, were selected for daily average calculation to distinguish between polluted vs. clean days. Was there large $PM_{2.5}$ gradient between the urban and rural sites? What was the impact of such $PM_{2.5}$ gradient if it existed?

- Reply: $PM_{2.5}$ concentrations in urban stations were higher than those in rural stations by 6.9 μg m^{-3} on average over 2016-2020. Under polluted conditions, the difference reached 14.4 μg m^{-3} on average. Such gradient may result in the overestimation of pollution for those rural stations and change the statistical results.
- We further evaluated the results based on the standard that $PM_{2.5}$ concentrations at all stations meet the criterion of clean or polluted (Fig. R1) and the standard that average $PM_{2.5}$ concentration of all urban stations and rural stations should meet the criterion of clean or polluted (Fig. R2).
- Compared with Fig. R3 (Fig. 1 in the revised manuscript), we found similar distributions and some minor difference in mean values. When $PM_{2.5}$ concentrations at all stations meet the criterion of clean or polluted, we found the mean values increased by 0.03-0.04 K for clean conditions but decreased by 0.14 K during daytime and 0.06 K during nighttime. When we used average $PM_{2.5}$ concentration of all urban stations and rural stations to determine clean or polluted, mean values decreased by 0.01 K for clean conditions and increased by 0.01 K and 0.06 K during daytime and nighttime, respectively.
- The changes are not notable and we added these comparisons in the revised manuscript.
- In the revised manuscript, we further added Table R1 to show the distribution of daily average urban and rural $PM_{2.5}$ concentration under clean and polluted conditions.
- We found that there were 17.07% overestimation in rural stations because of the gradient between urban and rural areas. However, we also observed that $PM_{2.5}$ concentrations were over 60 μg m^{-3} for most of considered days. Besides, we found pollution in urban and rural shows a good occurrence, and we thus believe using the daily mean $PM_{2.5}$ concentration averaged over all stations can properly represent the regional feature of aerosol pollution.
- Therefore, to better characterize the regional air quality and avoid effects of individual station, all stations within the administrative divisions of Beijing were selected to calculate mean $PM_{2.5}$ concentration to distinguish between polluted and clean days.

