Comments from Review 1:

The authors conducted a study to investigate the impacts of two types of extreme weather events on the U.S. ozone air quality in the future climate. To achieve this goal, they have made use of a chemistry model and an ensemble of CMIP5 models. They found the compound events have larger effects on the ozone concentrations and these events would become more frequent by 2050s. Overall, I think the topic is suitable for ACP. But the authors should improve the presentation quality and address the questions raised by the two reviewers before the paper can be accepted.

We thank the reviewer for the constructive comments to help us further improve the manuscript. Please see the detailed responses to your comments below.

Major Comments

1. The presentation quality of this manuscript needs to be improved. The authors made a lot of descriptive statements without a reference. I have tried to list some of them but the authors should go through the manuscript and check every sentence. Many figures are too compact. Please enlarge the font size and use subtitles to more clearly convey the main information of each plot. In the caption, briefly mention the data and model that you use. Otherwise, readers who just browse the paper will get lost.

Response:
Thank you for the suggestions concerning statements without references and presentations of figures. For statements without references, we have checked the manuscripts carefully and added several references. The minor comments below listed a few missing references, and those are responded (reference added) one-by-one below. Besides these specific instances, we have also added references in line 70-72, 189-190 and 451. For clarity of the figures, we have enlarged the legend text and switched the y- and x-axis in the scatterplots of Fig. 2 and Fig. 3. In Fig. 4 and Fig. 5, the US map is not necessary as it has been described in Fig. 1 so we have removed the map to allow each panel to be bigger. We have increased the line thickness and added subtitles to show more information for each plot in Fig. 6, Fig. S2 and Fig. S3. In the figure captions, we have added information of the model and observations used.

2. The authors should show more details about the capability of WRF-Chem in simulating the ozone variability under the extreme weather. These figures should be shown. Or maybe the authors have shown some of these figures in the supplement, but I don't read clearly since they seldom mention the data or the model used for each figure in the caption.
a). A map of ozone distribution w/o the influence of extreme weather events for both observations and model simulations.
Response:
We have generated a new set of figures (shown below; Fig. S1) as suggested (shown

below and added to the supplement). As shown below, from the model results, ozone with extreme weather is higher than ozone without extreme weather in the eastern US and the west coast. The spatial pattern of ozone with extreme events is similar to the spatial distribution of compound extreme days in Fig. 2f showing more extreme days (and likely more intense extreme events) in the eastern US and the west coast. This feature is, in general, consistent with the observations (bottom row), indicating that the model well reproduced the effect of extreme events on ozone.
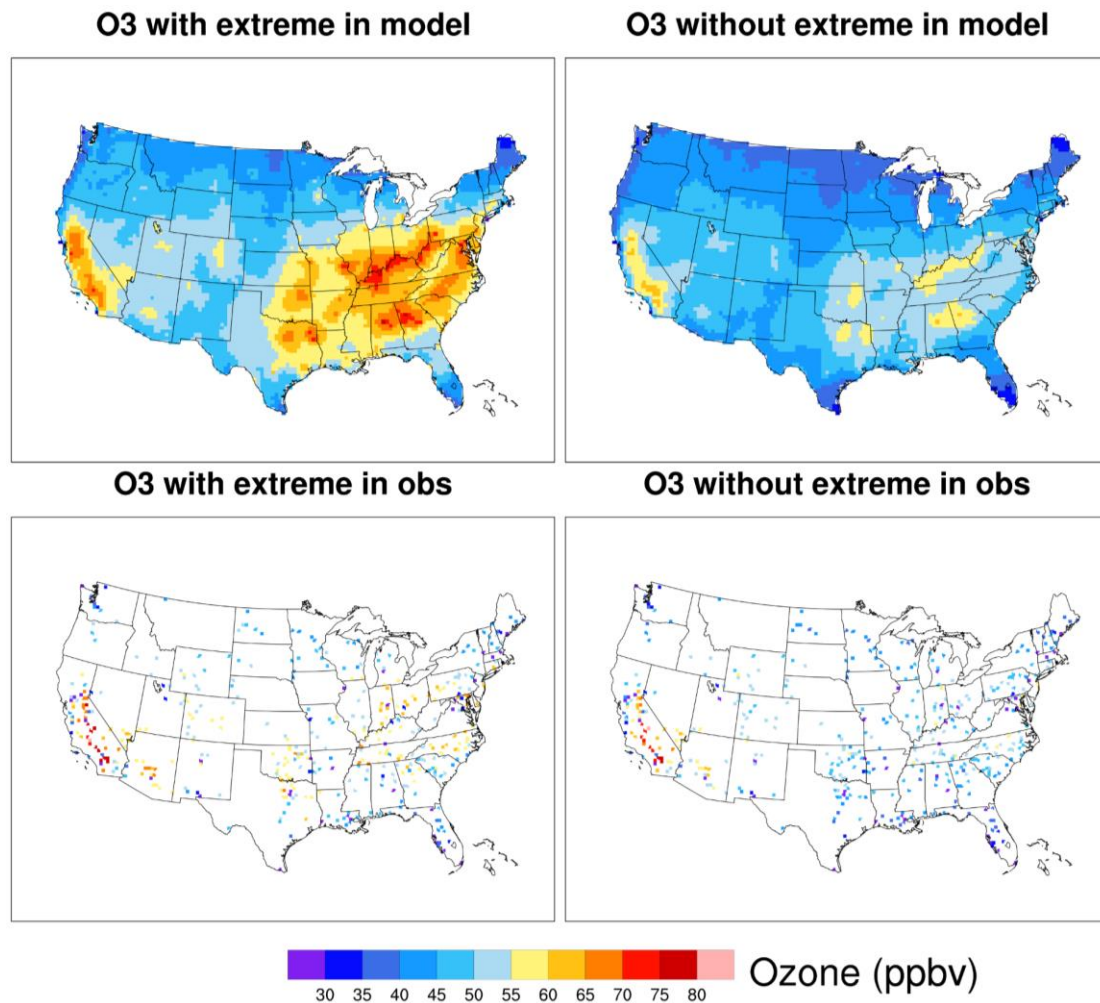
Fig. S1. Spatial distribution of summer ozone concentration with/without extreme events (heat wave, atmospheric stagnation and compound events) in model (WRF/Chem) and observations (NARR/AQS) during 2001-2010.

