Response: We thank the reviewer for the comprehensive review and insightful comments on the manuscript. For detail, please refer to the responses below.

The reviewer's comments are in black and the responses are in blue. If responses are from the revised manuscript, they are in italic.

1) The plot showing the zonal mean concentrations from CAM-chem should be revised.

(1) The focus of this paper is on the US, not the zonal mean. (2) The authors are using the high resolution WRF-CMAQ model to examine the ozone over the US. This model has different emissions than CAM-chem and different resolution. So why show zonal mean distributions in CAM-chem? To me, it only makes sense to examine the latitude height cross-sections using WRF-CMAQ. Moreover, I would suggest the authors show these cross-sections over the longitudes of interest (i.e., the US) as well as a more circumscribed latitude range (why show the S.H.?). In addition, I would recommend including a panel showing the sensitivity to reducing the methane concentration. The impact of methane (and the boundary conditions) seems to be essential to determining the future response of ozone over the U.S. In addition, some of the explanations given in this section seems speculative. There are a number of instances where the authors explain a feature of the figures without giving adequate proof.

Response: We would like to look at the distributions in global models and further narrow down to regional models because the boundary driver is from the global models. As the reviewer suggested, we removed south hemisphere and now only show the distributions in the northern hemisphere. We also explored the impact of methane on ozone, as the reviewer advised, shown below.

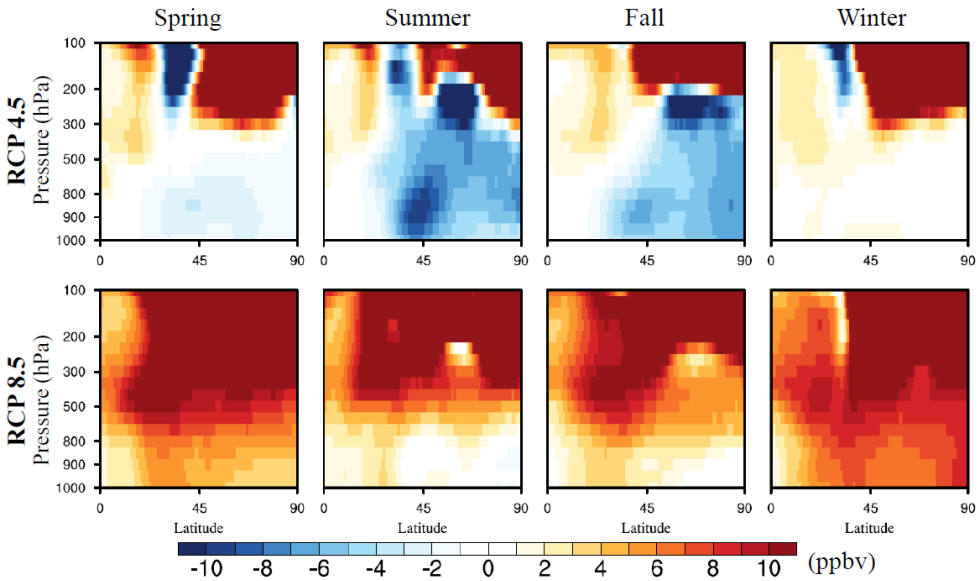


Fig. 4. Zonal mean vertical ozone changes from CAM-Chem under future climate (2057-2059 minus 2001-2004) for RCP 4.5 (top panel) and RCP 8.5 (bottom panel).

The impact of methane on ozone has been added to the manuscript (Page 13, Line 383-395) and also shown below:

*Considering the larger ozone increases in spring and winter in RCP 8.5 compared to RCP 4.5, and the large increase in methane concentrations in RCP 8.5, a sensitivity study was conducted to explore the impact of methane on ozone concentrations. Under RCP 8.5, the methane level in 2050 is 2,740 ppbv, which is 56% higher than the level in 2000 (1,751 ppbv). The simulations were conducted using CAM-Chem for the period of 2050s by maintaining the methane concentrations at 2000 level. Fig. 5 (i-l) shows ozone changes in 2050s compared to present climate under RCP 8.5 without methane increase. Compared to Fig. 5(e-h), in spring and winter, the ozone increases areas and magnitudes were dramatically reduced, leaving small areas of ozone increase resulting from titration effect. In summer and fall, much larger decrease (comparing Fig. 5i, k and f, g) occurs when methane concentrations maintain at 2000 level. The sensitivity study clearly addressed the significant role of methane concentrations play on ozone concentrations, and the impact could be as large as 4-8 ppbv.*

