

## Response to Reviewer 1

We would like to thank the reviewer for their detailed edits and comments that have improved this manuscript. Below you will find our response to each comment/question. The reviewer's comments/questions are in bold italic font.

***In the Abstract the authors state, “aqueous chemistry exacerbated the discrepancies of HCOOH leading to a net depletion”. Which discrepancies – between models or in relation to measurements at WFM?***

This sentence was edited to clarify which discrepancies we were referring to. This sentence was changed to:

“The addition of aqueous chemistry exacerbated the discrepancies of HCOOH with observations by leading to a net depletion within cloud water.”

***The cloud samples are collected in bulk at 12 hour intervals. Can nighttime vs. daytime differences in the cloud samples be teased out to understand potential accuracy of model differences in their prediction of the predominance of glyoxal production pathway (i.e., nighttime from the ozonolysis of an isoprene hydroperoxy aldehyde vs. daytime oxidation of the lumped peroxy radical XO<sub>2</sub> reaction with NO)? Perhaps the NO<sub>x</sub> emission timing and concentrations are too uncertain.***

This is an interesting question certainly worthy of investigation. However, most of the chemical production of glyoxal occurs upwind of WFM (see Figure S23) before subsequently being controlled by entrainment of background air. Therefore, differences in nighttime vs daytime oxalic acid concentrations measured at WFM are not necessarily indicative of associated nighttime vs daytime chemistry impacting oxalic acid concentrations. Due to the reason stated above, we do not believe that it is possible to test nighttime vs daytime model disagreement.

***The authors assume gas-to-droplet partitioning and do not consider aqueous phase production and then find that the gas-phase predictions are an order of magnitude lower than WFM cloud water measurements. Is there any understanding from Barth et al., (2021) cited in the manuscript as to which process produces more of the carboxylic acids in cloud water? Is it gas-to-droplet partitioning or in-cloud production?***

To evaluate whether gas-phase only chemistry can produce the measured organic acid concentrations, gas-to-droplet partitioning was applied to the measured aqueous-phase concentrations to estimate the gas-phase mixing ratios for HCOOH, CH<sub>3</sub>COOH, and oxalic acid. In section 4.3, we then considered gas + aqueous-phase chemistry in the Barth cloud chemistry model that represents both gas-to-droplet partitioning and chemistry. The results from these runs

show that gas-to-droplet partitioning was 100x and 10,000x greater for HCOOH and CH<sub>3</sub>COOH, respectively, than their aqueous-phase chemical reaction sources. Oxalic acid has no gas phase production chemistry, so any production will only have occurred in the aqueous phase. In this current study, cloud water chemistry depletes HCOOH concentrations while the Barth model in the Barth et al. (2021) paper, cloud water chemistry is the major source of HCOOH in cloud droplets, evidenced by strong increases in HCOOH in the aqueous phase while there were no increases in HCOOH during the gas phase only simulations. These differences highlight that cloud water can be either a net sink or net source of HCOOH in different scenarios. A few sentences were included that summarizes the above discussion which read:

“These model results imply that gas-to-droplet partitioning is the major source of HCOOH and CH<sub>3</sub>COOH in cloud water rather than chemical production within cloud droplets. This is confirmed by comparing the rate of gas-to-droplet partitioning to aqueous production, which is 100x and 10,000x greater for HCOOH and CH<sub>3</sub>COOH respectively.”

And:

“The differences in model results on different dates imply that cloud water chemistry can either be a net source or net sink of HCOOH depending on the given scenario.”

***In this paper, the description of model performance is generally subjective and colloquial. Phrases such as “strong agreement”, “fairly substantial disagreement” etc. are common. It would be better to provide quantitative assessment. For example, starting on Line 154, the authors state, “There is reasonable agreement of surface O<sub>3</sub> and PM<sub>2.5</sub> ...” Typically models are evaluated with a quantitative description of statistics such as normalized mean bias. Can the authors provide this assessment for PM<sub>2.5</sub> and O<sub>3</sub> in this simulation and provide context for model performance? I find it is difficult to tell in some areas (e.g., near Lake Michigan) what “reasonable agreement” is. Also, how is the bias/error in the general vicinity of the HYSPLIT trajectory ‘upstream’ of WFM, specifically?***

The WRF-Chem Evaluation section has been updated to be more quantitative in reporting model performance. Pearson correlation coefficients and Mean Bias Error (MBE) maps have been added to the supplemental material (Figures S7 and S8), and these same statistics have been added to Figure S9. Additionally, the text was updated to include these model performance statistics and how the model performs along the HYSPLIT trajectories. The section was updated to read as follows:

“Modeled O<sub>3</sub> exhibited a strong positive linear correlation ( $r > 0.8$ ) with observations across the model domain, but consistently exhibited a mean bias error (MBE) of 10+ ppbv on June 29th and July 1st (Figures S7 and S8). This high bias in O<sub>3</sub> production has been reported in other recent works (Travis et al., 2016; Schwantes et al., 2020; Place et al., 2023) which may be due to overestimated NO<sub>x</sub> emissions and/or improper representation of gas-phase organic chemistry.

