Please see below our point-by-point response (in blue) to reviewers' comments (in black). Quoted text from the revised manuscript is *in italic*.

Response to RC1

General Comments

Overall, the paper is interesting and uses an updated version of WRF-Chem with the Noah-MP LSM (i.e., the NASA-Unified Weather Research and Forecasting model with online chemistry). The paper covers relatively important aspects of the complex interactions between changes in reactive nitrogen (Nr) fluxes and ozone formation in the eastern U.S., and the relationships and implications of such changes with changing land/biosphere-atmosphere interactions. The paper highlights such interactions that are relatively less appreciated in the community in regards to land surface data/processes, Nr fluxes, and ozone. Indeed, the paper attempts to cover a lot of complex land-atmosphere-chemistry topics. However, many arguments and discussions, appear only cursory with minimal to no supporting scientific evidence/analyses or quantitative assessments provided for justification. I provide explicit comments below that pertain to some of these issues in the paper, and suggestions to improve them. Also, the manuscript also appears flawed in its grammar and sentence structure, where a thorough proof-read of the English writing should have been done prior to submission. To return the review comments in a timely manner, I can only provide the detailed grammatical and sentence structure errors in the Abstract and Introduction sections, with general issues in other sections that need to be thoroughly revised (see Technical Corrections below). The grammatical errors are persistent through the remaining manuscript, and it is imperative that the manuscript is thoroughly checked for grammar and English writing in all sections before submission of the revised version. Otherwise, in my opinion, the paper is not acceptable for publication.
Thanks for the overall positive comment on the paper content. Multiple authors who are native English speakers edited previous versions of this manuscript. More information supporting our arguments has been added. Careful proofreading of the revised manuscript has been conducted.

Specific Comments

2 Methods

Line 103: Model spinup description of "late April" is not specific enough. What was the exact number of days used for spinup in April?
"*Late April*" has been changed to "*25 April*".

Lines 144-145. OK, but seems that omission of detailed fertilizer inputs and bi-directional ammonia fluxes (BIDI-NH3) in the modeling system is an oversight considering all other advances. The BIDI-NH3 approach has been shown to have significant impacts on overall deposition of Nr species and aerosols. Can this in any way be rectified, or at least discussed in terms of model uncertainty using only a unidirectional approach in areas of significant ammonia fluxes (e.g., agricultural lands)? This becomes an issue later in the results section of the paper (i.e., Lines 351-353).

The decision of not applying a BIDI-NH_3 approach in this case was carefully made. We are aware of the potential uncertainty introduced by BIDI approaches which can result from the assumptions in calculating stomatal and ground compensation points, and good information for adequately addressing that uncertainty is lacking, especially at grid scale. The cited review papers summarize numerous published methods and datasets related to emission potential (for calculating compensation points) and compensation points – the data are overall highly variable and extremely limited for the US regions, making it difficult for them to be broadly applied to the US ecosystems we study here. Based on a few previous studies (previous L351-353), the uncertainty of not applying a BIDI approach was estimated to be a few percent over source regions. This estimated uncertainty can be highly uncertain because emission potential, compensation points, and NH_3 fluxes in these studies were not evaluated.

The WRF-Chem modeled NH_3 fields have been evaluated with data from the NADP/Ammonia Monitoring Network (AMoN, see revised Table S2 and Fig. S13) which to a certain degree can indicate the suitability of the applied NH_3 deposition method in this system.

The impacts of fertilizer inputs were not omitted but included in the CAMS anth emission input that is introduced in the following paragraph.

Lines 161-163: Agreed, but there has been recent work that addresses and quantifies this issue of using satellite retrievals to infer NOx emissions. See (Silvern et al., 2019; https://acp.copernicus.org/articles/19/8863/2019/) and Zhen et al., 2021; https://pubmed.ncbi.nlm.nih.gov/34149109/). Indeed there is much uncertainty given the increasingly large role of background NO2 emission sources.
Elguindi et al. (2020) discussed results in Silvern et al. (2019) to reach the point being referred to here. To make this clear, we added "*references therein*". Both Silvern et al. (2019) and Qu et al. (2021) are based on a global model which may not capture NO_x lifetime and budgets so well as fine-resolution systems. A few main sources of uncertainty were discussed by the authors of those papers.

Line 164: Please describe more about what plume rise approach is used and uncertainties. Plume rise has a major impact on wildfire emission transport.
The default WRF-Chem plume rise scheme based on Freitas et al. (2007) was used, as introduced by Grell et al. (2011). Grell et al. (2011) is now cited. The impact of plume rise vs. horizontal winds, emission intensity, and chemistry on wildfire emission transport and climate feedback on event-to-multiyear time scales is an active research topic being addressed by various communities (e.g., Veira et al., 2015).

Freitas, S. R., Longo, K. M., Chatfield, R., Latham, D., Silva Dias, M. A. F., Andreae, M. O., Prins, E., Santos, J. C., Gielow, R., and Carvalho Jr., J. A.: Including the sub-grid scale plume rise of vegetation fires in low resolution atmospheric transport models, Atmos. Chem. Phys., 7, 3385–3398, https://doi.org/10.5194/acp-7-3385-2007, 2007.

