Dear Editor:

We are grateful for the highly constructive comments from both reviewers. We have carefully considered the comments provided and made corresponding revisions to improve the manuscript. Below, we provide a point-by-point response to each comment. The original comments from the reviewer are in *italicized black text*, our responses are in normal font, and changes in the manuscript are highlighted in "blue". Line numbers referenced in the reviewers' comments correspond to the submission from November 2025, while those cited in our responses align with this revision.

In response to concerns raised by the reviewers, we have substantially revised and strengthened the manuscript in the following ways:

- **Clarified motivation and context**: We expanded the Introduction, Abstract, and Discussion to better situate this study within the existing literature, clearly articulating its role as both a standalone quantification of emission temporal resolution effects and as foundational work for future applications using high-frequency satellite observations.
- **Quantified implications for emission inference**: We added new analyses that explicitly connect modeled NO_2 column sensitivity to hourly emission variability by quantifying the potential bias in inferred NO emissions when sub-daily emission variability is neglected.
- **Added targeted analyses and figures:** In response to specific reviewer concerns, we introduced new analyses and supplemental figures that (i) verify time-zone handling of hourly emissions (Fig. S6), (ii) document the locations of surface monitoring stations (Fig. S8), and (iii) assess averaging-kernel interpolation (Fig. S9); in addition, we added a new main-text figure (Fig. 4) to more clearly show urban-rural contrasts, complementing Fig. 3.
- **Revised text for clarity and organization**: We refined the manuscript throughout to improve clarity and flow, including clearer descriptions of experimental design, emission temporal characteristics, and links between surface- and column-based diagnostics.

Thank you for the opportunity to revise the manuscript. We hope these changes have sufficiently improved the manuscript so that it is now acceptable for publication.

Best regards,
Madankui Tao

-- Response to Reviewer 1 --
Tao et al. implemented hourly- or daily-varying emissions into the version 0 Multi-Scale Infrastructure for Chemistry and Aerosols (MUSICA) model and evaluated the impact of 3 different time-varying emissions

inventories. These included (1) hourly variation in all emissions, (2) hourly variation in NO only, (3) daily variation in NO only. These were compared against (1) a base case using a global emissions inventory, (2) a case using the US EPA NEI monthly average inventory over the continental US, (3) monthly NEI with NO emissions reduced by 30%, and (4) monthly NEI with VOC emissions reduced by 30%. The model runs are compared against TROPOMI NO2, HCHO, and CO data, as well as surface data. The latter has a heavy focus on 6 US cities (New York, Atlanta, Houston, Chicago, Denver, and Los Angeles). Implementing hourly emissions does have a clear impact on simulated NO2, O3, and CO; however, the authors show that doing so does not uniformly decrease bias compared to satellite or in situ observations. The authors explore how the impact of hourly emissions depends on time of day and day of week, as well as factors in the model driving regional differences in sensitivity to NO emissions. They conclude that changing from monthly to hourly emissions will have significant impacts on NO2 and O3 concentrations, even if the overall emitted mass is similar.

This work is technically sound. My main comments are around the context/motivation for the work and organization of the paper to improve clarity. I therefore recommend publication after those issues are addressed.

We thank the reviewer for the positive assessment and constructive feedback. We have addressed the comments regarding context, motivation, and organization, and provide a point-by-point response below.

General issues

My first general comment is: what is the motivation or context for this work? Testing the impact of hourly-varying emissions seems like an intermediate step in a larger project. That's perfectly fine; it makes a lot of sense to publish the groundwork as its own paper to cite for later steps. However, it would be helpful to know whether there is a next step planned, and if so, what it is to put this work into context. Alternatively, if the case is that this was simply an effort to quantify the effect of hourly-varying emissions, then providing sufficient background to show that, while it may be expected that implementing these emissions would have a complicated effect on air quality simulations, the actual effect has not yet been quantified. Or, if it has been quantified but for other models, referencing that previous work and describing the motivation to test how MUSICA compares to those studies would be important context.

Thank you for this thoughtful comment. We tried to clarify the broader context regarding the motivation and scope of this work. This manuscript serves the dual purpose of (1) a standalone

quantification of the effects of hourly-varying anthropogenic emissions (i.e., demonstrating that neglecting hourly variations leads to non-trivial difference in the simulated time-averaged concentrations, with follow-on implications for relying on those simulated concentrations to infer emissions) and (2) foundational work to enable applications of MUSICAv0 using emission inventories with hourly temporal resolution.

As the reviewer surmised, our initial intent was not simply focused on the role of hourly emissions but to further interpret earlier findings regarding ozone sensitivity in the New York City region (Tao et al., 2022; 2025) by quantifying anthropogenic and wildfire influences on surface ozone in this region. We thus configured MUSICAv0 with the latest detailed emissions inventory (2017) available and adjusted the inventory to summer 2018. But the intriguing spatial differences in simulated NO_x and O_3 concentrations associated solely with the temporal resolution of anthropogenic emissions motivated us to extend our analysis to the broader CONUS region. In response to this constructive feedback from the reviewer, we have revised the manuscript in two main ways described below.