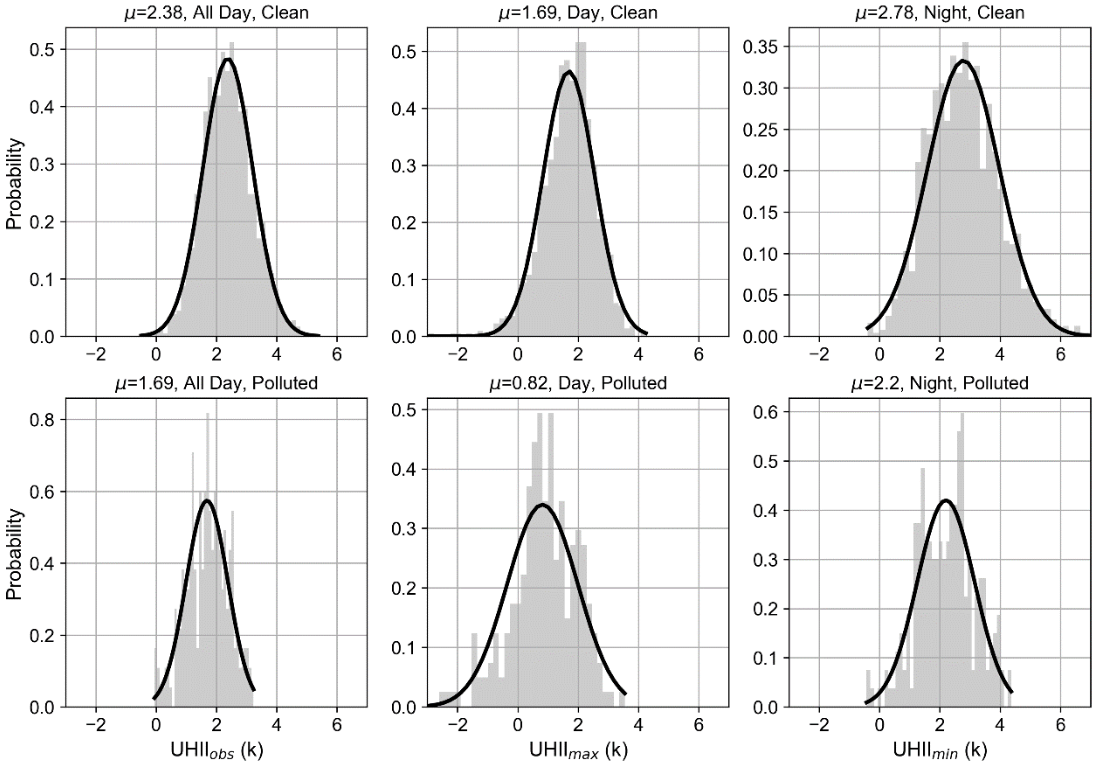

Figure R1: Probability distribution of UHII$_{obs}$ (a, d), UHII$_{max}$ (b, e) and UHII$_{min}$ (c, f) under different pollution conditions. Clean means $PM_{2.5}$ concentrations of all stations are below 75 μg m^{-3}. Polluted means $PM_{2.5}$ concentrations of all stations are equal or over 75 μg m^{-3}. The bold curve in each subgraph is normal distribution curve, and μ denotes the average value.