Grell, G., Freitas, S. R., Stuefer, M., and Fast, J.: Inclusion of biomass burning in WRF-Chem: impact of wildfires on weather forecasts, Atmos. Chem. Phys., 11, 5289–5303, https://doi.org/10.5194/acp-11-5289-2011, 2011.

Veira, A., Kloster, S., Schutgens, N. A. J., and Kaiser, J. W.: Fire emission heights in the climate system – Part 2: Impact on transport, black carbon concentrations and radiation, Atmos. Chem. Phys., 15, 7173–7193, https://doi.org/10.5194/acp-15-7173-2015, 2015.

Lines 171-172: And due to rising agricultural, plant fertilizer application and emissions.
Agreed. This information is included in the description here.

Line 191: I think this section is better titled: "Chemical observations from satellites, aircraft, and ozonesondes."
Changed as suggested.

Lines 215 – 231: While this is a very useful assimilation and comparison using satellite SM products, I think some uncertainty should be explained regarding lack of deeper soil moisture observations and understanding, and implications for drought. The top-level soil measurements at first 5 cm, woefully neglects the important impacts of rootzone SM on drought.
The impact of assimilating satellite surface SM on SM in deeper soil layers in part depends on the surface–subsurface coupling strengths of the used land systems (Kumar et al., 2009; Huang et al., 2022).

Land data assimilation that integrates surface SM (e.g., L-Band SMAP) as well as rootzone SM (e.g., P-Band AirMOSS and SNOOPI; thermal infrared ALEXI) and terrestrial water storage (e.g., GRACE and GRACE-FO) will likely lead to even more robust results. This is however not always true – see Figs. 8 and 9 in Hain et al. (2012). Related suggestions have been added to Section 4.

Hain, C. R., Crow, W. T., Anderson, M. C., and Mecikalski, J. R.: An ensemble Kalman filter dual assimilation of thermal infrared and microwave satellite observations of soil moisture into the Noah land surface model, Water Resour. Res., 48, W11517, https://doi.org/10.1029/2011WR011268, 2012.

Lines 219-223: This is a very broad statement, and no definitely understanding on how results presented here either qualitatively or quantitatively agree/disagree with the NADM (e.g., spatiotemporal comparisons). Please revise and be more explicit.
The NADM, as well as the USDM, is developed based on many sources of information by rotating authors. Therefore, it is partially subjective and designed to indicate various types of droughts - meteorological, agricultural, and hydrological. Inevitably, comparisons between SM data and the NADM/USDM have been typically qualitative.

Here, we now include state-level NADM drought extents and their temporal variability (Table S1) and discuss them together with Fig. 4a and the added standard deviations of SMAP SM in Fig. S3. For the period of case study #2, the Vegetation Drought Response Index maps are now shown together with a USDM map in Fig. S4.

Lines 243-245: Understood that long-term direct chemical flux measurements are limited, but what about the CASTNET (dry deposition) and NADP (wet deposition) networks?

As discussed in Huang et al. (2022) and references therein, dry deposition fluxes from the CASTNET dataset are partially model-based, which have known limitations and biases against eddy covariance flux measurements as well as fluxes estimated using other methods. Taking Referee #2's suggestion, we instead evaluated the modeled PM speciation with data from CASTNET and AQS sites, the modeled HNO_3 concentrations with CASTNET data, and the modeled NH_3 concentrations with NADP/AMoN observations. In addition, we compared the diurnal cycles of the modeled O_3 dry deposition velocity v_d at Harvard Forest for the study period with flux measurements reported in literature for previous decades. In the Supplement (Fig. S7), we mentioned: "*At Harvard Forest, WRF-Chem MJJ and measured $v_{d,o3}$ (during 1990–2000 June-July-August-September, Clifton et al., 2017) display similar diurnal cycles, with their daytime maxima and nighttime minima of 0.8–1.0 and <0.3 cm s^{-1}, respectively*".

The NADP/National Trends Network (NTN) nitrogen and sulfur wet deposition fluxes, as well as precipitation, have been analyzed and included in the revised Supplement.

Clifton, O. E., Fiore, A. M., Munger, J. W., Malyshev, S., Horowitz, L. W., Shevliakova, E., Paulot, F., Murray, L. T., and Griffin, K. L.: Interannual variability in ozone removal by a temperate deciduous forest, Geophys. Res. Lett., 44, 542–552, https://doi.org/10.1002/2016GL070923, 2017.

3 Results

Lines 318-339: I find parts of this section very speculative, and qualitative, where it is difficult to follow the emphasis (also due to writing issues, noted below).
This paragraph mostly describes the spatiotemporal variability of surface and column NO_2 based on quantitative results presented in several figures (Figs. 6, 7, and previous S5-S8/current S7-S10). The descriptions on the passive lightning tracer (no chemical reactions involved) and surface NO_2 measurements (known to be positively biased) have to be qualitative. We cannot find anything more specific from this reviewer's later comment on these lines.

