Reviewer 2

The revision have improved the significance of the paper and I have only some remaining comments/questions before I see the paper as ready for publication.

Author Response: We thank the anonymous reviewer again for insightful and critical feedback on our manuscript. We originally envisioned this work being included in two papers: 1 demonstrating the model's ability to predict O_3 in the UGRB and 1 focused on VOC and NO_X sensitivity. Based on the comments and suggestions, we have conducted the sensitivity analysis and included it in this study. We have fully explored both VOCs and NO_X in this revision based on the comments and suggestions provided. Further, we have carefully reviewed the entire paper to 1) reduced unneeded repetition, 2) correct grammatical errors), 3) improve the flow of the text, and 4) address other minor text errors. We truly believe that this work will have a much larger impact on the community after the suggested additions/revisions.

1. **Reviewer Comment:** Why were the boundary conditions only updated every 24 hours if the global model output is available every 6 hours?

   Author Response: We apologize for the confusion. This was an error in the original manuscript. In fact, the boundary conditions were updated every 6 hours, and this has been revised accordingly on line 188: "*To account for the transport of chemical species into the model domain, 6-hourly data from the Community Atmosphere Model with Chemistry (CAM-CHEM; Emmons et al. (2020)) were used in the simulations.*"

2. **Reviewer Comment:** My question about the spin-up period was not really answered and at least a 1-2 day spinup should be considered in any study. At the least it should be mentioned that the first days might be affected by spin-up effects and that results for these days needs to be taken with caution.

   Author Response: We apologize for not fully addressing your original comment. In practice, 1-2 days of spinup is ideal. However, given the computational expense of the simulations and the growing number of sensitivity simulations, additional simulation days preceding the study period for spin-up were not considered. However, O_3 does not start ramping up until nearly 24 hours into the simulation, and the results of the peaks in O_3 are generally qualitatively similar between days providing some confidence that a longer spin-up period would not have changed the key outcomes of this work. To address this in the manuscript, a statement regarding the spin-up period has been added on line 197 under Section 2.4: "*A spin-up period was not explicitly considered in this study owing to the computational expense of each simulation and that O_3 generally does not start increasing until nearly 24 hours into the simulation; however, the results from the first day should still be viewed with caution.*"

3. **Reviewer Comment:** Evaluation of winds: I would not necessarily say that the model does a great job at simulating WS, specifically for the last day. The authors also did not provide information on wind direction. I suggest rephrasing "good qualitative agreement"

to something like "reasonable qualitative agreement" or similar. Also please comment on the evaluation for wind direction.

Author Response: Thank you for the comment. The comment on the last day is valid; however, on this day, O_3 levels are quite low and below the 70 ppb threshold commonly used to indicate a high O_3 event. The key is that the wind speeds are well simulated on the days of interest, i.e., the days with elevated O_3 contents. We feel that mentioning the small negative wind speed bias on the final day would be a distraction for the reader based on this.

We have changed the wording from "good qualitative agreement" to "*reasonable qualitative agreement*" in line 360.

Regarding the wind direction, the reviewer is correct that we did not provide information on validating the wind direction in the model simulations. The reason is the same as focusing the wind speed analysis on metrics such as the occurrence of wind speeds below a certain threshold over a direct correlation analysis. The key wind speed criterion is weak winds. We have shown in the paper that the model exhibits satisfactory performance with regard to low wind speeds. As such, we are far less concerned about the direction because if the wind speed is weak, the biases in the wind direction will have minimal impact on chemical transport. Moreover, at low wind speeds, wind direction observations can vary considerably and are less reliable. For these reasons, we feel that it is best to exclude a discussion of modeled versus observed wind directions in the manuscript. Future work is currently under way focused on detecting precursor leaks from oil and natural gas wells, and in this work we are focused on wind speed and detection because the work is intended to be generally applicable to all weather conditions, not just those conducive to high O_3 levels.

4. **Reviewer Comment:** Line 254: The photolysis option 4 should not be used because it gives higher ozone for this case study but because it is an updated parameterization based on recent advanced in the understanding of photolysis rates.

   Author Response: This is a very good point, and our original wording certainly did not make this clear. The text in the manuscript has been updated on line 222 to read as follows: "*Where possible, the same namelist options were used for both models. However, regarding the photolysis option, the simulations with MOZART used photolysis option 4, which is the updated TUV photolysis option based on recent advanced in the understanding of photolysis rates that was configured to work with only a few chemistry mechanism schemes in WRF-Chem v3.9.1, while the RACM simulations used photolysis option 1, which is the Madronich photolysis scheme.*"

5. **Reviewer Comment:** Line 300ff: Models struggle getting the boundary layer correct and specifically so for early morning. This will impact the model evaluation. At this time of the day, I would assume the measurements to be strongly impacted by local effects because of limited mixing which I would not expect the model to represent due to coarser spatial

resolution. So if there is a nearby source at the Boulder station, the measured signals might be dominated by it rather than representative of the regional characteristics.

Author Response: The reviewer has a good point. We utilized the mobile lab data to confirm that the methane concentration at the site was similar to the methane in the larger grid. While methane is not a perfect proxy for VOCs, it generally tracks the smaller alkanes (ethane to pentane) fairly closely and the smaller alkanes represent the majority of the VOC mixing ratio. This has been explained more clearly on line 439 that reads as follows: "*The WYDEQ Boulder data were within 25% of the data collected by the mobile lab near the monitoring site (Fig. D1 in Appendix D). This observation indicates that the difference between the simulated and observed NMHC mixing ratios is not the result of anomalously high mixing ratios at the Boulder site, and thus the NMHC mixing ratio measured at the Boulder site is an accurate representation in the region.*"

6. **Reviewer Comment:** Line 305: I am puzzled by the use of the word "emission factor". An emission factor typically represents the amount of a compound released, e.g. grams of a compound per biomass burned. What is shown here is simply the ratio of the observed and modeled value. I recommend using a different term, e.g. scaling factor or similar. I also caution that you are using ambient concentrations to correct the emissions without knowledge about the age of the airmasses and how representative they indeed are of fresh emissions.