Note that the 2017 EPA NEI used in this study is appropriate for a typical summer day and will likely not represent the actual emissions of the heatwave period caused by the stagnation event. Heatwaves can increase demand on the grid (Maia-Silva et al., 2020; Stone et al., 2023) and therefore increase NO<sub>x</sub> emissions due to greater combustion of fossil fuels from power generation (Chen et al., 2015), which are not represented by the 2017 NEI. Given the potential low bias in modeled NO<sub>x</sub> emissions, the high bias in modeled O<sub>3</sub> is even more perplexing, highlighting the complex chemistry involved in O<sub>3</sub> production. Importantly, the modeled MBE for O<sub>3</sub> is < 10 ppbv for central Missouri on June 29th, and Western New York on July 1st, locations that were upwind of WFM according to the HYSPLIT trajectories. This indicates that O<sub>3</sub> chemistry was well represented in the air mass that traveled to WFM.

PM<sub>2.5</sub> model predictions performed worse compared to O<sub>3</sub> with many linear correlation values exhibiting null or negative values and MBE exceeding 10 µg m<sup>-3</sup>. Similar to O<sub>3</sub>, model MBE was < 10 µg m<sup>-3</sup> for Missouri and much of Chicago on June 29th and Western New York on July 1st. Three air quality monitoring sites in New York, measuring O<sub>3</sub>, PM<sub>2.5</sub>, and 2 meter temperature, were chosen for time-series evaluations of WRF-Chem, including Pinnacle State Park (PSP) in the Southern Tier of New York, Queens College in New York City, and measurements at the old ski lodge below the summit of WFM (Figure S9). More information about the data collected at these sites can be found in Brandt et al. (2016) and Ninneman et al. (2020), while Pearson correlation values and MBE statistics can be found in Figure S9. WFM tends to show the lowest linear correlation with observations. This is likely due to WRF-Chem underestimating the elevation of WFM (1483m) by over 700m, and therefore not properly accounting for the topography in the region (Figure S10). PSP shows the lowest MBE values with high correlation coefficients ( $r > 0.7$ ) for O<sub>3</sub> and 2m temperature. Finally, Queens college saw the strongest correlation coefficients for O<sub>3</sub> and 2m temperature ( $r > 0.85$ ), but exhibited large positive biases for O<sub>3</sub> and PM<sub>2.5</sub>. The causes behind these overpredictions remain unclear but are beyond the scope of this work.”

## **Detailed comments:**

***Line 51: I think these are all modeling studies with the exception of Blando and Turpin, 2000 which is a literature review of plausibility. It would be good to cite as evidence a glyoxal–OxAc-cloud SOA reference that is experimental (lab or field).***

Observational/experimental studies were added to the list of citations including Sorooshian et al 2006, Carlton et al 2007, and Tan et al 2010.

***Line 94: “a manuscript regarding organic acid measurements is forthcoming” I find it difficult to assess some of the quantitative description in the absence of constraints on measurement uncertainty.***

Additional information was added to the methods sections that includes detection limits, which reads:

“While the exact detection limits of the organic acid analysis is currently being determined, a conservative estimate of 50 ug L<sup>-1</sup> for all 3 organic acids is used, based on the lowest concentration calibration standard. It is worth noting that the concentrations of the 3 organic acids investigated in this study are well above this conservative detection limit, with concentrations of 113, 111, and 23 times greater than the lowest concentration standard used in the calibrations respectively.”

***Line 196: are these gas-phase or aqueous-phase photolysis rates? How were the different?***

Line 196 was updated to clarify gas phase photolysis rates. The O<sub>3</sub> and H<sub>2</sub>O<sub>2</sub> aqueous phase photolysis rates are 1.5x greater than their gas-phase counterparts to represent the increased pathlength in the cloud droplets.