Lines 343-346: So are the differences in dry deposition contributions a result of overestimated wet deposition in other literature/models, or underestimations of wet deposition based on WRF-Chem? This is a confusing and rather contradictory argument.
Both are reasons. Nr wet deposition was underestimated in WRF-Chem (see added evaluation in Table S2 and Fig. S11) and overestimated in some other studies/models, so contributions of dry deposition to the total in WRF-Chem are larger than in those other studies/models and possibly overestimated. We adjusted these two sentences and inserted descriptions on the new wet deposition evaluation results here.

Lines 341-365: Again, this paragraph largely compares the results from this modeling system based on WRF-Chem to other literatures, and has some rather speculative arguments. I think much of this paragraph could be trimmed, improved writing (see below), and improved discussion. Its plausible this section is more an assessement of the WRF-Chem modeling evaluated generally, and mainly qualitatively against other models, and some measurements. Not sure what is new here.

It's very common and important to compare results from one's own study with literature. We have added and discussed the new evaluation results and broken down this paragraph.

Lines 367-379: I think this could be improved by directly comparing the modeled spatial Nr deposition to critical load thresholds for different vegetation types and how it has changed from 2018-2023.

Simkin et al. (2016) critical load thresholds, as well as the lower and higher limits of the 95% confidence interval of these thresholds, are compared with the modeled Nr deposition fluxes for a subset of model grids. The estimated exceedances are now presented in Fig. S16.

Lines 393-394: I also think the much higher (lower) correlation coefficients against NO2 (HCHO) in year 2020 deserve to be highlighted and discussed briefly.

Fig. 11 shows that in 2020 O_3-NO_2 correlations were higher whereas O_3-HCHO correlations were much lower than those in other years. This model-based result suggests an overall stronger NO_x-sensitive regime in 2020 partly due to COVID impacts. This sentence has been changed to: "*Daytime surface O_3 concentrations exhibit more robust spatial correlations with early afternoon (19 UTC) NO_2 columns than HCHO columns, especially for 2020 due to COVID, with correlation coefficient r of 0.54 (0.62) and 0.40 (0.07) for all years (year 2020), respectively (Fig. 11)*". Related sentences in the abstract and Section 4 have also been updated.

Lines 425-428: Here it would be good to briefly describe the controlling parameters on decreasing CUO in the past 2018-2023, and that projected to continue to decrease in the future climate. Seems much uncertainty here, and justification is needed for discussion. If the eastern U.S. is projected to become wetter climate in the future, it would suggest increasing CUO, unless ozone concentrations decrease enough in proportion. Even with decreasing anthropogenic NOy it plausible that future increases in some GHGs, e.g., CH4, could lead also to widespread increases in ozone concentrations in the future, thus exacerbating CUO increases under a projected wetter climate in Eastern U.S.

The text here describes the spatiotemporal variability of CUO during 2018-2023, driven by various factors such as the changing emissions, land cover types, environmental and vegetation conditions. Many of these factors are intrinsically interconnected (see Section 1) and the spatially and temporally varying land-atmosphere coupling strengths are discussed in Sections 2.2.2 and 3.3.1. In terms of model uncertainty, the first two case studies offer insights into the SM impacts on regional O_3 and Nr and how one may improve models in these aspects. In this paper, we also indicate the model's incapability of accurately representing the impact of stratospheric intrusions on (near-)surface O_3 and the impact of omitting spatial variability in CO_2 forcing on photosynthesis and O_3 uptake. In addition, the estimated overall temporal changes in CUO from this work were discussed together with conclusions in Clifton et al. (2020), which were reached from one global climate model running with the RCP8.5 (now specified in paper). In Clifton et al. (2020), the changes in drought conditions from past to future as well as their associated uncertainty were not explicitly discussed. However, Cook et al. (2020) presented the projected SM, runoff and precipitation changes based on experiments with CMIP5/RCP8.5 and CMIP6/multiple SSPs. Those multimodel based results suggest drier soil conditions in future in many of the eastern US regions that could contribute to the overall decreasing CUO trends. Results in Cook et al. (2020) have been cited in Intergovernmental Panel on Climate Change (2021) and now also included in our discussions.

Cook, B. I., Mankin, J. S., Marvel, K., Williams, A. P., Smerdon, J. E., and Anchukaitis, K. J.: Twenty-first century drought projections in the CMIP6 forcing scenarios, Earth's Future, 8, e2019EF001461, https://doi.org/10.1029/2019EF00146, 2020.

Intergovernmental Panel on Climate Change: the Sixth Assessment Report, Summary for Policymakers, https://www.ipcc.ch/report/ar6/wg1, 2021.

Lines 447-449: This seems a relatively small impact of the SMAP SM DA on surface ozone bias and RMSE, relative to other much larger controlling factors on ozone formation and loss processes.

We now more clearly describe the DA impacts, which are non-trivial. "*Due to increased upwind pollution contributions whereas weakened local emissions and production, both enhancements and reductions by up to ~4 ppbv in daytime surface O_3 levels (not shown in figures) are found in the New England region (40.5–43°N, 70–74°W). Across the New England region, WRF-Chem daytime surface O_3 performance for 14 July was improved in 31 out of 50 of the model grids where AQS data were available, with the largest improvement of ~1.8 ppbv*".