   Author Response: Yes, this is a very good point, and sorry for the confusion. All instances of "emission factor" have been revised to "emission adjustment factor" in the revised paper.

7. **Reviewer Comment:** Line 314: It needs to be explained here why the dry deposition was turned off (or at least referred to the later discussion on this)

   Author Response: The use of dry deposition for the sensitivity studies was not mentioned until later in Section 4, where the results from the sensitivity simulations are analyzed and discussed. To address this, on line 392, the text has been reworded to read as follows: "*RACM does not adjust the dry deposition rate over snow-covered surfaces, hence the dry deposition is likely too high in RACM and could explain the underestimate of O_3. Thus, we turned off dry deposition to mimic the very slow deposition of gas-phase species over a snow-covered surfaces (i.e., RACM_ddOff).*"

8. **Reviewer Comment:** NOx Sensitivity Simulations: As stated, I am worried that the VOC measurements at the Boulder site might be significantly impacted by local effects and by adjusting the VOCs in the model for the entire area based on these values might not be truly representative of the region. Have the authors tested the NO_X sensitivity also with the base emissions? At the very least they should comment on the changes in the $HCHO/NO_2$ ratios and potentially chemical regimes that arise from an increase in the VOC emissions.

   Author Response: The reviewer brings up an important point about utilizing data from a single station to adjust VOC emissions. We have added Table B1, which shows the

observed-modeled VOC ratios for the other observations sites in the basin where canister data are collected. Further, in line 275 the following text has been added to the manuscript: "*Table B1 in Appendix B shows ratios of canister-observed speciated VOC mixing ratios to the simulated values. The Moxa Arch site is relatively far from the emission sources and is not representative of the main O_3 formation region. The other sites show variability, although all sites show significant underestimates of both VOCs and NO_X. The Boulder site has the largest underestimates of reactive BTEX species, while some species have larger underestimates at other sites. Given this comparison and the site-to-site variability in the model–observation comparison, VOCs measured at the Boulder site appear to be a reasonable, though aggressive, basis for adjusting the emissions in the model.*"

We also thank the reviewer suggesting that we focus also on NO_X sensitivity alone. We have conducted two more simulations for each chemical mechanisms, which are described on line 501: "*To explore the sensitivity to NO_X vs. VOC further, additional simulations were conducted where only NO_X and only VOCs were adjusted from the baseline run (Fig. 16). The simulation where only NO_X was adjusted results in moderate increases in the simulated O_3 mixing ratios at all sites and across the entire study period. When NO and NO_2 emissions were kept at their baseline levels and VOC emissions were adjusted, the simulations produce slight increases in O_3 at some sites, especially with the MOZART chemistry scheme, but at other sites the O_3 mixing ratios are not always elevated. Rather, in these cases, the timing of O_3 formation changes, with increases seen earlier in the day and sometimes even lower peaks in the modeled O_3 mixing ratio compared with the baseline run. These results are interesting and somewhat unusual, but further analysis was not pursued because by increasing VOCs while not adjusting NO_X, the VOC:NO_X ratio for this model run is far outside what is actually observed in the basin and is thus considered highly unrealistic.*"

Further, we have also looked at the changes in the HCHO:NO_2 ratio. On line 512, the following discussion has been added: "*To further investigate this, the formaldehyde:NO_2 (HCHO:NO_2) ratio for all VOC and NO_X sensitivity simulations is presented in Fig.* **??**. *This ratio has been used in previous studies as a proxy for VOC-limited and NO_X-limited conditions (Liu et al., 2021). Here, the ratio is well above 1 during the high O_3 events for all simulations, with the only decrease being observed for the simulations with only increased NO_X emissions. These results further suggest that O_3 formation in the basin is strongly controlled by NO_X availability (Liu et al., 2021)*"

9. **Reviewer Comment:** The wording for some of the added parts would need to be proof-read and improved"

   Author Response: We have carefully reviewed the entire manuscript to revise and improve the wording. Several paragraphs have been removed to eliminate redundancy, and a lot of the text has been revised to ensure grammatical correctness and conciseness.

References

Emmons, L. K., Schwantes, R. H., Orlando, J. J., Tyndall, G., Kinnison, D., Lamarque, J.-F., Marsh, D., Mills, M. J., Tilmes, S., Bardeen, C., Buchholz, R. R., Conley, A., Gettelman, A., Garcia, R., Simpson, I., Blake, D. R., Meinardi, S., and Pétron, G.: The Chemistry Mechanism in the Community Earth System Model Version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, e2019MS001 882, https://doi.org/10.1029/2019MS001882, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001882, e2019MS001882 2019MS001882, 2020.

Liu, J., Li, X., Tan, Z., Wang, W., Yang, Y., Zhu, Y., Yang, S., Song, M., Chen, S., Wang, H., et al.: Assessing the Ratios of Formaldehyde and Glyoxal to NO2 as Indicators of O3–NO x–VOC Sensitivity, Environmental Science & Technology, 55, 10 935–10 945, 2021.