First, we more clearly situate this work within the existing literature and our quantitative assessment of emission timing effects in the introduction (lines 58-65): "Prior regional chemical transport model (CTM) studies have examined the sensitivity of simulated surface air pollutant concentrations to sub-daily (hourly or diurnal) variations in anthropogenic emissions across a range of modeling frameworks (e.g., WRF-CMAQ, CHIMERE). These studies show that the temporal allocation of emissions can influence pollutant concentrations and their daytime-nighttime contrasts, particularly for species with strong diurnal variability such as NO_x and O_3 (e.g., Guevara et al., 2025; Jo et al., 2023; Menut et al., 2012; Shen et al., 2023). Despite their increasing relevance for interpreting retrievals from geostationary satellite missions such as Tropospheric Emissions: Monitoring of Pollution (TEMPO), a systematic quantification of how hourly anthropogenic emissions affect daily or monthly averages of both surface and column concentrations over the United States remains lacking."

Second, we quantify the potential impact of the simulated concentration differences arising solely from the temporal resolution of the emissions for the inference of emissions from tropospheric columns retrieved from space. We now introduce a simple scaling analysis to estimate the potential bias in the NO emission magnitude inferred from NO_2 columns retrieved once per day if they rely on relationships simulated in models that neglect hourly emission timing. This analysis is summarized in Section 4.3 and visualized in the new Fig. 8. Additional detail is provided in Text S6, with regional- and city-scale statistics reported in Table S6.

My second general comment is around the organization of the paper. The abstract focuses only on the impact of implementing hourly emissions. However, the case studies with 30% smaller NO and VOC emissions as part of the urban case studies are also a significant part of the results. It would help the reader

if this part was introduced in the abstract, as well as if the last paragraph of the introduction was clearer that (a) the urban case studies included emission reductions scenarios and (b) why those are included or how they tie back into the hourly emissions test.

Thank you for this comment. Our core scientific objective is to quantify the impacts of resolving hourly anthropogenic NO emissions on simulated surface and column air pollutants. The additional sensitivity experiments and urban case studies are included to aid interpretation of this primary objective, as clarified in the first paragraph of Section 3: "The idealized 30% emission reduction experiment is included to provide a benchmark for the magnitude of effects associated with emission temporal resolution, while the urban case studies are used to illustrate these effects at the local scale (e.g., a single model grid cell), where responses are most pronounced and easiest to interpret."

We also revised the Abstract and Conclusions to more clearly introduce the role of the idealized emission reduction and urban case studies. Specifically, we added the following sentence to the Abstract: "While comparable in magnitude to a uniform 30% NO emission reduction (grid-cell level differences of - 12 % to +9 %, -7 to +3 ppb for O_3), the spatial response patterns differ with location-specific timing of emissions and meteorology. For example, higher morning NO_x concentrations and stronger NO_x-saturated O_3 suppression occurs in Los Angeles compared to New York City."

We also revised the first paragraph in Discussion and Conclusions to integrate them with the previously summarized key takeaways: "We demonstrated that incorporating hourly variations in NO emissions produces substantial changes in NO_x and O_3 concentrations, even in monthly averages, comparable in magnitude to those associated with an idealized 30% reduction in NO emissions (Figs. 3 and 4). These changes, however, are nuanced, differing between urban and rural areas and between the eastern and western U.S., reflecting region-specific emission patterns, photochemistry, and meteorological conditions (Fig. 5). Monthly mean daytime surface O_3 concentrations over the CONUS in July differ by up to 7 ppb (11% for NEI_hourly relative to NEI_monthly; Fig. 5), large enough to affect model-based conclusions regarding regional attainment of the U.S. NAAQS for O_3. In the western CONUS, monthly mean surface NO_2 increases by up to ~6 ppb during the daytime. In contrast, the eastern CONUS with its higher NO and biogenic VOC emissions, larger dry deposition flux rates, more humid conditions, and shallower PBLH—all of which favor a shorter NO_2 lifetime—experiences daytime NO_2 decreases of up to ~1 ppb (Fig. 5). We also find differences in O_3 production regimes across cities: for example, a shift toward more NO_x-saturated conditions in Los Angeles but toward stronger NO_x sensitivity in New York City when hourly variations in NO emissions are included (Figs. 6 and 7)."

Minor comments

Lines 24-25: "We conclude that models used to infer NOx emissions from monthly mean concentrations may alias hourly emission variations into the inferred magnitude of emitted NO." It's not clear to me what this means, particularly since the paper as written does not include any quantification of how inferred emissions would change. The results as given all focus on forward model results, rather than the inverse inference of emissions from concentrations (monthly mean or otherwise). Making the jump from forward effects to inverse effects is difficult to do mentally, as there's a lot of interaction between grid cells with transport and chemical nonlinearity that make this difficult to intuit. If you want to retain this conclusion, a new discussion section that walks the reader through your chain of reasoning would be necessary to support this statement. Since this comes back in the discussion & conclusions, see my final comment for a more specific suggestion there.