There are many factors controlling models' O_3 performance. Improving one or more processes via DA often does not improve (or even degrade) O_3 performance due to the impacts of other error sources. And in such cases, free-running systems perform better for wrong reasons. It is encouraging to find that, in this case, O_3 performance was overall improved via the DA. It is unclear what exact "other controlling factors on O_3 formation and loss" this reviewer referred to and how large this reviewer estimates their impacts may be for this period/area.

Lines 465-466: This does seems more important locally (e.g., Northern Virginia), and would be much easier to see if paired ozone spatial bias plots (against AQS) were provided for both the Model no-DA vs. Model DA, instead of having to qualitatively compare the surface AQS obs against contour plots in Figure 14. Ultimately, I am concerned of the statistical significance of these ozone changes, and think some quantification of the significance is needed here.

A set of figures (Fig. S18) has been added to help better understand the DA impact on the modeled daytime O_3 interannual variability. Specific figure contents are: 1) the differences between Figs. 14i and 14h; 2) the *p* values of Student's *t*-tests that compare no-DA and DA daytime surface O_3 in July 2022 and July 2018 in all model grids; and 3) scatterplots of the modeled (no-DA and DA cases) vs. AQS daytime surface O_3 interannual differences in/near Northern Virginia. The added information helps to demonstrate that SM DA impacts on the modeled O_3 fields can vary year-by-year, due to many factors such as observation availability, the performance of SM in the no-DA case, and land-atmosphere coupling strength. In this paragraph, we now explicitly note that Northern Virginia is one of the subregions where SM DA impacts on the modeled daytime O_3 are strong for both July 2022 and July 2018.

Lines 484-508: In this section 3.2, a provided map of irrigated vs. non-irrigated lands is necessary to interpret the changes in Figure 15. Also, very difficult to interpret the noisy signal of Nr deposition, and as above comment, the significance of these changes are strongly in question for relevance and understanding. Suggest a statistical significance test is included on these changes, otherwise, the results presented here are very questionable.

Noah-MP's irrigation fraction input based on Salmon et al. (2015) is now shown in Fig. 1c. The soil type map (previous Fig. 1c) is now Fig. S1. Student's *t*-tests comparing daily full/reduced/no irrigation Nr deposition fluxes in all model grids suggest that the most meaningful irrigation impacts (where $p<0.05$, areas marked in green in Fig. 15) on Nr deposition are in/near the irrigated lands in the Carolinas.

Lines 520-522: I do not follow this argument, ¼-1/3 as large? This could stem from writing issues here, but very difficult to take anything from this argument scientifically.
Changed to: *"…only ¼–1/3 of its impact on free tropospheric O_3"*. See the added Fig. S21 for surface-level stratospheric O_3 tracer results. Stratospheric impact on O_3 aloft based on the WACCM stratospheric O_3 tracer is mentioned in the previous sentence. We recommend using the horizontal and vertical gradients of these O_3 tracer fields to help understand the stratospheric O_3 impacts instead of interpreting their absolute values as the stratospheric contributions to O_3.

Lines 529-540: These are very weak scientific arguments, and is only very cursory here with the ozone profiles in Fig. 17b and WRF-Chem/AQS spatial maps of ozone concentrations in Figure 17c-j. There is really no evidence provided here really isolating the elevated ozone with fire plume transport vs. other sources of extra regional ozone and precursor transport, when simply using clean (unrealistic) vs. base simulations in Figure 18. Indeed, these interactions are known to be very complex regarding ozone concentrations. More evidence and potentially source apportionment or sensitivity studies (e.g., simply fires on vs. fires off) would be needed to associate these areas with fire plumes vs. other sources of important precursors, and the related enhancements in daytime surface ozone formation.
The "clean BC" simulation is designed to help indicate upwind source (both fire and non-fire) impacts on the study area. Aside from stratospheric intrusions and Canadian fires, cross-state transport of pollution is a policy-relevant topic lately catching lots of attention: e.g., https://www.npr.org/2024/06/27/nx-s1-4996428/supreme-court-good-neighbor-plan , and many other media sources. A sentence has been added to make this point clearer. The fact that different scales of transport (e.g., trans-Pacific, stratospheric intrusions, and interstate) can be dynamically and chemically coupled to impact the western US O_3 was demonstrated in previous work (Huang et al., 2013).

An additional simulation "Sen" using chemical boundary conditions (BCs) from WACCM with an alternative fire emission input has been conducted. Fire emission has been identified as one of the most important configurations in global wildfire modeling. There is no standard fire-off NCAR/WACCM product for use as chemical BCs, which would also represent unrealistic conditions for the study period anyway.

The base and two BC sensitivity simulations as well as WACCM stratospheric O_3 tracers (see the added daily maps in Fig. S21) together help determine the impacts of stratospheric intrusions and transported Canadian fire (and other) plumes on surface O_3 during 13-16 June 2023.