b). Show if WRF-Chem can simulate the interannual variability of high ozone. If the R2 is too low, maybe using WRF-Chem is not a good strategy.

Response:
To evaluate the capability of WRF/Chem in simulating the inter-annual variability of high ozone, maximum daily 8-hr ozone concentration beyond 70 ppbv is used to obtain the annual mean value of high ozone concentration in summer. For observations, only grids with data for at least five years or more are used. As shown below in Fig. S2,

interannual variability of high ozone is higher in the western US and the eastern US in the WRF/Chem simulations and observations. Correlation coefficient based on the observation stations is 0.29, with statistical significance at 95% confidence level. If we calculate the correlation coefficient only for the eastern US (east of 90° W), the correlation coefficient is 0.43, with statistical significance at 95% confidence level. These results show that WRF/Chem can reproduce the spatial distribution of interannual variability of high ozone, which is an important criterion for its use in the present study of the impacts of extreme weather events on high ozone in the U.S.
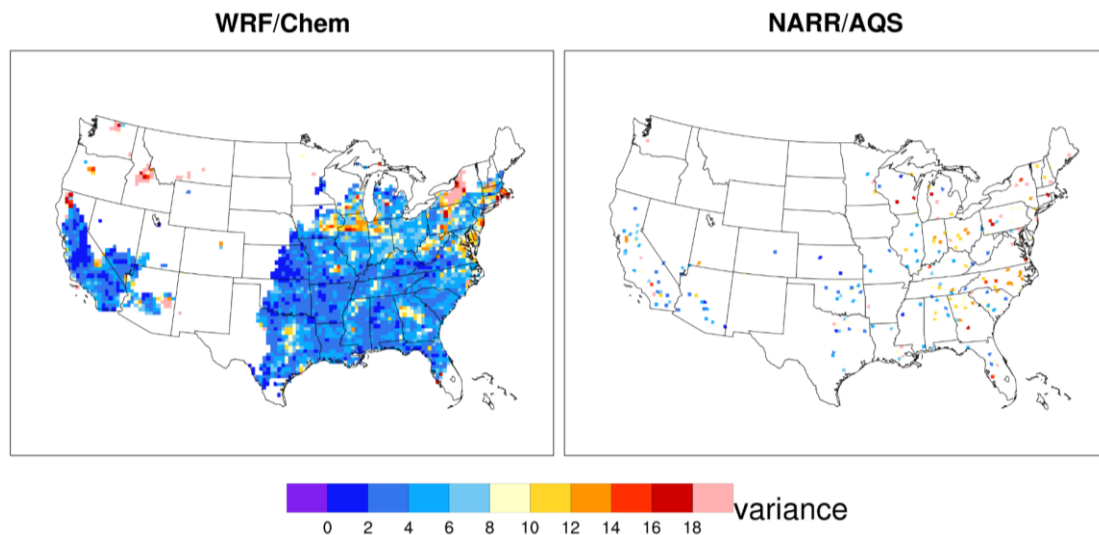
Fig. S2. Spatial distribution of variance of annual mean high ozone in each grid over the US for WRF/Chem simulations (left) and observations from NARR/AQS (right) during 2001-2010. Only grids having five years or more data are used.