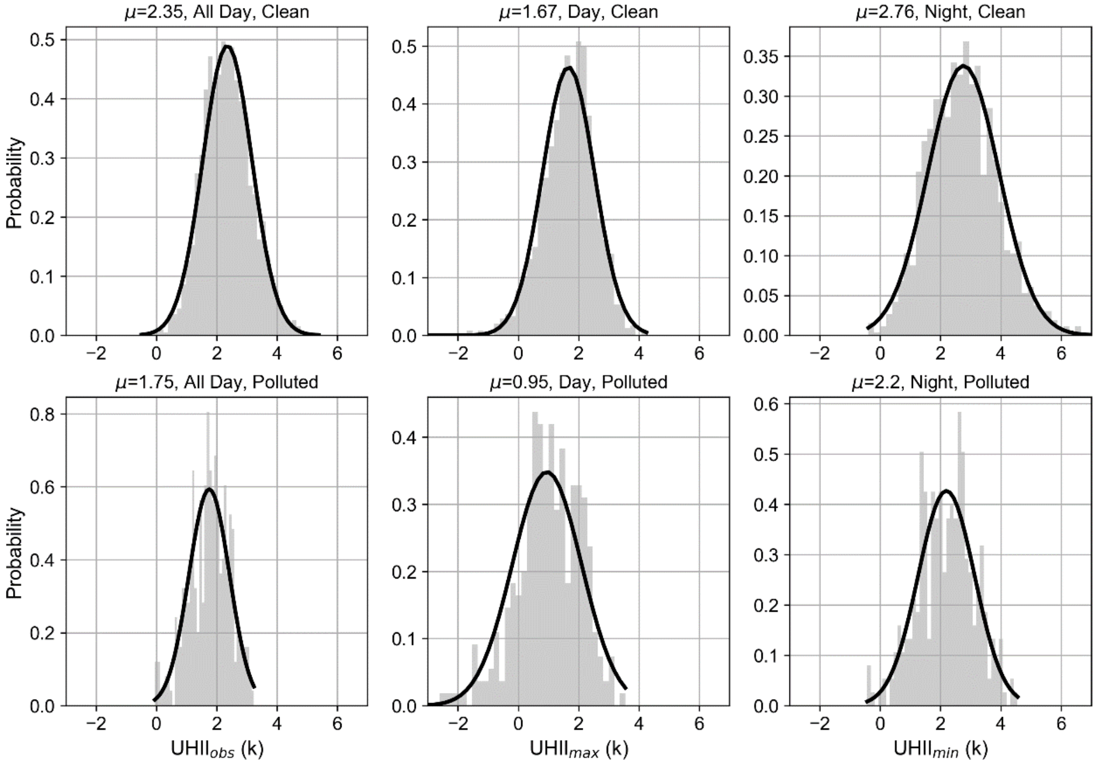

Figure R2: Probability distribution of UHII$_{obs}$ (a, d), UHII$_{max}$ (b, e) and UHII$_{min}$ (c, f) under different pollution conditions. Clean means both average $PM_{2.5}$ concentrations of all urban stations and those of rural stations are below 75 μg m^{-3}. Polluted means both average $PM_{2.5}$ concentrations of all urban stations and those of rural stations are equal or over 75 μg m^{-3}. The bold curve in each subgraph is normal distribution curve, and μ denotes the average value.

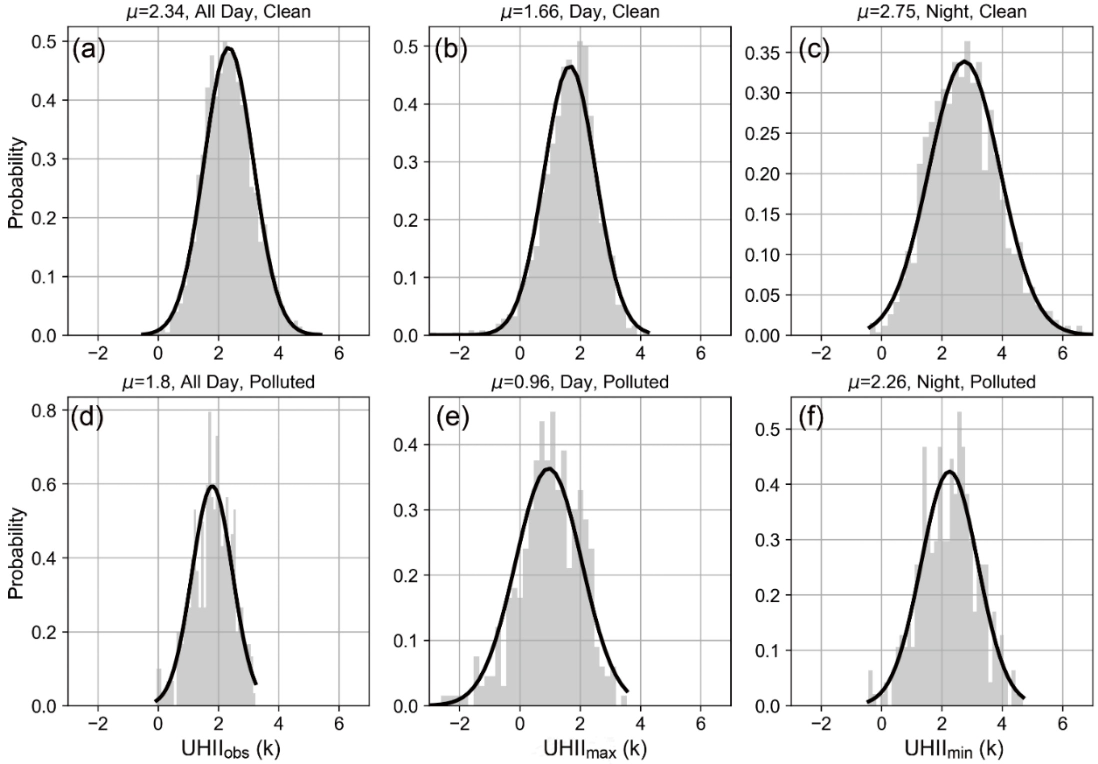

Figure R3: Probability distribution of $UHII_{obs}$ (a, d), $UHII_{max}$ (b, e) and $UHII_{min}$ (c, f) under different pollution conditions. Clean means average $PM_{2.5}$ concentration of all stations is below 75 μg m^{-3}. Polluted means average $PM_{2.5}$ concentration of all stations is equal or over 75 μg m^{-3}. The bold curve in each subgraph is normal distribution curve, and μ denotes the average value.