Huang, M., Bowman, K. W., Carmichael, G. R., Pierce, R. B., Worden, H. M., Luo, M., Cooper, O. R., Pollack, I. B., Ryerson, T. B., and Brown, S. S.: Impact of Southern California anthropogenic emissions on ozone pollution in the mountain states: Model analysis and

observational evidence from space, J. Geophys. Res. Atmos., 118, 12,784–12,803, https://doi.org/10.1002/2013JD020205, 2013.

Lines 538-539: Would suggest adding more recent literature on the importance of fires, N deposition, and implications for downwind ecosystems. This is a growing field of importance. https://doi.org/10.1016/j.scitotenv.2022.156130; https://library.wmo.int/records/item/62090-no-3-september-2023; see Pages 7-8
Fires radiative/ecosystem impacts is indeed a growing field of importance. We cited the Koplitz et al. paper published in 2021 which described an earlier CMAQ-based study on this topic. A more recent HTAP3-Fires multimodel experiment paper covering this topic (Whaley et al., 2024, submitted to GMD in July 2024) is now also cited.

Summary and Suggested Future Directions

Lines 580-595: I find these arguments significantly broad and not well supported by the presented results in this paper. From what is presented, it is very difficult to determine, where this WRF-Chem configuration performed "remarkably" better than other platforms. Better identification, quantitative comparisons, and examples of improved results are needed. I assume much of this comment is pertaining to the inclusion of Land DA for SM and different simple case studies such as irrigation switches and turning off chemical LBCs, i.e., Clean scenario (Section 3.1-3.3). However, as presented, it is rather cursory arguments, which are not fully apparent how much better this system is able to represent the interactions of Nr and ozone formation.
Operational air quality forecasting systems can undergo multiple times of upgrades within years. Technical notes and peer-reviewed papers have been produced to document these upgrades and their impacts on model performance. There are such papers published in or currently under review for GMD, some of which are cited in Section 1. In such papers, old and new versions of models were usually run for a short period of time and the model results were compared with observations to demonstrate the effectiveness of the model upgrades.

This study serves different purposes, as noted in Section 1. To help determine O_3 spatiotemporal variability and the sources and processes controlling it during a multi-year period (including surface-atmosphere interactions which have growing importance and are understudied), high-resolution simulations with relatively consistent configurations and stable performance throughout the study periods are needed. As indicated in various sections of the paper and reiterated here, referring to AQS data, the model's O_3 performance is stable and overall better than what's reported in many previous studies (see some examples in Section 1, and language here has been adjusted). Certainly, many factors can contribute to successful model simulations, but through case studies, we highlight the importance of several of them and the needs to further investigate them. We do not deemphasize the importance of other factors controlling the models' O_3 performance, and in fact they were carefully considered (benefiting from test runs for short periods) as the baseline simulation was being set up.

The model simulation described here is already extended to 2024 (the TEMPO era), running on a routine basis. The study has connections with various communities, as well as implications for updating other models in the aspects being highlighted. This point is now explicitly made in this section along with extended uncertainty discussions.

Technical Corrections:

******Detailed grammatical errors and suggestions only shown here for Abstract and Introduction sections******

Abstract

Lines 15-20: Grammatical error. Run-on sentence, and needs revision.
Revised.

Line 20: Grammatical error. This statement "compared with and related to" is redundant.
"Compare with" has been removed.

Lines 23-24: Grammatical error. Remove comma.
Done.

1 Background, motivation, and goals

Lines 42-44: Grammatical error. The sentence structure is very awkward, and needs revision.
Revised.

Line 47: Grammatical error. Change "O3 via the aerosol radiative" to "O3 via aerosol radiative". Lines 47-48: Grammatical error. Remove "the" in "via the aerosol radiative effects".
Done.

Lines 48-54: Grammatical error. Run-on sentence, and needs revision.
Revised.

Line 55: Suggest changing "would be" to "is".
Done.

Lines 58-59: Grammatical error. Awkward sentence structure, and cannot understand the connection the author is making with "…and carbon dioxide (CO2) concentration as well as plants' physiological conditions."
Changed to: "*closely interact with multiple other interconnected environmental stressors (e.g., temperature, humidity, precipitation, soil moisture, SM, and carbon dioxide, CO_2) and plants' physiological conditions*".

Lines 62-64: Grammatical errors and inappropriate verbiage. "…continue to decrease there due..", "…for studies on Nr and O3, attention should…", and "imported".
This sentence has been broken down and reworded.

Line 68: Grammatical error. Need comma, "…and the estimated background O3, as well as…"

This sentence has been broken down and reworded.

Lines 74-75: Grammatical error. Run-on sentence, and needs revision.
This sentence has been broken down.

Line 76: Grammatical error. Awkward sentence structure. "…limits the capability of understanding air quality there and evaluating…"
Changed to: "*limits our capability of understanding air quality there and evaluating…*"

Line 78: Awkward verbiage. Suggest changing "is anticipated to" to "will".
Changed to "*can*".

Lines 80-85: Grammatical error. Run-on sentence, and needs revision.
Revised.