c). Show a map of the fraction of ozone episodes that are driven by heatwave and stagnation in observations and the model simulations. If this fraction is too low, how will this affect the final conclusion of this study?

Response:
We have generated a new set of figures (shown below and added to supplement) for the fraction of high ozone episode driven by extreme events in summer. A high ozone episode is defined by the maximum daily 8h average (MDA8) ozone exceeding 70ppbv. We calculated all high ozone episodes in model and observation and plotted the fraction of episodes that are driven by extreme events. High fractions mainly occur in the eastern US and the west coast, which is consistent with Fig. 2 (main manuscript) as well. The spatial pattern of fraction of high ozone events that are driven by extreme events is comparable between WRF/Chem and observation. For instance, high fractions mainly occur in the eastern US in both WRF/Chem and observation, with fractions of 0.5-0.9 in most areas.
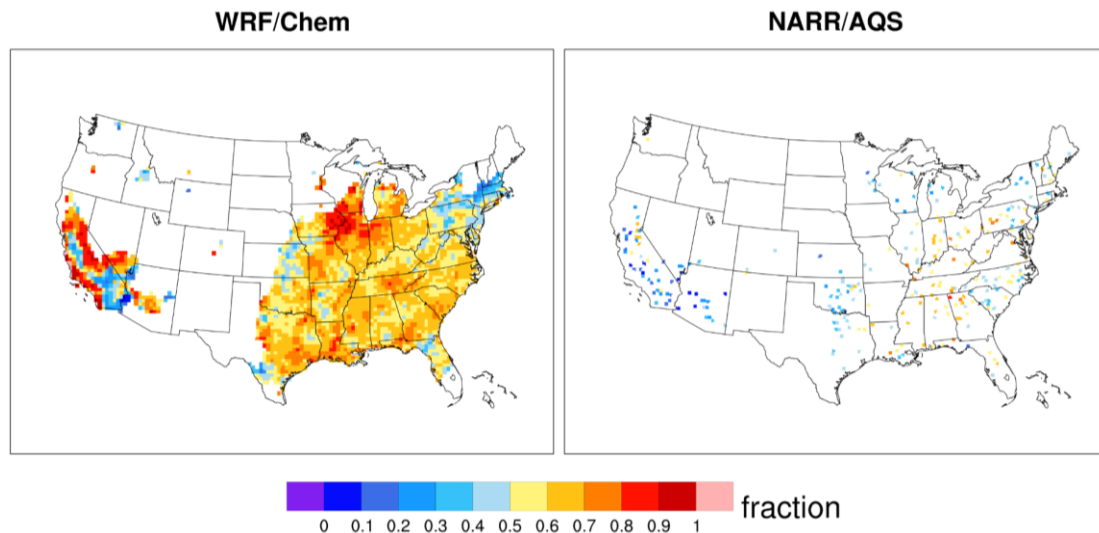
Fig. S3. Spatial distribution of the fraction of high ozone episodes that are driven by extreme events (heat wave, atmospheric stagnation and compound event) in model (left) and observation (right) during 2001-2010. Only grids having 10 days or more with high ozone are calculated. Blank areas in the model distribution correspond to areas with no or very few occurrences of high ozone episode.

3. As I read from the paper, the future meteorology used by WRF-Chem is downscaled from one single climate model. How will the meteorological changes from this single model be compared to these from the ensemble of CMIP5 models?