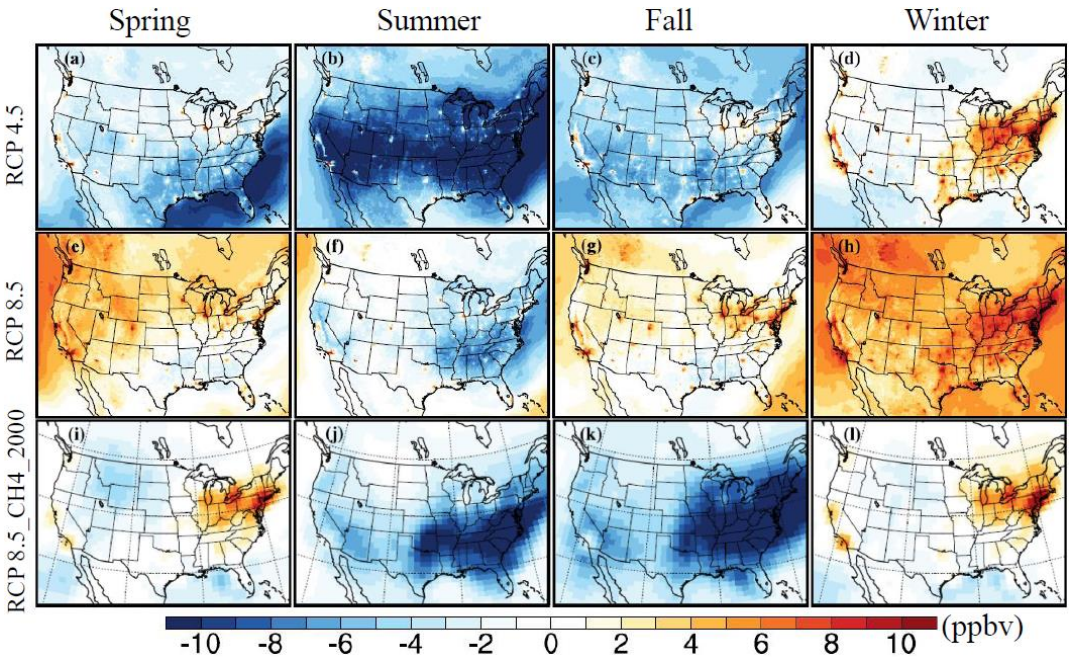


Fig. 5. Seasonal mean surface ozone changes from CMAQ outputs under future climate (2057-2059 minus 2001-2004) for RCP 4.5 (a-d), RCP 8.5(e-h), and the bottom panel (i-l) shows ozone changes by the end of 2050s without methane increases in RCP 8.5 (ozone in 2050s with 2000s methane concentrations – ozone in 2000s).

2) The slides showing the PV and their interpretation are not convincing. (1) Increased PV can be indicative of more stratospheric influence, but on the other hand it might simply suggest increased cyclonic flow. (2) It is not clear that this change in PV between the simulations is robust. The authors need to test its significance to determine if it is indeed robust, but point (1) suggests that even if robust an interpretation of this figure is difficult.

Response:

Thanks for the reviewer's comment on the PV. Considering the reviewer's points, we removed the PV plot and further checked the results. The STE increase in both RCP 4.5 and RCP 8.5, but only RCP 8.5 shows strong ozone increase in western US. Thus, the STE may not the major driver but the increased methane in RCP 8.5 plays the most significant roles. As the reviewer suggested in the comment 1, we added methane sensitivity study to explain the ozone increases in RCP 8.5 (Fig. 5 shown above).

3) More thought needs to be taken in testing the significance of the author's conclusions. This is particularly true for Figure 7. However, it also applies to interpretation of almost all the author's figures. I'm not suggesting the authors necessarily change their figures (although in many cases it would be convenient to show where the results are significantly different), but to be careful of the interpretation. For example, I have no problem in the author's showing in Figure 9 differences in mean ozone and % differences over 75 ppbv and 60 ppbv; however, when the authors claim that there has been a change between one period and another they need to show that this change is not a statistical fluke (i.e., that it is significant).

Response:

We thank the reviewer for pointing out the importance of statistical significance. We have added the statistical significance in Fig. 8 (shown below). The descriptions have also been revised in the manuscript (Page 16, Line 475-502), also shown below. Fig. 7 has been deleted as the reviewer suggested in the comment 2.