Table R1: Distribution of daily average urban and rural $PM_{2.5}$ concentration (unit: μg m^{-3}) under clean and polluted conditions. Here, $PM_{2.5\_average}$ represents average $PM_{2.5}$ concentrations of all stations; $PM_{2.5\_urban}$ represents average $PM_{2.5}$ concentrations of all urban stations; $PM_{2.5\_rural}$ represents average $PM_{2.5}$ concentrations of all rural stations.

	$PM_{2.5\_average} \geq 75$ (369 days)	$PM_{2.5\_average} < 75$ (1373 days)
$PM_{2.5\_urban} \geq 75$	366 (99.19%)	18 (1.31%)
$PM_{2.5\_urban} < 75$	3 (0.81%)	1355 (98.69%)
$PM_{2.5\_rural} \geq 75$	306 (82.93%)	12 (0.87%)
$PM_{2.5\_rural} < 75$	64* (17.07%)	1361 (99.13%)

*: These 64 days consist of 4 days with $PM_{2.5\_rural} < 50$, 8 days with $50 \leq PM_{2.5\_rural} < 60$ and 52 days with $60 \leq PM_{2.5\_rural} < 75$.

Section 3.1 – Fig. 2 discussion: A figure or table shows the average $PM_{2.5}$ conc. over urban/rural areas under each prevalent wind directions should be provided to support the argument.

- Reply: We have added a table shows the average $PM_{2.5}$ concentration over urban and rural areas under each prevalent wind directions in the manuscript.

Table R2. Average $PM_{2.5}$ concentration (unit: μg m^{-3}) in urban and rural areas under each prevalent wind directions.

Wind directions	Easterly	Southerly	Westerly	Northerly
Urban $PM_{2.5}$	58.23	53.88	52.24	49.49
Rural $PM_{2.5}$	50.82	47.34	44.68	43.31

Section 3.2 – in Fig. S2, how did the authors derive the observed time series of UHII? Was it the average UHII at the same days/hours from 2016 to 2020, or other? Since the modeled UHII reflected the heavy polluted condition, why not compared the modeled UHII with the observed one under the pollution condition?

- Reply: We conducted model simulations of a typical haze event that occurred in January 2010 in Beijing, and Fig. S2 (Fig. S5 in the revised manuscript) is the comparison done with observations during this haze event.
- It is under the heavy polluted condition.
- In the revised manuscript, we also added an evaluation of UHII during a light polluted case in 2018 (Fig. R4, Fig. S6 in the revised manuscript).

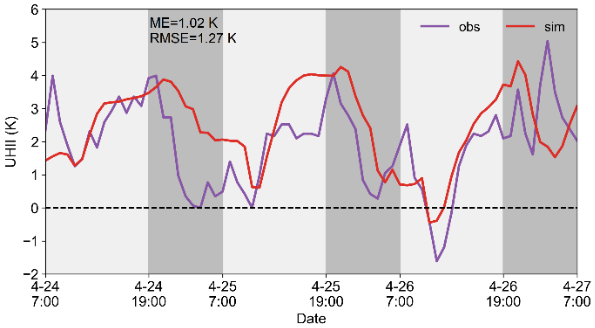

Figure R4: Observed and simulated UHII by AF in Case_2018 in Beijing. Observations are obtained from the stations listed in Table S1.

In Line 158: "…and differences in values are generally within the trusted range". What is the trusted range of UHII comparison? How did modeled wind and temperature compare to the respective observations?

- Reply: There is no specific definition for the trusted range of UHII comparison.
- However, according to previous explorations, if the model can successfully reproduce the temporal variation of UHII and the mean bias and root mean square error are smaller than 2 K, the results are acceptable.
- In our simulation, we successfully reproduce the temporal variation of UHII, and the mean bias and root mean square error are 1.16 K and 1.47 K in Case_2010 and 1.02 K and 1.27 K in Case_2018, respectively.
- The modeled wind and temperature for the 2010 case have been shown in our previous work (Gao et al., 2016).
- We also added the performance for Case_2018 in Fig. R5 (Fig. S4 in the revised manuscript).