Response:
We have generated a new figure (shown below) for the extreme weather days in the WRF/Chem simulations and the CMIP5 multi-model mean. In this study, WRF/Chem is downscaled from the CESM-NCSU climate model. Meteorological parameters in the CESM-NCSU simulations have been compared with those of CESM in CMIP5 by Glotfelty et al. (2017), who showed consistent performance between the two models for variables including 2-meter temperature, 10-meter wind speed and precipitation. From Fig. R1, extreme weather days in the WRF/Chem simulations and CMIP5 mean show similar spatial distribution, but their intensities differ significantly. For instance, both WRF/Chem and CMIP5 mean show high heat wave days in the western US and the south central US, high stagnation days and compound event days in the western US and the central US. Generally, the CMIP5 mean shows higher heat wave days and lower stagnation days and compound event days compared to WRF/Chem for the whole US. Comparing Fig. R1 with Figure 2 of the main text, we can see that the CMIP5 mean better captures the broader areas of heat waves in the U.S. than WRF/Chem, but the heat wave days in the high centers of southwest and lower Mississippi are much lower than the observed. The CMIP5 mean also simulated much lower number of stagnation and compound extreme days compared to observations and WRF/Chem. Hence, overall, weather extremes are better simulated by WRF/Chem than the CMIP5 mean.
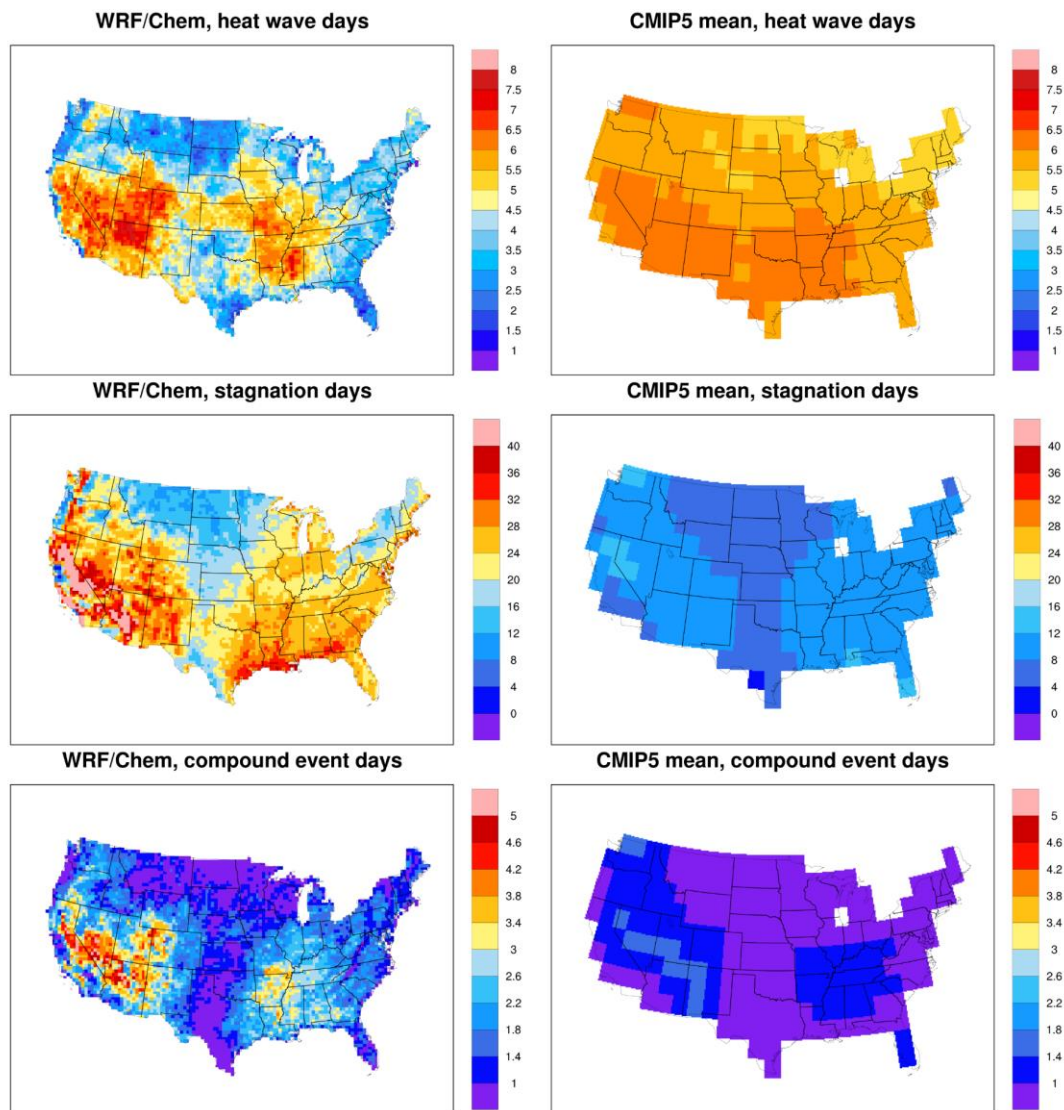
Fig. R1. Distribution of mean number of extreme weather days in summer of 2001-2010 from WRF/Chem simulations (left panels) and CMIP5 simulations (right panels) for heat wave days (top row), stagnation days (middle row) and compound event days (bottom row). Mean values of CMIP5 ensembles were used for comparison with the WRF/Chem simulations over the US.