We agree that the inference we made in the original manuscript required additional fleshing out. As noted in our response to the reviewer's later comment (Lines 505-507), we have added a new quantitative analysis to directly address this issue. Specifically, we follow Lamsal et al. (2014) to estimate from our simulations how unresolved hourly emission timing can be aliased into the inferred NO emission magnitude when interpreting monthly mean tropospheric NO_2 columns. This analysis is presented in a new subsection (Section 4.3) as well as in the Discussion and Conclusions, with full methodological details in Text S6 and results summarized in our new Figure 8 and Table S6.

Lines 66-67: "We select July 2018 due to the availability of trace gas retrievals from the TROPOspheric Monitoring Instrument (TROPOMI) and concurrent field campaign observations." Given that the focus of the paper is on hourly emissions, why not select July 2024, when TEMPO was observing? What field campaign observations are available in 2018 that wouldn't be available during a TEMPO-observed time period? It's not clear that the SLAMS data was a short term set of observations, and there does not seem to be any other field data used in the paper. Also, why was a winter month not tested as well, given the strong impact of photolysis and meteorology on the results? (I presume the reason is that summer remains the season with the most O3 exceedences, but this should be stated explicitly.) This answer can likely tie back to my first general comment, i.e., how this study fits into a larger project, if it does.

Thank you for catching that our original wording did not sufficiently clarify the rationale for the selected period. We removed the reference to "concurrent field campaign observations," as this study does not rely on campaign-specific measurements; as the reviewer notes, the SLAMS data provide long-term regulatory surface observations and are not unique to 2018.

We have revised the manuscript to explicitly state that "We select July 2018 as a representative summer month for our analysis because it coincides with the availability of TROPOspheric Monitoring Instrument (TROPOMI) observations beginning in May 2018. Our summer focus reflects active photochemistry during this season, when surface O_3 concentrations are typically highest and most likely to exceed the U.S. National Ambient Air Quality Standards (NAAQS) (Tao et al., 2022, 2025)."

Sect. 2.1: It would be easier to follow if this section began with a short paragraph introducing the reader to the experimental design, that CAMS will be used as the baseline global emissions, NEI for the more detailed CONUS emissions, and that different variations of NEI emissions will be used to test the model response to different emissions scenarios. The rest of the section is fine, adding just that extra context up front would help understand why the two different inventories are used.

Thank you for this suggestion. We have revised the first paragraph of Section 2.1 to explicitly introduce the experimental design and clarify the roles of these two anthropogenic emissions inventories:

"To evaluate the sensitivity of simulated air pollutant concentrations to the temporal resolution of anthropogenic emissions, we use CAMS-GLOB-ANT as the global baseline inventory and replace it with the U.S. Environmental Protection Agency (EPA) National Emissions Inventory (NEI) over the CONUS, which provides more detailed (higher spatial resolution; hourly-varying) regional emissions. We then conduct a series of simulations that differ only in the temporal resolution of NEI emissions (monthly, daily, or hourly) to isolate the model response to emission timing while holding July total monthly emissions constant."

Lines 143-144: "...as in NEI_monthly, but applying daily NO emissions with weekday-weekend differences (NEI_daily_NO)." So do NEI_hourly and NEI_hourly_NO not include weekend/weekday differences? Please be clear about which emissions scenarios include what temporal variation - it might be clearer to add columns for "temporal resolution" (e.g., monthly, daily, hourly) and "weekend/weekday differences" (yes or no) to Table 1.

Thank you for this clarification request. We have clarified this distinction in Section 2.1 lines 157-160: "Hourly NEI emissions (NEI_hourly and NEI_hourly_NO) include day-specific diurnal variability (and therefore distinguish between weekdays and weekends), whereas NEI_daily_NO removes within-day (diurnal) variability (Table 1)." and in Table 1 with new columns to make the temporal characteristics of each emissions scenario explicit.

Line 163: "We use measurements collected from State and Local Air Monitoring Stations (SLAMS)..." A map showing how many stations are available for each grid cell would help clarify how representative the SLAMS data is of the regions shown in Fig. 1. This is important for interpreting Fig. 2.

Thank you for this suggestion. To clarify the representativeness of the SLAMS observations used here, we have added a supplemental figure (Fig. S8) showing the spatial distribution of SLAMS monitors for each species, along with a table summarizing the number of available monitors within the individual model grid cell corresponding to each selected city. We also added the following text in lines 194-199: "The locations of SLAMS used in this study for each species, along with their distribution across CONUS regions and the number of monitors within selected urban grid cells, are shown in Fig. S8.

We select six monitoring stations as examples to examine diurnal patterns at individual sites (Table S5), chosen based on their proximity to major city centers across the CONUS and the availability of continuous NO_2 and O_3 measurements throughout July 2018. Nearby stations with similar diurnal behavior were excluded to avoid redundancy. We do not average across sites as we prefer to preserve distinct local features given differences in monitor availability across cities (Fig. S8)."