*Under RCP 8.5, the mean MDA8 shows increases across US except Southeast, during the entire period compared with non-heat wave period, and the increase are all statistically significant, ranging from 0.3 ppbv to 2.0 ppbv. The ozone exceedance of 60 ppbv and 75 ppbv during the non-heat wave period is on average 1-8% and 0-4% lower than the entire period respectively. The daily maximum temperature (TMX) under this scenario is statistically higher during the entire period than non-heat wave period, ranging from 0.8 to 2.0 °C. One of the major reasons Southeast does not show statistically significant increase in MDA8 is its position adjacent to the ocean and its small diurnal temperature variations (Fig. S2 in the supplement). Under RCP 4.5, statistically significant MDA8 increase occurs in five regions, however, the increase magnitudes (maximum of 0.7 ppbv) are much smaller than RCP 8.5. The 95% confidence interval of the MDA8 differences between entire period and non-heat wave period was also shown in Fig. 8. The upper 95% limit indicates 0.4 ppbv to 1.5 ppbv increase under RCP 4.5 and 1.2 ppbv to 3.2 ppbv*

*increase under RCP 8.5 was resulted from heat waves without including Southeast (increase of 0.6 ppbv). Even though the temperature increase in RCP 4.5 due to heat waves is statistically significant, the ozone precursors including NMVOC and NOx decreased dramatically and the methane emissions decrease by ~10% (Table 2) as well. As is explained in section 5.2, methane is the major contributor in ozone increase in RCP 8.5, and without enough ozone precursor emissions in RCP 4.5, the heat waves may not play as significant role as it does in RCP 8.5.*