We also displayed in Fig. R2 the changes of extreme weather event days in summer in mid-century compared to historical periods (2046-2055 minus 2001-2010) from the WRF/Chem simulations (left panels) and CMIP5 mean (right panels). The changes of extreme weather days in WRF/Chem and CMIP5 mean have similar spatial pattern, particularly for the changes in heat wave days, but the intensity differs significantly, with WRF/Chem projecting larger changes than CMIP5 mean in general (note the difference in scales used in the WRF/Chem and CMIP5 mean panels). This is not surprising given that (1) the CMIP5 mean historical extreme event days are lower than that of WRF/Chem and observations; (2) the ensemble mean changes should, in general, be smaller than those projected by individual models; and (3) model uncertainty has important contributions to the overall uncertainty in projecting climate change before

the mid-century. Hence it is important to include a section on the CMIP5 results in the paper to provide the multi-model context. We have added some discussions of the comparison of WRF/Chem and CMIP5 mean in the revised manuscript.
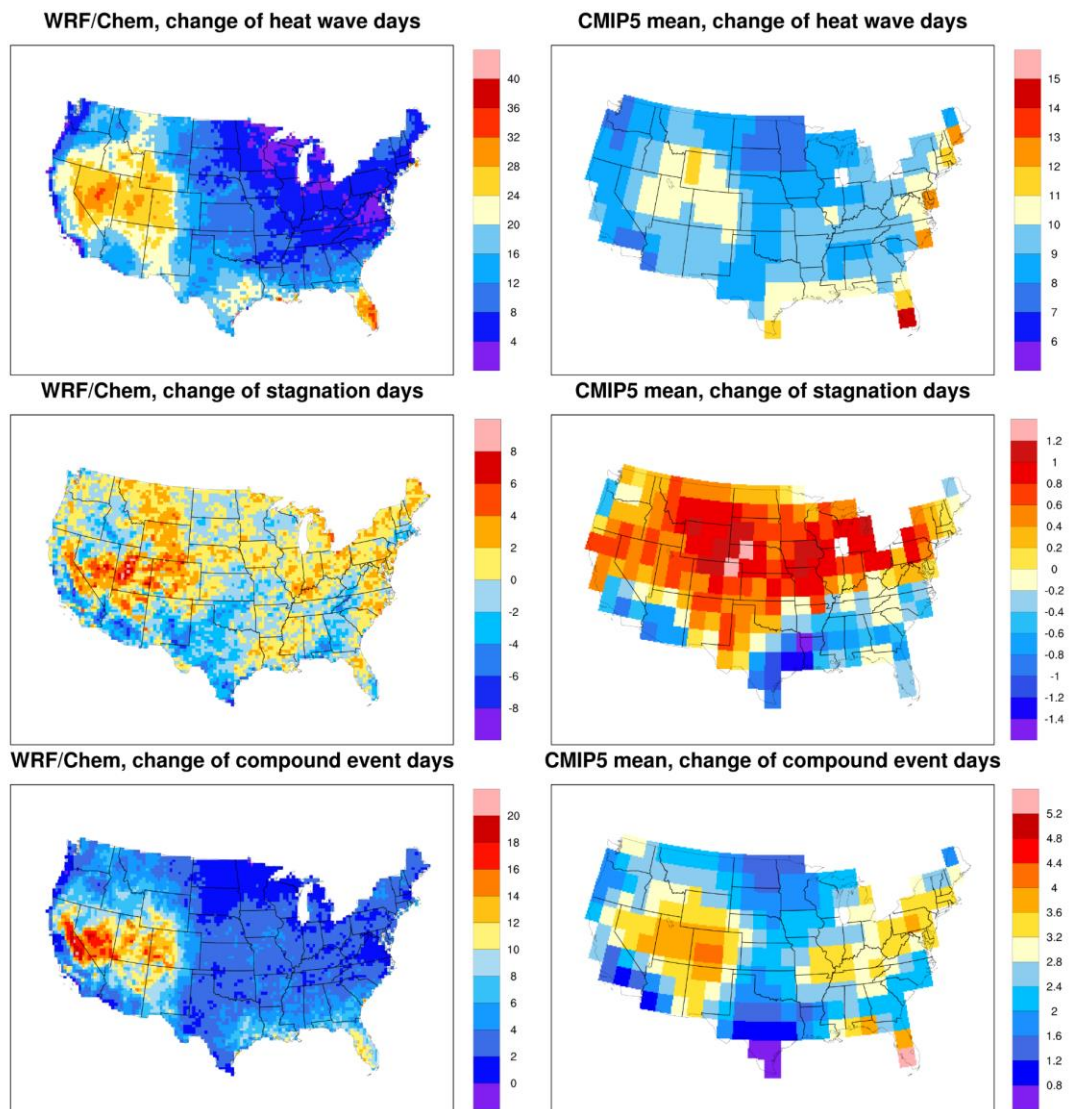
Fig. R2. Changes of mean number of extreme weather days (2046-2055 minus 2001-2010) in summer from WRF/Chem simulations (left panels) and CMIP5 simulations (right panels) for heat wave days (top row), stagnation days (middle row) and compound event days (bottom row). Mean values of CMIP5 ensembles were used and the viewport is zoomed to the US in order to compare with WRF/Chem simulations.