Lines 190-191: "...and apply the TROPOMI AKs, linearly interpolated vertically to the MUSICAv0 vertical resolution..." Unless the MUSICA profiles have significantly better vertical resolution than the TROPOMI AKs or the MUSICA profiles aren't reasonably smooth, it's usually better to interpolate to the AK levels, as this avoids interpolation issues around the sorts of sharp jumps in the AKs that occur when a pixel is partly cloudy. This is also something to use care with when averaging the AKs together to the grid cell scale. Since it isn't specified whether you filtered the TROPOMI data according to any QC variables, it's impossible to know whether clouds will impact the results. (So please also clarify what TROPOMI filtering criteria were used.) A supplemental figure comparing the averaged and interpolated AKs to the original AKs for a few different grid cells (perhaps one clear and one partly cloudy) would help give the reader confidence that the averaging of multiple pixels' AKs together to the grid cell level and interpolation to the MUSICA levels is not introducing unphysical smoothing of the AKs.

Thank you for this comment. The TROPOMI quality filtering criteria are specified in the Methods: "We select pixels with quality assurance greater than 0.75 (Table S4),…", which is a standard and relatively strict threshold that applies to all of the retrievals we use here (NO_2, HCHO, and CO). We have added the following for further clarification: "…which largely excludes cloudy and partly cloudy pixels (van Geffen et al., 2020; Lange et al., 2023)."

Interpolating the TROPOMI AKs to the MUSICAv0 vertical grid allows a single, consistent set of AKs to be applied across all MUSICAv0 simulations, ensuring that differences among sensitivity cases arise solely from changes in emissions. In addition, the vertical resolution of the TROPOMI retrievals is not substantially finer than that of the MUSICAv0 model as the model-based a priori profiles provided with the Level-2 products are comparable (34 layers for tropospheric NO_2 and HCHO, 50 layers for atmospheric CO; all *a priori* profiles use TM5-MP, versus our 32 layers in MUSICAv0).

We add a new supplemental figure comparing the original and interpolated TROPOMI averaging kernels to confirm that we are not introducing unphysical smoothing with this interpolation. Corresponding text and Fig. S9 have been added to the revised manuscript in lines 214-219: "We calculate the average between 1 and 2 p.m. local time (to approximate 1:30 p.m. values) for each region and apply the TROPOMI AKs, linearly interpolated vertically to the MUSICAv0 vertical resolution, to ensure a consistent application of AKs across all MUSICAv0 sensitivity simulations when calculating modeled VCD_{Trop} of NO_2 and HCHO and VCD_{Total} of CO. Our interpolation of the TROPOMI AKs to the MUSICAv0 vertical grid generally preserves their vertical structure without introducing unphysical smoothing (Fig. S9)."

Lines 198-199: "Across the six CONUS regions, spatial correlations between modeled and observed surface concentrations are stronger for NO2 and O3..." For which model configuration(s) are these correlations calculated?

Thank you for this clarification request. We revised accordingly (now lines 279-280): "…between observed and modeled surface concentrations in the BASE simulation (with the global CAMS-GLOB-ANT emissions)…"

Lines 215-216: "CO concentrations increase relative to the BASE simulation..." Could point to Fig. 3 here - in general, if text is discussing multiple figures within the same paragraph, it is helpful to include liberal cross references to figures or tables to be clear which results are being discussed.

We added an explicit reference to Fig. 3 and have also reviewed and revised the full text to include figure cross-references where multiple figures are discussed within the same paragraph.

Line 223-224: "Compared to the BASE case, the NEI_monthly simulation worsens the model underestimates of NO2 VCDTrop by 16-21% (~1×1014 molecules/cm2)..." I've also seen in the past that models have lower background NO2 tropospheric columns that satellite observations, going back to OMI. This might be due to incorrect NO2:NO ratios in the free troposphere

(https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2018GL077728) or the stratospheric/tropospheric separation in the satellite products. If you think your modeled stratosphere is accurate, you could see if the total NO2 columns have less bias - that would remove the strat/trop separation algorithm as a source of uncertainty (though it then adds in the stratospheric column as a new source). Otherwise this could be part of the discussion.

We thank the reviewer for highlighting prior evidence of low model biases in background tropospheric NO_2 columns and the potential roles of free-tropospheric NO_x chemistry and retrieval uncertainties. We have expanded the discussion below in lines 257-265:

"Biases in trace gas columns do not always match those at the surface. For instance, surface NO_2 shows high biases in some regions especially for the BASE simulation, while NO_2 VCD_{Trop} are consistently biased low relative to TROPOMI (Fig. 2). These different biases in the surface versus column can arise from several factors that are independent of surface emission magnitude, including (i) vertical sensitivity and representativeness, given that satellite columns reflect vertically weighted pixel means whereas surface observations are point measurements; and (ii) retrieval uncertainties, including assumptions in the air mass factor calculation and the stratosphere-troposphere separation applied to derive tropospheric NO_2 columns (van Geffen et al., 2020). In addition, prior global chemistry models have been shown to underestimate background NO_2 VCD_{Trop} relative to satellite products, partly due to uncertainties in free-tropospheric NO_x chemistry and partitioning (e.g., NO/NO_2 ratios and partitioning among reactive nitrogen species) (van Geffen et al., 2022; Shah et al., 2023; Silvern et al., 2018, 2019)."