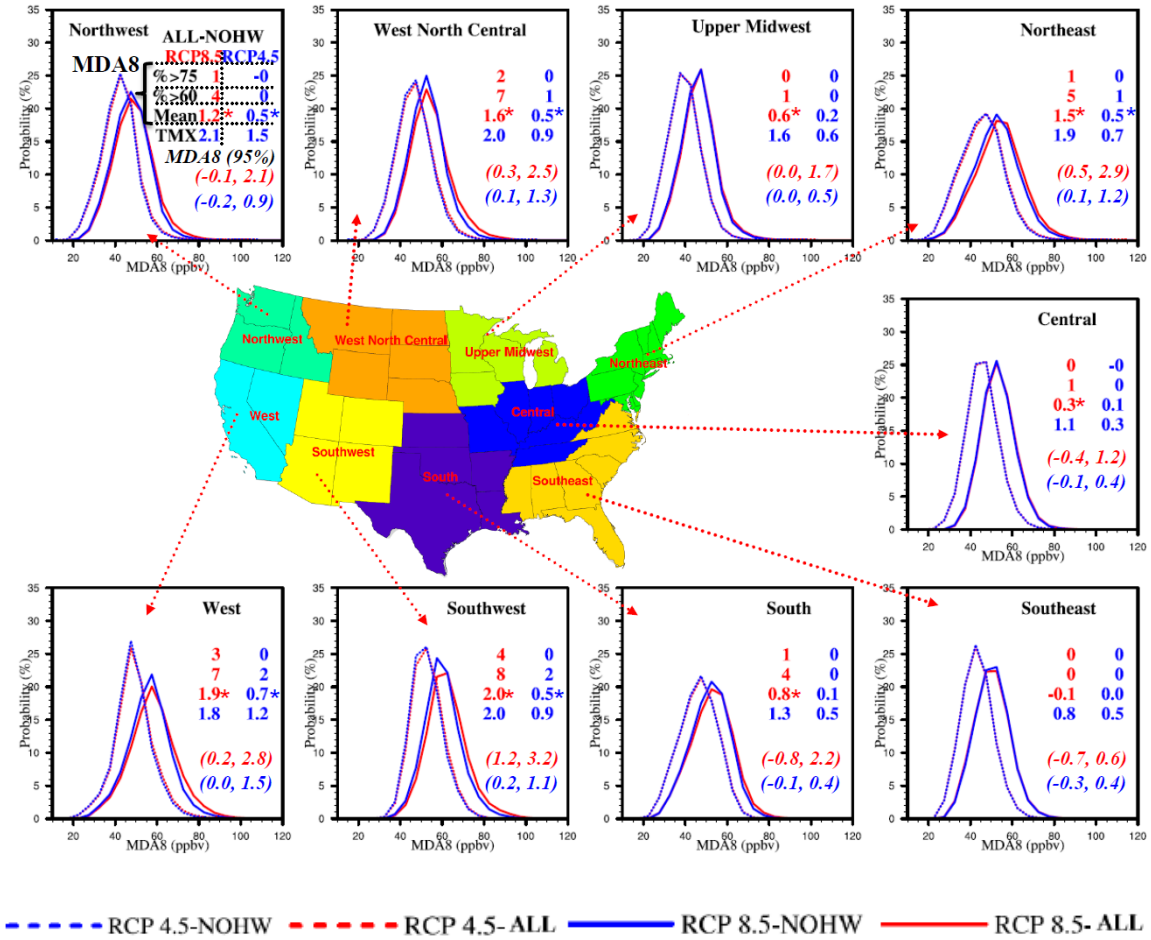


Fig. 8. Distributions of MDA8 during the entire period (referred to as ALL) and non-heat wave period (referred to as NOHW) for RCP 4.5 and RCP 8.5 from June to October. There are two columns of numbers (top four rows), and they represent percentage of MDA8 ozone exceeding 70 ppbv (top row) and 60 ppbv (second row), mean MDA8 ozone (third row, with unit of ppbv) and mean daily maximum temperature (TMX, fourth row, with unit of °C) for both scenarios. The bottom two rows numbers in italic in parenthesis represent the 95% confidence interval of MDA8 differences (ppbv) between entire period and non-heat wave period under RCP 8.5 (red numbers) and RCP 4.5 (blue numbers). Statistical significance was tested for the mean MDA8 differences and marked with star to indicate statistical significant at the level of 0.05. All TMX mean differences are statistically significant.

Minor Comments:

1) The descriptions of CAM-chem and WRF are not at all consistent. While a great-deal of detail is given on the WRF parameterizations the parameterizations within CAMchem are not mentioned.

Response:

We added the descriptions of CAM-Chem in the revised manuscript (Page 5, Line 131-146), which is also shown below.

*The atmospheric chemistry integrated in the atmosphere component CAM4 in the CESM is referred to as the CAM-Chem. The descriptions and parameterizations have been discussed in detail by Lamarque et al. (2012). In summary, the major physics used in CAM4 include Zhang-McFarlane deep convection scheme (Zhang and McFarlaneb, 1995), Hack shallow convection scheme (Hack et al., 2006) and Holtslag and Boville (1993) planetary boundary layer process. The atmospheric chemistry was adapted from Model for OZone And Related chemical Tracers (MOZART) – 4 and bulk aerosol model was used in CAM-Chem (Emmons et al., 2010; Lamarque et al., 2005). The CAM-Chem has been widely used and evaluated on its representation of atmospheric chemistry in the atmosphere (Aghedo et al., 2011; Lamarque et al., 2012; Lamarque et al., 2011b; Lamarque and Solomon, 2010; Lamarque et al., 2011a). The atmospheric chemistry is computed at the same resolution (horizontal and vertical) as the atmosphere model. In order for the performed simulations to be consistent with the simulations performed for CMIP5 (without chemistry; Meehl et al., 2012), the simulated chemical fields do not affect the simulated climate, eliminating the risk of generating a different climate than the original CESM simulations.*

2) Could the authors clarify the relationship between the models? I assume the CAMchem meteorological initial and boundary conditions were used to drive WRF and that the WRF meteorology was used in CMAQ? Please be explicit here.

Response:

CESM/CAM-Chem outputs are used to provide initial and boundary conditions for WRF and CAMQ. WRF outputs are used as the meteorological input for CMAQ. This relationship has been revised in the manuscript (Page 8, Line 189-193) and shown below:

*Dynamical downscaling is a technique that uses the outputs from global climate or chemistry models to provide the initial and boundary conditions for the regional models. In this study, three hourly global climate (CESM) and chemistry (CAM-Chem) model outputs are used to provide the*

*initial and boundary conditions for regional climate (WRF) and chemistry (CMAQ) simulations, respectively.*

3) Page 11321, line 13. "present climate (1850-2005): : :" This entire period doesn't really correspond to the "present climate".

Response:

"present climate" has been changed to "historical simulations" (Page 6, Line 173).

4) Species mapping. Which chemical mechanism is used in CMAQ? CAM-chem has many more species than indicated in this mapping.

Response:

The carbon bond mechanism in CMAS was used in this study. This is true that many more species are available for mapping between the two chemical mechanisms. However, because the short-lived species will quickly adjust to the specification of concentrations of the long-lived chemical constituents through the boundary conditions, only those long-lived constituents are mapped.

5) The algorithm used to calculate emissions in CMAQ was not altogether clear to me. Did CMAQ use emissions as calculated in 2005 for the RCPs? Were these RCP emissions then scaled using the EPA emission inventories and SMOKE for the years 2001-2005? Or did CMAQ use EPA emissions inventories for the years 2001-2005?