Lines 88-97: This is too long for a bulleted list, particularly difficult to read in bullet 3). Suggest separating it out of the paragraph and shortening into more bullets to make easier to read and understand.
The three bullets correspond to Sections 3.1, 3.2, and 3.3 of the results section. We have changed the structure of the sentences/phrases related to bullet 3).

Results

Lines 318-339: I find this section needs significant writing improvements, as discussed above. Also, it would be best to break this paragraph up into multiple paragraphs.
See response to your earlier comment on these lines.

Lines 341-365: Writing needs significant improvement and needs multiple paragraphs.
See response to your earlier comment on these lines.

Lines 381-431: Writing needs significant improvement and sentence structure needs substanitial improvement. Currently it is difficult to follow the arguments.
Note: Similar writing improvements are needed through the remaining results section.
See response to your earlier comment on this paragraph. These lines have been broken down and revised.

Summary and Suggested Future Directions

Technically, this section also needs similar significant writing improvements and is very cumbersome to read. Highly recommend thorough proof-reading in revised manuscript.
This section has been revised according to your earlier comments and proofread.

Response to RC2

The MS (egusphere-2024-484) by Huang conduced model simulation of Reactive nitrogen in and around the northeastern and Mid-Atlantic US, and analyzed its influence on O3 and plant, it shows a lot of model simulation work by considering different model setups and also analyzing so many components, which follows Huang's previous studies as list in references. However, the performance of model simulation, especially for dry/wet deposition and the influence of O3 on plants should be furthered carefully evaluated, which can be potential large uncertainty for this study.
Thanks for the summary. Please see below for the added model evaluation work and discussions on sources of model uncertainty.

other comments as:

Line 111-113, "Noah-MP's CO2 forcing for 2018, 2019, 2020, 2022, 2023's warm seasons were set to 410, 412, 415, 420, and 423 ppmv, respectively, based on measurements at the Mauna Loa Observatory and its nearby Maunakea Observatories for part of 2023", why this study choose GHG background values of Mauna Loa as the CO2 forcing for the urban area, which can have much higher CO2 concentration as >450 ppm, and affect photosynthesis of plants.
The CO_2 forcing for the Noah-MP land surface model is typically set as a constant value and therefore including year-to-year changes in that forcing is already an advance.

The increases in CO_2 are seen at the Mauna Loa Observatory and across the globe at similar speeds of approximately 2-3 ppmv year^{-1} for recent years, according to observations from the NOAA GML network, satellites, and model/analysis fields (e.g., https://nasaviz.gsfc.nasa.gov/5194; https://gml.noaa.gov/ccgg/trends/gl_gr.html; https://gml.noaa.gov/webdata/ccgg/CT2022/CT2022.global_AGR.pdf). We recognize the different magnitudes and seasonal variability of rural and urban CO_2 (Fig. 11 in Karion et al., 2020), which in some years present anomalies due to COVID (Weir et al., 2021), and these were not accounted for in our configurations. However, the impact of ignoring these differences (tens of ppmv) on the photosynthesis-based dry deposition estimates is likely to be very small according to independent global model sensitivity analysis (e.g., Fig. 12 in Sun et al., 2022; Fig. 7 in Silva et al., 2023), and is worth future investigations with finer-resolution models. The need to develop high-quality, spatially and temporally varying CO_2 forcings for Noah-MP, especially in their longer-period simulations, was brought up at a recent Noah-MP Users' International and similar occasions. This is now also mentioned in Section 4.

Karion, A., Callahan, W., Stock, M., Prinzivalli, S., Verhulst, K. R., Kim, J., Salameh, P. K., Lopez-Coto, I., and Whetstone, J.: Greenhouse gas observations from the Northeast Corridor tower network, Earth Syst. Sci. Data, 12, 699–717, https://doi.org/10.5194/essd-12-699-2020, 2020.

Silva, S. J., Burrows, S. M., Calvin, K., Cameron-Smith, P. J., Shi, X., and Zhou, T.: Contrasting the biophysical and radiative effects of rising CO_2 concentrations on ozone dry deposition fluxes, J. Geophys. Res. Atmos., 128, e2022JD037668, https://doi.org/10.1029/2022JD037668, 2023.

Sun, S., Tai, A. P. K., Yung, D. H. Y., Wong, A. Y. H., Ducker, J. A., and Holmes, C. D.: Influence of plant ecophysiology on ozone dry deposition: comparing between multiplicative and photosynthesis-based dry deposition schemes and their responses to rising CO_2 level, Biogeosciences, 19, 1753–1776, https://doi.org/10.5194/bg-19-1753-2022, 2022.

Weir, B., et al.: Regional impacts of COVID-19 on carbon dioxide detected worldwide from space, Sci. Adv., 7, eabf9415, https://doi.org/10.1126/sciadv.abf9415, 2021.

Line 155, "missions Database for Global Atmospheric Research version 5 based on the Community Emissions Data", please illustrate what the spanning years for EDGAR v5.0 for these pollution species.
Added "*for the years after 2015*".

Line 165-174, the authors mentioned the annual variations of different pollution species, it's much better to illustrated them with time series figure than worlds.
Agreed. That information is indicated in Fig. 6a.