***The authors focus on a high-pressure stagnation event and seek to make associations with NO<sub>x</sub>. Electricity sector emissions increase during meteorological events like one studied here. Such events are often referred to high electricity demand events and there is increased reliance on peak load units. This may help the authors make their connections to NO<sub>x</sub>.***

This is a very interesting point brought up by the reviewer. Heat waves are well known to increase NO<sub>x</sub> emissions from the power sector. Additionally, the 2017 National Emissions Inventory is designed to represent a typical summer day and thus does not properly represent emissions from the power sector due to elevated temperatures. However, WRF-Chem overestimates O<sub>3</sub> mixing ratios in most regions of the domain during the pollution event, revealing a complex chemical system of potentially underestimated NO<sub>x</sub> emissions from the power sector occurring at the same time as high bias of modeled O<sub>3</sub> mixing ratios. We have added discussion of this phenomena to the WRF-Chem evaluation in section 3.2.2. The section reads:

“Modeled O<sub>3</sub> exhibited a strong positive linear correlation ( $r > 0.8$ ) with observations, but consistently exhibited a mean bias error (MBE) of 10+ ppbv on June 29th and July 1st (Figures S7 and S8). This high bias has been reported in other recent works (Travis et al., 2016; Schwantes et al., 2020; Place et al., 2023) which may be due to overestimated NO<sub>x</sub> emissions and/or improper representation of gas-phase organic chemistry. Note that the 2017 EPA NEI used in this study is appropriate for a typical summer day and will likely not represent the actual emissions of the heatwave period caused by the stagnation event. Heatwaves can increase electrical demand (Maia-Silva et al., 2020; Stone et al., 2023) and therefore increase NO<sub>x</sub> emissions due to greater combustion of fossil fuels from the power generation (Chen et al., 2015), which are not represented by the 2017 NEI. At the same time, there is a high bias of modeled O<sub>3</sub> mixing ratios, highlighting the complex chemistry involved in O<sub>3</sub> production.”

***Incorporating information used to Figure 8 into Figure 10 that directly compared concentrations from the observation-based estimates to the model predictions would be helpful.***

We are a bit unsure what the author is referring to with this comment. The blue and red dotted lines represent the observations made at Whiteface Mountain, allowing for direct comparison with model results. Additionally, information plotted in Figure 8 was used to initialize the gas+aqueous model that produced Figure 10. The first two sentences of Section 4.3 were edited to help clarify this:

“Cloud chemistry can alter organic acid concentrations distinct from gas phase chemistry alone. This section examines the impacts of aqueous chemistry by investigating both total mixing ratios and aqueous concentrations of HCOOH and CH<sub>3</sub>COOH using mixing ratios near WFM to initialize the model (Figure 10).”

***Line 390: I think it is more precise to start this sentence with “Model predictions suggest strong isoprene emissions ...”***

Line 390 was updated to be more precise as the Reviewer suggested.

## **Editorial:**

***Line 155: “PM 2.5” has a space and the number is not subscript***

This line was updated to include a subscript.

***Starting at line 172: “The complex geography of the Adirondack Mountains are likely not properly captured with a 12kmx12 km horizontal grid resolution ...” “geography” is singular and “are” is plural. Also, I am sure there must be a paper describing the complex micrometeorology of WFM. Why not describe is more precisely?***

The word “are” was changed to “is”. Unfortunately, there is very little research that has looked at the micrometeorology at WFM, but in general mesoscale flow includes mountain-valley winds and the mountain summits located above the PBL during night and morning and within the PBL during afternoons. In addition, when using a 12 km x 12 km horizontal grid mesh in WRF-Chem, the topography is not well represented. WRF-Chem underestimates the elevation of WFM by ~700m. Therefore, the mesoscale motions in WRF-Chem may not include upslope and downslope winds between 700 and 1500 m affecting the ability to compare properly between the model and observations at the summit. The authors have added a new figure to the supplemental material (Figure S10). Additionally, a sentence was added to the main text that reads:

“This is likely due to WRF-Chem underestimating the elevation of WFM (1483m) by over 700m, and therefore not properly accounting for the topography in the region (Figure S10). By using a 12 km x 12 km horizontal grid mesh in WRF-Chem, the topography is not well represented resulting in the modeled WFM summit to be underestimated by ~700 m and affecting the capabilities of WRF-Chem to represent mountain-valley winds and timing of when the summit is above and within the PBL (Giovannini et al 2020).”

***Many of the References are formatted incorrectly: Herckes et al., seems to have a personal note. Sometimes Atmos. Phys. Chem. references have “publisher: Copernicus GmbH”, sometimes they do not.***

The citations in the reference section have been corrected. This seemed to be caused by strange errors within the first author’s citation manager.