Lines 228-230: "These discrepancies reflect vertical distribution heterogeneity and diurnal variation, as column comparisons are for 1:30 p.m. local time, while surface comparisons include all times of the day." Not necessarily, see my previous comment.

We removed this sentence and revised accordingly (see reply to previous comment directly above).

Lines 244-245: "...changes in NO2 VCDTrop and CO VCDTotal largely mirror the surface patterns (Fig. 3b)." Fig. 3b shows NO2 and HCHO columns, not CO. Were CO columns supposed to be in that figure? Even if not, it would make sense to include them.

Thank you for pointing this out. We have now added the model-simulated total column CO to both Fig. 3b and Fig. S1b for completeness and revised the text accordingly.

Fig. 1 caption: "July mean nitric oxide (NO) emissions from the adjusted 2017 National Emissions Inventory..." What does "adjusted" mean here? Is that just the weekday shift and regridding? If so, please call that "adjusted NEI" in sect. 2.1 so that readers know this is how you will refer to it elsewhere.

Yes, by "adjusted NEI" we mean the 2017 NEI that has been shifted to align with the 2018 weekday-weekend calendar and regridded (mass-conserving) to the model grid, as now defined and clarified in Section 2.1 lines 140-143: "The 2017 hourly NEI data are then shifted by one calendar day so that the weekday-weekend cycle matches the 2018 calendar, then re-gridded (mass-conserving) to the unstructured ne0CONUSne30x8 horizontal resolution using NCAR-developed tools (National Center for Atmospheric Research, 2022a, b). All subsequent references to NEI in this study refer to this adjusted product unless otherwise noted." We have also updated the Fig. 1 caption to explicitly refer to the "the 2017 National Emissions Inventory adjusted to 2018".

Fig. 2: If most of the SLAMS data are in the six urban areas shown in Fig. 1, is it really fair to label Fig. 2a with regions, or should panel a be labeled with the city names instead? This is where the map of SLAMS stations per grid cell would help.

Thank you for raising this point. While major urban areas contribute substantially to the available SLAMS data, many observations are also located outside the labeled city centers, for example, along the U.S. West Coast beyond the Los Angeles area (including the San Francisco region) and in dispersed locations across the Mountain region that are not centered on Denver. For this reason, labeling Fig. 2 solely by city names would not accurately represent the spatial sampling of the observations. Instead, we have added Fig. S8, which shows the full distribution of SLAMS monitors and provides city-level monitor counts to better contextualize the regional statistics.

Lines 310-313: "These results highlight the need for improved representation of diurnal cycles, particularly boundary-layer dynamics that strongly shape near-surface concentrations (Adams et al., 2023), as biases in daily range and peak timing increase uncertainty and limit the application of model simulations, particularly in the era of high-temporal-resolution observations from satellites and ground-based networks." I agree that PBL dynamics are a likely explanation, but they aren't the only one. Has there been any evaluation of the PBL height in the MUSICA dynamical core? (It's not clear to me if the Adams paper evaluated this.) Another direct source of error would be an incorrect diurnal profile in the NEI emissions - how is the diurnal cycle of emissions calculated, and how well has it been validated? What about representativeness error - Fig. 4 seems to be comparing a whole grid cell to a single SLAMS station in each

case, do you know whether there are highly local effects that could be in the observation but not in the model grid cell?

Thank you for this thoughtful and detailed comment. We agree that boundary-layer dynamics are not the only potential contributor to the diurnal discrepancies discussed in this section, and that additional sources of uncertainty warrant explicit discussion. In response, we have expanded the manuscript text below to clarify the role of boundary-layer processes in MUSICAv0, summarize what is known about their evaluation in the CESM framework, and explicitly acknowledge other contributing factors, including uncertainties in emission temporal profiles and representativeness differences between model grid cells and point observations.

See lines 327-339: "It is challenging for models to accurately represent the complex diurnal processes that shape near-surface concentrations, including boundary-layer evolution (Adams et al., 2023). Our MUSICAv0 simulations use the standard CAM6 physics configuration, including the unified Cloud Layers Unified By Binormals (CLUBB) planetary boundary layer parameterization (Bogenschutz et al., 2018; Danabasoglu et al., 2020), which has been evaluated within the CESM2/CAM6-chem framework. While the scheme generally captures the broad diurnal and seasonal structure of boundary-layer behavior, earlier work reveals overly strong nocturnal mixing and errors in the timing of the morning transition, with the direction and magnitude of these biases varying across regions and seasons (Holtslag et al., 2013; Schwantes et al., 2022; Stjern et al., 2023). Such biases may influence the simulated diurnal amplitude and timing of near-surface concentrations and their relationship to the underlying temporal profile of emissions, such as nighttime accumulation and the timing of morning concentration peaks. Errors may also arise from the NEI diurnal profiles applied in this study, which are sector-specific and derived from activity-based temporal allocation methods whose accuracy at local scales remains uncertain. In addition, comparisons between model grid-cell means and individual SLAMS sites inevitably involve representativeness differences that may contribute to mismatches."