Response:

We revised the description. We use SMOKE to run 2005 emissions based on NEI, and then use the ratios to scale the emissions in 2001-2004. The revised descriptions (Page 8, Line 238-242) are as below:

*As 2005 represents the start year of RCP scenarios in US, the 2005 US EPA's National Emission Inventory* was processed by Sparse Matrix Operator Kernel Emissions (SMOKE) 2.7. The 2005 emissions were used to scale back the emissions from 2001-2004.*

6) Page 11324 "However, these statistical methods have not been used in climate studies". Assuming CAM-chem was not forced by meteorological analysis, but was run in the historical period using GCM winds and that WRF was forced by CAM-chem meteorology it is hard to understand the emphasis the authors place on paired space and time evaluation (p11324). The model will not replicate the observed meteorology so it makes little sense to compare against

* http://www.epa.gov/ttn/chief/net/2005inventory.html#inventorydata

observations taken at the same time (in fact monthly averages probably make sense). The authors do not say what the temporal resolution of their emission dataset is. I am assuming it does not include variations on the daily timescale. So this emphasis on paired space and time evaluation is confusing. I assume paired space-time comparisons means comparing at the same point and time as the measurements were taken. Please clarify how the model evaluation was made and the meteorology used to drive the WRF-CMAQ simulations.

Response:

Thanks for the reviewer's comments.

The reviewer is right, and the meteorology does not replicate the observed meteorology in a climate study. However, we did find great improvement compared with observations after downscaling, as stated in Gao et al., 2012. In global simulations, usually monthly mean emissions are provided. However, the CMAQ emission inputs require hourly data, and diurnal variations are more likely to improve the air quality simulations. Thus, in the emission dataset, we did include the daily and hourly variations. In addition, the temporal resolution of AQS is hourly to daily, and we used the available AQS data (hourly basis) to conduct evaluation. Thus, the paired time and space comparison means hourly comparison over the observational sites.
This has been revised (Page 10, Line 284-287) in the manuscript.

7) The authors do not say what timescale they compare the model and the observations or using what ozone metric (i.e., is it MDA8?) They do not say where the measurements are from or what part of the country is covered. More information is needed here.

Response:

Hourly ozone was used to for the comparison as explained by response to comment 6. This has been updated in the manuscript (Page 10, Line 288-292), shown below.

*All the observations from the US EPA Air Quality System (AQS*) are used to evaluate the present climate period from 2001-2004. A statistical evaluation of the pairing of the gas species outputs (CO, NO2 and O3) in time (hourly) and space (observational sites the corresponding model grids) between CMAQ outputs and AQS datasets is shown in Table 3.*

* http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm

We revised Fig. 1 and added the distributions of observational sites, shown below. The red points (~1200), the gray triangles (~450) and black squares (~450) represent the observational sites of O3, NO2 and CO, respectively.