Section 2.2.3 ground-based observations, why the site based PM2.5 PM10 the components SO42- NO3- NH4+ were not compared with model simulations, which can support your model's performance regarding atmospheric chemical reaction, dry/wet deposition and pollution emissions.
The following model evaluation work has been added, along with discussions:

1) Wet deposition fluxes of SO_4, NO_3, and NH_4, as well as precipitation, evaluated with NADP/NTN data;
2) Surface SO_4, NH_4, and NO_3 concentrations evaluated with CASTNET (remote/rural) and AQS (urban/suburban) observations;
3) Surface HNO_3 concentrations evaluated with CASTNET observations; and
4) Surface NH_3 concentrations evaluated with NADP/AMoN data.

Additionally, literature on surface speciated aerosol trends based on the IMPROVE observations (Hand et al., 2024) is now also cited in the Supplement.

Hand, J. L., Prenni, A. J., and Schichtel, B. A.: Trends in seasonal mean speciated aerosol composition in remote areas of the United States from 2000 through 2021, J. Geophys. Res. Atmos., 129, e2023JD039902, https://doi. org/10.1029/2023JD039902, 2024.

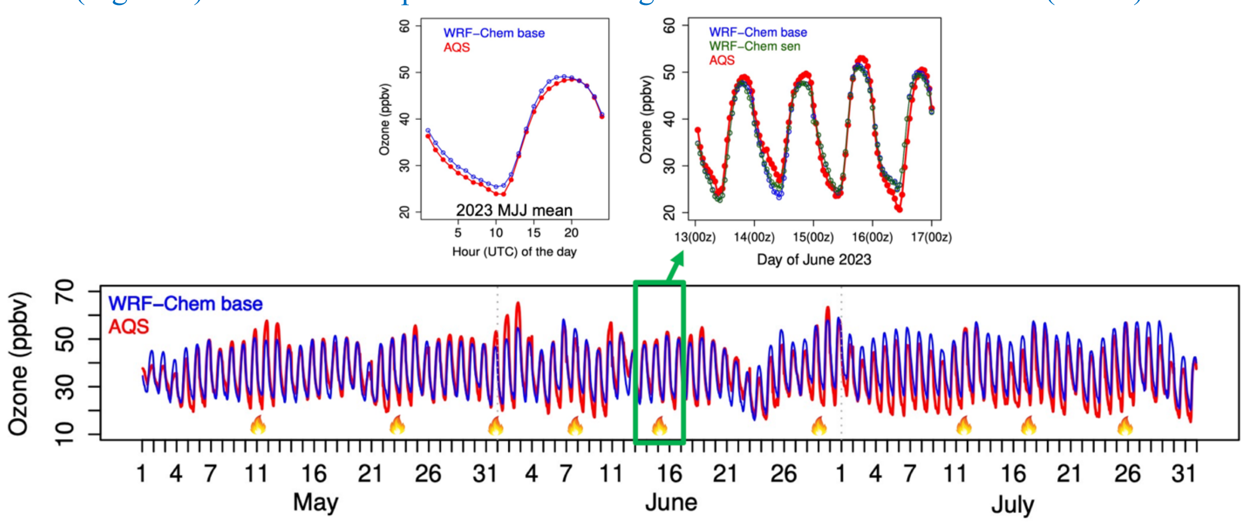
Section 2.3.2 I am still wondering whether the plant models in your study can well represent the harmful O3 effect on stomate, where the parameters and plant model structure can largely affect your evaluation.
As introduced in Section 2.1, O_3 vegetative impacts were dynamically modeled by applying two separate factors $F_{p,O3}$ and $F_{c,O3}$ (which are linearly related to CUO) to photosynthesis and stomatal conductance rates, and an O_3 flux threshold to account the ability of plants to detoxify O_3 was applied.

Key sources of uncertainty include: 1) methods to calculate stomatal conductance/resistance, including model structures; 2) slopes and intercepts used to estimate $F_{p,O3}$ and $F_{c,O3}$ for a limited number of plant types; and 3) the O_3 flux threshold used to account for the ability of plants to detoxify O_3. 1) has been discussed in depth in many previous studies including Huang et al. (2022), Sun et al. (2022) and references therein. The Ball-Berry type of approaches have shown advantages over multiplicative and Medlyn based approaches. Some of the inputs of the stomatal models such as SM could be improved through data assimilation, as highlighted in Huang et al. (2022). Improving the model's CO_2 forcings and assimilating other datasets are encouraged (Section 4), according to both reviewers' comments. 2) and 3) were taken from literature with evidence based on measurements, which can certainly be uncertain for this case, and we note in Section 4 that evaluating and improving these parameters for more types of plants at various growth stages in future is encouraged.

Sun, S., Tai, A. P. K., Yung, D. H. Y., Wong, A. Y. H., Ducker, J. A., and Holmes, C. D.: Influence of plant ecophysiology on ozone dry deposition: comparing between multiplicative and photosynthesis-based dry deposition schemes and their responses to rising CO_2 level, Biogeosciences, 19, 1753–1776, https://doi.org/10.5194/bg-19-1753-2022, 2022.