Lines 338-339: "West-East Contrasts in Surface Pollutant Responses" - this sentence fragment looks like it was supposed to be a subsection title?

We thank the reviewer for catching this. This phrase was intended as a subsection title and has now been formatted accordingly.

Lines 376-378: "Collectively, these processes shorten the NO2 lifetime against deposition in the eastern CONUS, contributing to greater NO2 accumulation in the western CONUS." Need to be careful about over-

generalizing here, and differentiate between drivers that are consistent month to month or year to year (e.g., average VOC concentrations) and ones that vary with meteorology. Was there potentially some kind of blocking high or stagnation event during this time period that might be a factor in this west/east dipole?

Interesting point about the west/east dipole. In response to this comment, we evaluated daily stagnation using the meteorological criteria of Wang and Angell (1999) and find that stagnation conditions are rare during July 2018 and do not exhibit a persistent west-east contrast, supporting our original interpretation. We also softened our language to avoid over-generalizing.

The revised text in lines 381-386 reads: "This contrast in NO_x and O_3 concentration responses partly reflect systematic differences in meteorological conditions across the CONUS during July 2018 (Fig. 5c). In the eastern CONUS, shallower daytime PBLH can limit vertical mixing while enhancing both photochemical reactions and surface deposition (Y. Wu et al., 2024). Together with higher humidity in the eastern CONUS, these processes may contribute to a shorter NO_2 lifetime in the eastern CONUS during July 2018. We do not find evidence for a persistent east-west dipole in air stagnation, defined according to the meteorological criteria of Wang and Angell (1999)."

Fig. 5 caption: please confirm that the PBLH is given in meters above ground level, not meters above sea level. I assume so from the magnitude, but it follows the topography well enough that this would be good to clarify.

Yes, we confirm that in CESM, PBLH represents the diagnosed planetary boundary layer height above the ground surface (meters above ground level). This definition aligns with the meteorological and modeling interpretation of PBLH as the height of the lowest atmospheric layer influenced by surface processes, and the magnitude and diurnal cycle in our output are consistent with expected AGL values. We now state this explicitly in the main text "planetary boundary layer height (PBLH; m above ground level)" and the Figure 5 caption "(PBLH; m above ground level)".

Lines 433-434: "These MUSICAv0 simulations indicate that urban areas in both the western and eastern CONUS are generally NOx-saturated, with surface O3 increasing under reduced NO emissions (Fig. 3)." It's difficult to see urban/rural differences in Fig. 3; the O3 panels show decreases everywhere to my eye. Either zoomed-in maps on urban and surrounding areas or box plots that show the change in O3 for urban and rural grid cells separately would make this point clearer.

Thank you for this helpful suggestion. We agree that the urban-rural contrast in the O_3 response is not easily discernible at the CONUS scale shown in Fig. 3. We have now added zoomed-in panels for two representative urban regions, Los Angeles and New York City, to the new Fig. 4. These panels show surface NO_2 and O_3 differences for NEI_hourly_NO − NEI_daily_NO and NEI_monthly_m30anthroNO − NEI_monthly. We also added panels c-d for CAMS_m30anthro – BASE and CAMS_m30bio − BASE in Fig. S1 that zoom in to the two cities. The regional views show that urban cores exhibit O_3 increases under reduced NO emissions, consistent with NO_x-saturated chemistry.

Lines 437-438: "...suggesting that these urban centers likely remain within a NOx-saturated regime but are approaching the transition toward NOx sensitivity (Fig. 7)." It's not clear to me how you come to this conclusion from Fig. 7. Are you making this argument because in LA the 30% reduction moves the high NOx points over toward the peak O3 production, so if further reductions occur, they should move into the NOx-limited regime? If so, please so that explicitly and explain how that applies to NYC, as the high-NOx points for NYC largely seem to stay at the same P(O3) level.

First, we revised Fig. 7 to improve visual clarity, particularly for the lowest concentration bins where most of the data points fall. The time-of-day progression from morning, when NO_x concentrations are highest, to mid-day and afternoon, is also now much clearer. We revised the text to explicitly state that our interpretation is based on shifts in the NO_x concentrations associated with peak $P(O_3)$, rather than changes in peak $P(O_3)$ magnitude itself. We also add references to the subplots for clarity.

The revised text in lines 449-452 now reads: "Compared to NEI_monthly (Fig. 7a, d), reducing anthropogenic NO emissions by 30% (NEI_monthly_m30anthroNO; Fig. 7c, f) shifts the NO_x concentration bins associated with high $P(O_3)$ toward lower NO_x in both Los Angeles and New York City. For example, in Los Angeles, the highest-NO_x conditions (> 30 ppb) shift into lower NO_x bins (< 30 ppb), bringing them closer to the NO_x range associated with maximum $P(O_3)$."