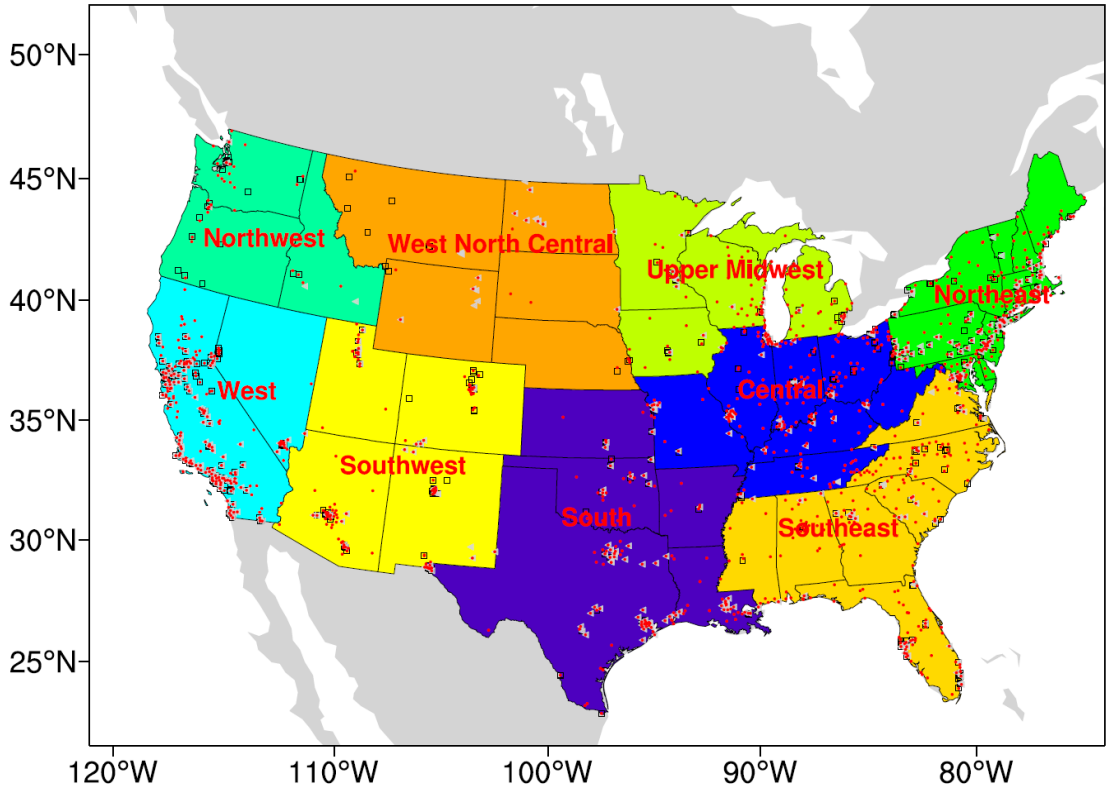


Fig. 1. 12 km by 12 km simulation domain with nine climate regions in US. The red points (~1200), the gray triangles (~450) and black squares (~450) represent the observational sites of O3, NO2 and CO, respectively.

8) What do the author's mean that a metric is the least biased (p 11325, l 19)?

Response:

We listed the formula for all the metrics in the supplement (also shown below).

Based on US EPA (2007), if the denominator only contains the observational data, higher bias could be achieved when the observation values are small. Thus, the most biased metrics could be MNB/MNE due to the observation data at the denominator, and the least biased metrics are MFB and MFE because of the summation of model output and observation data at the denominator.

*USEPA: Guidance on the Use of Models and Other Analyses for Demonstrating Attainment of Air Quality Goals for Ozone, PM2.5. and Regional Haze, EPA-454/B-07e002, 2007.*

$$MFB = \frac{2}{N}\left(\sum_{1}^{N}\frac{Model - Obs}{Model + Obs}\right)\times 100\% \qquad \text{Equation 1}$$

$$MFE = \frac{2}{N}\left(\sum_{1}^{N}\frac{|Model - Obs|}{Model + Obs}\right)\times 100\% \qquad \text{Equation 2}$$

$$NMB = \frac{\sum_{1}^{N}|Model - Obs|}{\sum_{1}^{N}Obs}\times 100\% \qquad \text{Equation 3}$$

$$NME = \frac{\sum_{1}^{N}|Model - Obs|}{\sum_{1}^{N}Obs}\times 100\% \qquad \text{Equation 4}$$

$$MNB = \frac{1}{N}\sum_{1}^{N}\frac{(Model - Obs)}{Obs}\times 100\% \qquad \text{Equation 5}$$

$$MNE = \frac{1}{N}\sum_{1}^{N}\frac{|Model - Obs|}{Obs}\times 100\% \qquad \text{Equation 6}$$

9) It is not clear how Figure 6 and 9 were created. Were they created by averaging the ozone over each region and then showing the distributions of regionally averaged ozone, or all points within a region separately into the distribution or how? The methodology changes the interpretation somewhat. Please clarify.

Response: We used all the model grids in each region without average over region. This has been added in the revised manuscript (Page 14, Line 405).

10) I'm having a difficult time understanding the author's explanation on pages 11332, line 16 through 11333, line 8. They are trying to explain why the three regions have show relatively little impact of heat waves on ozone. However, their argument has to do with the number of heat waves in these regions and/or their duration. It is perfectly possible to have a small number of heat waves but for each wave to have a large impact on ozone. Thus, the argument for the impact of heat waves on ozone should not depend on the number of heat waves.

Response:

We revised the figure based on the reviewer's major comment 3, and added statistical significance test. The sample size of heat wave days is mostly much smaller than non-heat wave days, and the percentage of heat wave days to non-heat wave days ranges from 5% to 17% under RCP 4.5, and 21% to 35%. The sample size between the entire period (including heat wave and non-heat wave period) and non-heat wave period is similar, thus we re-draw the MDA8 distributions using entire period (including heat wave and non-heat wave period) and non-heat wave period, shown below. The explanations have been revised as well, and the daily maximum temperature has been also shown in the figure below to evaluate its impact. Basically, in RCP 8.5, the daily maximum temperature shows statistically significant increase, leading to statistically higher ozone in most of US regions. The detailed descriptions (Page 16, Line 483-502) are as below:

*Under RCP 8.5, the mean MDA8 shows increases across US except Southeast, during the entire period compared with non-heat wave period, and the increase are all statistically significant, ranging from 0.3 ppbv to 2.0 ppbv. The ozone exceedance of 60 ppbv and 75 ppbv during the non-heat wave period is on average 1-8% and 0-4% lower than the entire period respectively. Under RCP 8.5, the TMX is statistically higher during the entire period than non-heat wave period, ranging from 0.8 to 2.0 °C. One of the major reasons Southeast does not show statistically significant increase in MDA8 is its position adjacent to the ocean and its small diurnal temperature variations (Fig. S2 below). Under RCP 4.5, statistically significant MDA8 increase occurs in five regions, however, the increase magnitudes (maximum of 0.7 ppbv) are much smaller than RCP 8.5. The 95% confidence interval of the MDA8 differences between entire period and non-heat wave period was also shown in Fig. 8. The upper 95% limit indicates 0.4 ppbv to 1.5 ppbv increase under RCP 4.5 and 1.2 ppbv to 3.2 ppbv increase under RCP 8.5 was resulted from heat waves without including Southeast. Even though the temperature increase in RCP 4.5 due to heat waves is statistically significant, the ozone precursors including NMVOC and NOx decreased dramatically and the methane emissions decrease by ~10% (Table 2) as well. As is explained in section 5.2, methane is the major contributor in ozone increase in RCP 8.5, and without enough ozone precursor emissions in RCP 4.5, the heat waves may not play as significant role as it does in RCP 8.5.*