Section 3.1 usually the model simulated results should first compare with observations (not all species, depends on what observations the authors have as illustrated in method section) to verify the performance of model. It's easy to run the model and analyzed model simulations, but it's hard to tell us whether the simulations from your model parameter and emission setup were reliable. Here on line 407, I just notice your comparison with surface O3, and I am not that confident your model can well simulate the spatial-temporal variations of O3 changes, as you only displayed the averages of a period, instead of hourly observations, with considerable bias. Have you considered the impact of stratospheric intrusion on ozone enhancement in the lower troposphere with upper O3 boundary condition scheme?

Section 3.2 focuses on interannual differences in O_3. In case study #3, we added a timeseries plot of the domain-mean observed and WRF-Chem hourly surface O_3 during 13-16 June 2023 at AQS sites (current Fig. 17c). This plot is now discussed together with O_3 timeseries for MJJ 2023 (Fig. S24) and the model performance during other fire events in MJJ 2023 (below).



The challenge for regional chemistry models, especially those systems without accurate dynamic upper chemical boundary conditions, to well capture the impacts of stratospheric intrusions on O_3 enhancements had been mentioned here in Section 3.2 as well as our previous studies, and now also explicitly in case study #3. This challenge contributes to the slight negative biases in

WRF-Chem daytime O_3 for this case, but the model still did an excellent job in reproducing the observed hourly O_3. An important message of this paper is that proper updates on WRF-Chem related to land and land-atmosphere interactions are transferrable to other regional models, including those running with dynamic upper chemical boundary conditions (now mentioned in both Section 3.3.3 and Section 4).

Section 3.2 irrigation approaches, on line 486, the "Ozone perturbs gross primary productivity more strongly (up to 20–30%) than transpiration", as I mentioned above, whether there are observation-based study that displayed similar results? because the plants model can not well simulate the feedback between O3 and plant. To me The GPP decreased by 20-30% only caused by O3 is not reliable, see the situations in China and India, large O3 concentrations occurred in summer, but the influence on crop production did not change too much. From my experience, even the influence of O3 on plants have not been well investigated by field observations, how can it be well represented by model equation and structure?
Please see our response to your earlier comment on Section 2.3.2. Also, the modeled GPP from the baseline simulation (including O_3 impacts) was compared with the MODIS Terra/Aqua 8-day GPP product (also known to be uncertain despite its wide usage, especially for locations with frequent cloud cover and high GPP) for June 18-25, 2022. Both the model and MODIS indicate that GPP was <0.04 kgC m^{-2} over dry croplands and 0.10-0.12 kgC m^{-2} over humid forest regions.

A key point from this paper is that the O_3 impacts on surface fluxes and vegetation are sensitive to various environmental factors. We are not sure under which conditions "in China and India, large O_3 concentrations occurred in summer, but the influence on crop production did not change too much". This finding may be cited if more information can be provided. Fig. 3 in Lombardozzi et al. (2015) shows 20-year average ozone impacts on GPP across the globe - for some places in the US, Asia, and Africa, these impacts were estimated to be >25%. Some of us are aware of multiple papers on the O_3 impacts on various types of ecosystems by TOAR-II vegetation team members such as Pandey et al. (2023) for India as well as a few in-review and in-preparation papers for this special issue. Results from these studies are/may also be informative.

Pandey, D., Sharps, K., Simpson, D., Ramaswami, B., Cremades, R., Booth, N., Jamir, C., Büker, P., Sinha, V., Sinha, B., and Emberson, L. D.: Assessing the costs of ozone pollution in India for wheat producers, consumers, and government food welfare policies, Proc. Natl. Acad. Sci., 120(32), e2207081120, https://doi.org/10.1073/pnas.2207081120, 2023.

Response to Dr. Owen Cooper's comments

1) When discussing the impact of COVID on ozone it would be helpful to cite a new paper published in the TOAR-II Community Special Issue. Putero et al. (2023) show that ozone decreased in 2020 at high elevation sites across the western USA, and also at four high elevation sites in the eastern USA. Another paper that is relevant to the COVID period is Steinbrecht et al. (2021) who show that ozone also decreased in the free troposphere of northern mid-latitudes.
Both papers are cited now.

2) When discussing long-term ozone trends across the USA, the papers that are currently available in the peer-reviewed literature are out-of-date, as they all seem to end in 2014 or 2015. However, EPA provides regular ozone trend updates on the following webpage: https://www.epa.gov/air-trends/ozone-trends . They focus on the 98th percentile, or the annual 4th highest MDA8 ozone value.
Newer papers and websites on O_3-concentration trends are also cited now. In addition, we emphasize the importance of analyzing flux-based metrics that are more relevant to assessing O_3 vegetation impacts.

3) As summarized in the TOAR-II "Guidance note on best statistical for TOAR analyses", the TOAR community is abandoning the expression "statistically significant" for the reasons described by Wasserstein et al. (2019). Please follow these recommendations and replace "statistically significant" on line 396 by describing your confidence in this result.
Changed to: "*where the p values of the correlation tests are lower than 0.05*".