Fig 7: the green and blue are hard to distinguish; using different markers as well as colors may help.

Thank you for this helpful suggestion. We now use both distinct marker shapes and colors in Fig. 7 to differentiate the three daytime windows. In addition, we revised Fig. 7 to use equal spacing per NO_x concentration bin along the NO_x axis and to visualize sample size (through box width and annotated counts), ensuring that bins at low NO_x are clearly visible and not obscured by the binning scheme. We also increased the transparency of afternoon points to reduce overlap and improve boxplot visibility.

Lines 454-455: "Using hourly NO emissions shifts the distribution and peak of P(O3) toward higher NOx concentrations in Los Angeles and toward slightly lower NOx in New York City, with more pronounced changes in Los Angeles (Fig. 7)." It's not clear from Fig. 7 that peak P(O3) does shift to higher NOx in LA - to me, it looks like the difference between the top left and top middle panels is that the 15-20 ppb NOx box actually has its P(O3) decrease slightly, with peak P(O3) still around the 5-10 ppb NOx range.

Thanks for this point. We have revised the following text in lines 454-457: "In contrast to the 30% NO reduction case (NEI_monthly_m30anthroNO), using hourly NO emissions (NEI_hourly_NO) increases the density of values in the highest NO_x bins in Los Angeles, while maintaining similar $P(O_3)$ values in those bins (Fig. 7b). This extends the high-NO_x tail of the $P(O_3)$-NO_x distribution into more strongly NO_x-saturated conditions (Fig. 7b)."

Lines 456-457: "...the NEI_hourly_NO simulation shifts the P(O3) distribution in Los Angeles toward higher NOx levels, indicating more strongly NOx-saturated conditions." This statement needs a bit more clarity: it looks like you're specifically talking about the long tail at the right end of the distribution, i.e., the hourly emission run has more points in the highest NOx concentration bins with correspondingly low P(O3). Since the immediately previous sentence was discussing peak P(O3), not the whole P(O3) distribution, it took me a few reads to catch the shift from discussing peak P(O3) to its general distribution.

Thank you again for helping us clarify the wording. We revised the text in 454-457 included above.

Lines 459-460: "...the morning reduction dominates the monthly mean O3 change (Fig. 6b)." Please clarify - do you mean that the decrease between about 9 and 14 local for LA is big enough that it makes the overall change in the monthly O3 negative?

Yes, that is our intended meaning. We have revised the text to clarify this point: "The enhanced high-NO_x tail in the NEI_hourly_NO simulation reflects larger morning NO_x concentrations, when the chemical regime is most NO_x-saturated and O_3 production is suppressed. Although conditions become more NO_x-sensitive later in the afternoon (Fig. 7b), the larger O_3 decrease during late morning to early afternoon (~9-14 local time; Fig. 6b) dominates the net monthly mean O_3 concentration response (Fig. 4a)."

Lines 485-486: "Incorporating hourly NO emissions amplifies the spatial gradients in simulated surface NO2 concentrations between urban and rural areas..." As mentioned previously (in the comment for lines 433-434), the current set of figures don't provide a clear visualization of urban/rural differences.

Thank you for this comment. As discussed in response to the earlier comment on Lines 433-434, we have added zoomed-in panels for Los Angeles and New York City as the new Fig. 4.

Lines 505-507: "Our results demonstrate that hourly variations in emissions can produce nontrivial impacts on NO2 and O3 concentrations even without changes in the total emission magnitude. Neglecting such variability may introduce biases when interpreting monthly mean concentrations from once-daily satellite overpasses..." There is some missing model description and analysis that weakens this claim. First, I did not see where it was shown or started that the hourly NEI emissions emitted the same total mass of pollutants as the monthly average one, which is important for the reader to understand that the differences in the model outputs are solely due to emission timing. Second, a worked example of how the timing of emissions affects the relationship between total emissions and early afternoon columns would help make the point: something simple like the analysis in Sect. 3.2.4 of Lamsal et al. (2014, https://acp.copernicus.org/articles/14/11587/2014/), but using the NEI_hourly model simulation as the "observed" columns and the NEI_monthly as the "modeled" values to show the error in retrieved emissions would help solidify that argument.

We thank the reviewer for this helpful comment. We have revised the manuscript and added new analyses to explicitly address both points.

First, we clarify that July total anthropogenic emissions are conserved at the grid-cell level in lines 137-140: "We then conduct a series of simulations that differ only in the temporal resolution of NEI emissions (monthly, daily, or hourly) to isolate the model response to emission timing while holding July total monthly emissions constant."

Second, to address the reviewer's request for a worked example, we added a new analysis following the proportional scaling framework of Lamsal et al. (2014). Our new analysis provides a simple estimate of how emission timing alone can affect the relationship between total emissions and NO_2 columns, as suggested by the reviewer. These results are described in detail in Text S6, summarized in a new subsection (Section 4.3), and illustrated in new Figure 8, with regional- and city-scale statistics reported in Table S6.