Minor Comments

Line 41. 'in US' should be 'in the US'

This has been revised.

Line 42. 'RCP 8.5' should be 'the RCP 8.5'

This has been revised.

Line 53-54. 'high ozone episodes are not eliminated'. You should define 'high ozone episodes' Line

We have added a definition of 'high ozone episodes' – episodes with maximum daily 8h average (MDA8) ozone concentration over 70ppbv.

Line 76-77. Missing reference.

We added two references (Mitchell, 1989; Schimel et al., 2000) after this statement.

Line 78. 'govern ozone and its changes'? Not clear.

Ozone and its changes may be confusing. we have changed it to 'govern ozone concentration'.

Line 80. Should be 'the presence of high precursor emissions'.

This has been revised.

Line 178-179. Need a reference for the definition of climate regions.

We have added a reference (Karl and Koss, 1984) for the nine climate regions in the US.

Line 241-242. Need a reference.

We have added a reference (Neelin et al., 2013) after this statement.

Line 242-244. Need a reference.

We have added a reference (Wang and Angell, 1999) after this statement.

Line 248-249. Need a reference.

We have added a reference (Zhao and Khalil, 1993) after this statement.

Line 282-283. The R shown in Figure 3 is only 0.3-0.4. Why this leads to the conclusion that WRF/Chem reasonably reproduced the observed ozone extremes?

Although the R value is not very high, we have used the rho-test (α=0.05) to test the correlation in Figure 3, and they are statistically significant. Therefore, we conclude

that WRF/Chem can reasonably reproduce the observed ozone extremes.

Line 318-321. The authors use too many quantitative words here, such as 'well captured' and 'considerable skill'. I am not very convinced that WRF/Chem can well simulate the ozone extremes in the US without further evidence.

Response: We have revised the words we used to discuss the model skill, since there isn't an absolute criterion to judge whether certain features are "well captured" or whether the model skill is "considerable". We now say "reasonably captured" instead of "well captured" and "skillful" instead of "considerable skill". We note, however, that our discussion is not limited to qualitative, as we discuss specific biases quantitatively. We also hope that by addressing your comments using Figures R1 – R3 above, we have provided further evidence that the simulations of ozone by WRF/Chem are comparable to observations to warrant their use in analysis of the impacts of extreme events on high ozone events.

Line 396-398. Is this supported by a reference?

We have added a reference (Vingarzan, 2004) after this statement.

Line 416- 418. Is this simulated by a chemistry model?

Yes, it is simulated by WRF/Chem, WRF model coupled with chemistry.

Figure 1. Use subtitles that can more clearly convey the main information. For example, "(a) NARR" can be changed to '(a) NARR, heatwave'.

This has been revised.

Figure 3. Switch the x and y axis.

This has been revised.