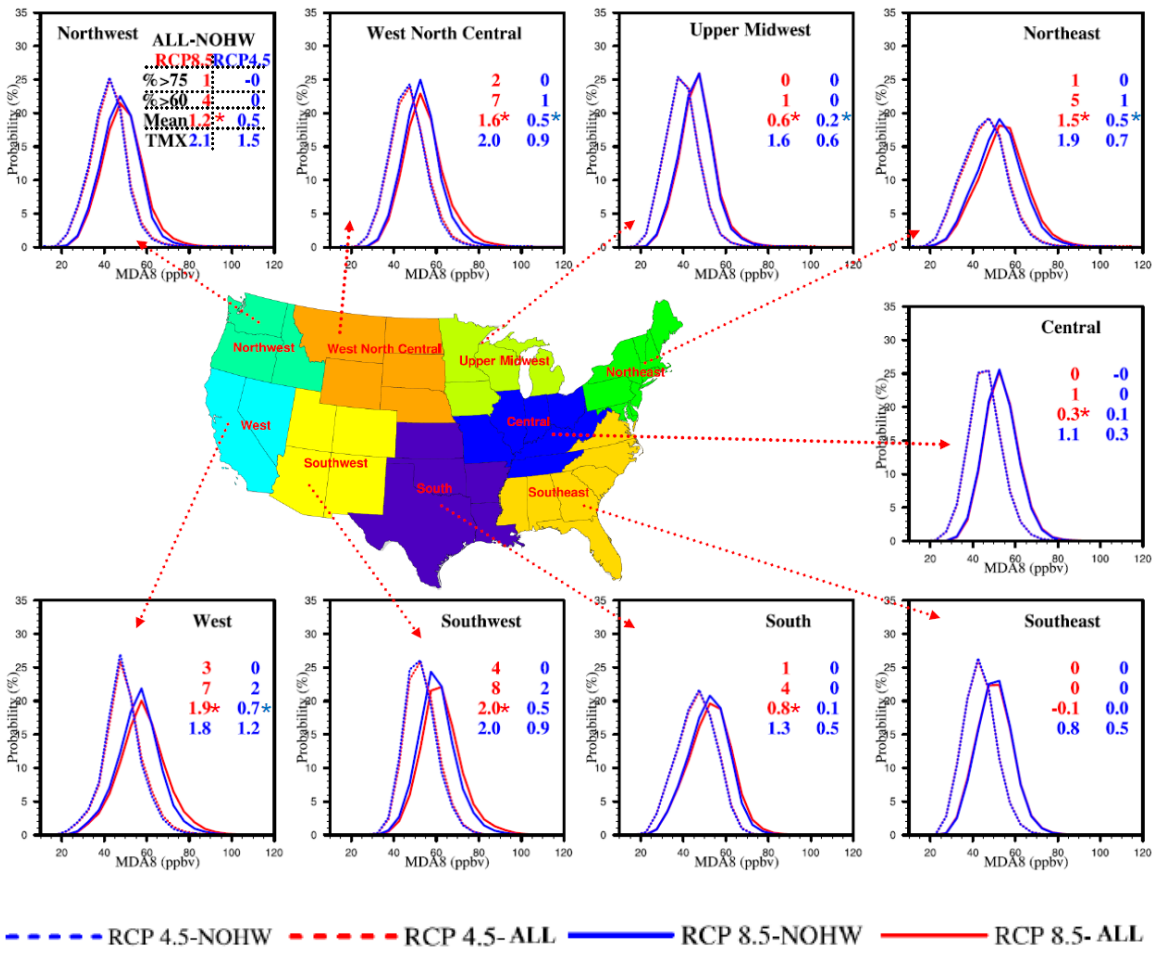


Fig. 8. Distributions of MDA8 during the entire period (referred to as ALL) and non-heat wave period (referred to as NOHW) for RCP 4.5 and RCP 8.5 from June to October. There are two columns of numbers, and they represent percentage of MDA8 ozone exceeding 70 ppbv (top row) and 60 ppbv (second row), mean MDA8 ozone (third row) and mean daily maximum temperature (TMX) for both scenarios. Statistical significance was tested and marked with star to indicate statistical significant. All TMX mean differences are statistically significant.

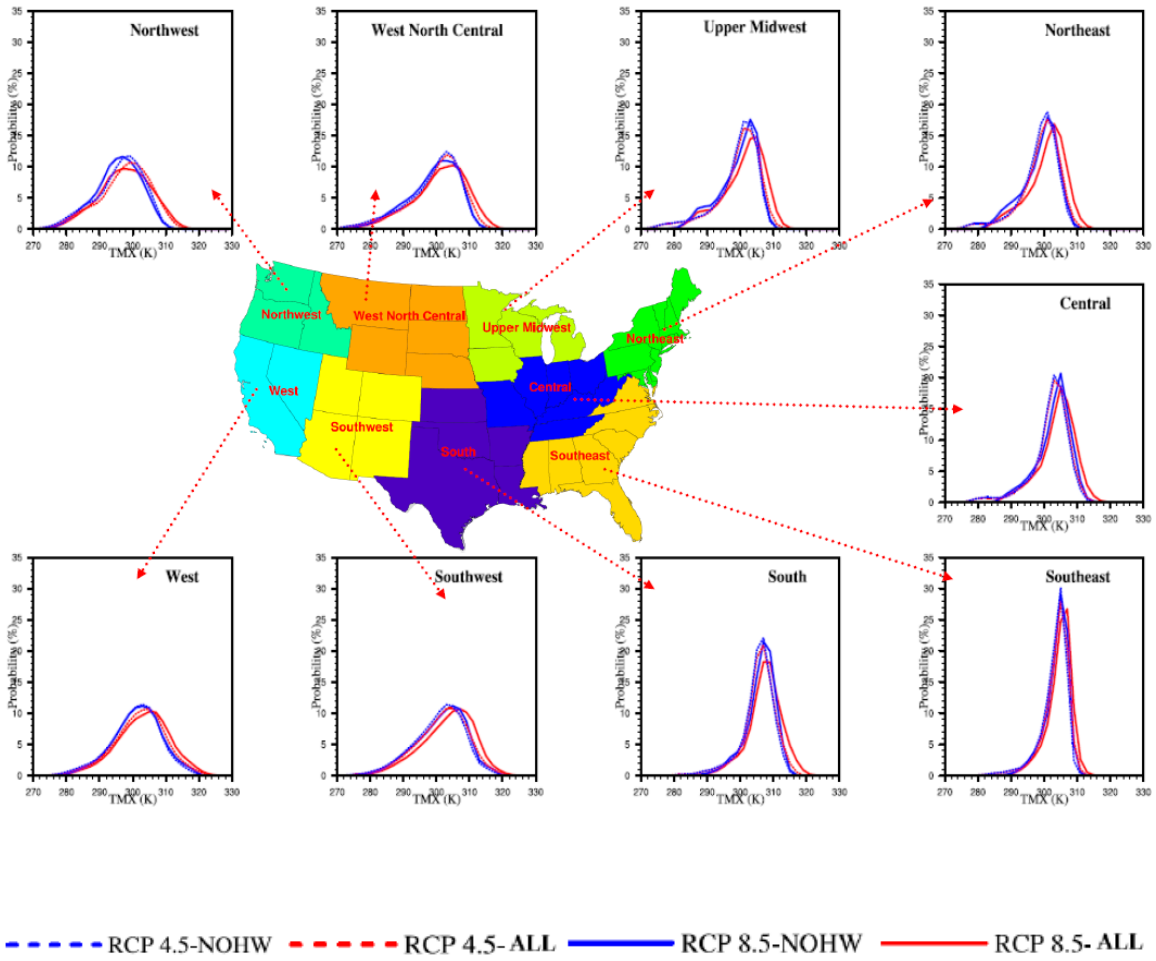


Fig. S2. Distributions of daily maximum temperature (TMX) during the entire period and non-heat wave period for RCP 4.5 and RCP 8.5 from June to October.