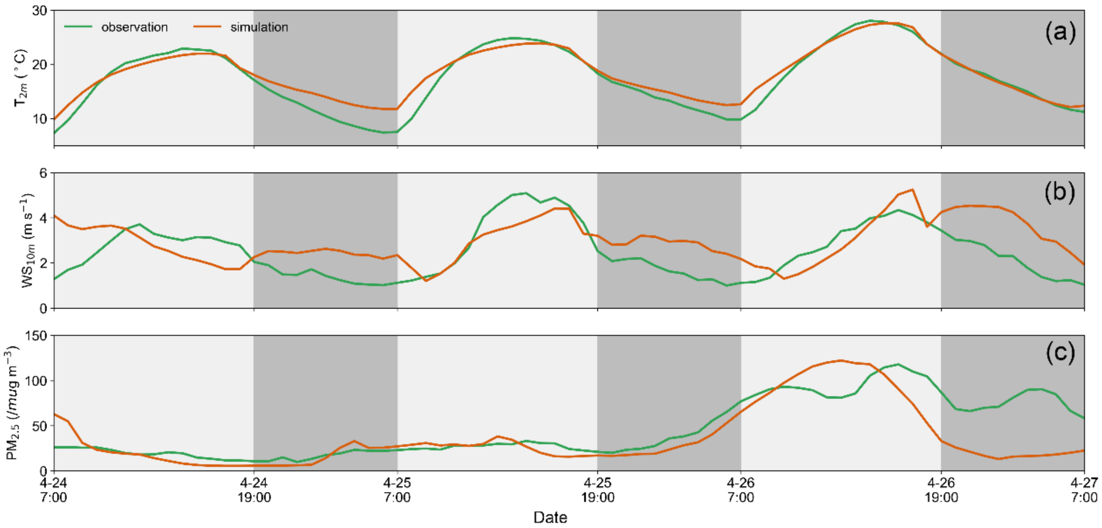

Figure R5: Simulated and observed 2m air temperature (a), 10m wind speed (b) and near-ground $PM_{2.5}$ concentration (c) in Case_2018. Observations are obtained from the stations listed in Table S1 and Table S2.

Line 168: how was heat storage calculated?

- Reply: In WRF-Chem model, heat storage is calculated with land surface model, and we applied Noah land surface scheme for non-urban grids and Urban Canopy model for urban grids. In Noah land surface scheme, heat storage is calculated using the following equations:

$$G = \left(1 - F_{veg}\right)G_b + F_{veg}G_v$$

$$G_b = \frac{2\lambda_{isno+1}}{\Delta z_{isno+1}}\left(T_{g,b} - T_{isno+1}\right)$$

$$G_v = \frac{2\lambda_{isno+1}}{\Delta z_{isno+1}}\left(T_{g,v} - T_{isno+1}\right)$$

- where, F_{veg} denotes fractional vegetated area, G_b and G_v are heat storage for bare ground and vegetated ground, respectively, and λ_{isno+1} represents thermal conductivity of the surface layer of snow or soil; z_{isno+1} is layer thickness of the surface layer of snow or soil, T_{isno+1} represents temperature of the surface layer of snow (when $isno+1<0$) or soil (when $isno = 0$), and $T_{g,b}$ and $T_{g,v}$ stand for ground surface temperature at bare ground fraction and vegetated fraction, respectively.
- In Urban Canopy model, heat storage is calculated using

$$G = G_0 + 2\int_0^{z_r}\left[\frac{\partial(\rho_b c_b T_b)}{\partial t}\right]d_z$$

- where, G_0 is the surface heat flux into the ground per unit area, including roof and road, and ρ_b, c_b, and T_b are density, specific heat, and temperature of buildings.
- We have added these descriptions in the revised manuscript.