--- Response to Reviewer 2 ---
In "Contrasting Air Pollution Responses to Hourly Varying Anthropogenic NOx Emissions in the Contiguous United States", Tao et al. address a timely scientific question regarding how the potential lack of temporal variability of emissions in global chemical transport models shapes the resulting concentrations or related metrics, such as vertical column densities or ozone production rates. The study

is carefully designed to test the hypothesis that monthly, daily, or, especially, hourly variation in emissions over the U. S. alter the diurnal cycle and spatial distributions of the emitted species and other species formed from them. This work is timely especially on account of TEMPO, which will provide higher temporal resolution in remotely-sensed column densities that are frequently used for revising estimates of emissions. Doing so without accounting for hourly variation in emissions could embed errors in the emissions refinements that could have been addressed with existing estimates.

This manuscript describes reasonable differences for daily and monthly variation in emissions with reasonable explanations. The authors support their analysis throughout the work with excellent contextualization with other literature. The amount of explanation in the supplemental material is suitable. The changes in concentrations observed when incorporating hourly emissions have unexpected differences from the western to the eastern portion of the U.S., and the authors introduce regional and local data analysis to demonstrate and explain the potential causes.

We thank the reviewer for this positive assessment.

My primary concern regards the implementation of hourly emissions. The supplemental information describes that a CMAQ-ready version of the NEI 2017 inventory was transformed for integration into MUSICAv0, which seems like an efficient approach. Yet, the unexpected patterns of increase in the west and decrease in the east made me curious.

We thank the reviewer for raising this important concern regarding the implementation of hourly emissions and for carefully examining the spatial and temporal patterns in Figures 3-7 of the submitted manuscript. When we first found this result, our initial thought was also that there must be an error in the implementation of the emissions, so we fully agree that these findings merit careful examination.

We added more description of our treatment of the hourly NEI emissions, which are derived based on the approach described in the main text section 2.1: "We first process the 2017 NEI (U.S. Environmental Protection Agency, 2022), the most recent pre-COVID inventory available at the time of this study, to hourly resolution on a ~0.1° × 0.1° grid over the CONUS (Text S3) using sector-specific diurnal profiles to capture within-day as well as day-to-day variations. The 2017 hourly NEI data are then shifted by one calendar day so that the weekday-weekend cycle matches the 2018 calendar, then re-gridded (mass-conserving) to the unstructured ne0CONUSne30x8 horizontal resolution using NCAR-developed tools (National Center for Atmospheric Research, 2022a, b)."

In the responses below, we provide additional detail to address the reviewer's questions regarding time-zone handling in the hourly emissions implementation and the interpretation of the relevant figures.

Figure 4a shows the change of concentration of CO (the species shown that is most like an emissions tracer) and, to a lesser extent NO2, when changing from monthly to hourly. An abrupt shift from negative to positive changes is evident with the demarcation of time zone from Eastern to Central. The shifts from Central to Mountain and Mountain to Pacific are not as prominent, but the deepening of the increase is suggestive of the offset in time contributing to the differences. Further investigation of other figures to understand whether a time shift in hourly emissions associated with time zone may be inducing these results was challenging to reconcile.

We thank the reviewer for thinking carefully about these features as we also initially suspected an error in the local time application of the emissions when we examined the map showing surface concentration changes (referred to as Figure 4a by the reviewer but more likely corresponds to Figure 3a in the submitted manuscript).

With a fresh look, we once again revisited the implementation of hourly emissions and confirmed that emissions are read by the model in hourly UTC time, as required by the MUSICA framework, and that the prescribed diurnal variability at pixel level is correctly preserved. We added a new figure (Fig. S6), discussion in Text S3, and the following to lines 149-151: "We verified that hourly NEI emissions are read by MUSICAv0 in UTC time with the prescribed diurnal and weekday-weekend variability preserved in local time at the grid-cell level (Text S3; Figs. S6 and S7)."

In addition, the spatial response in Figure 3a does not cleanly follow time-zone divisions in a uniform manner: transitions across the Central-Mountain and Mountain-Pacific boundaries are weaker and spatially heterogeneous, and substantial variability exists within individual time zones. Instead, the concentration changes closely mirror large-scale gradients in meteorological conditions, most notably surface temperature, planetary boundary layer height, and surface relative humidity (Figure 5c), supporting the interpretation that the observed west-east contrast reflects regional differences in atmospheric sensitivity to emission timing.

For instance, if Figure 6a is at local time for LA and for NYC, it suggests that at midnight the hourly emissions of NO are typically above the monthly average in LA whereas in NYC the emissions are below the monthly average, which would be more reasonable to expect. In Figure 6b, the diel plots, which are defined as being in local time, do not have this offset in time.

Thank you for carefully examining Figures 6 and 7, and especially for catching a plotting error! The apparent enhancement of NO emissions near local midnight in Figure 6a was indeed caused by a plotting script issue related to time-coordinate handling, rather than an error in the emissions themselves. In the submitted Figure 6a only, the hourly NO emissions were converted from UTC to local time for interpretation, but the converted local-time coordinate was not reassigned to the array storing the emissions prior to plotting. Consequently, the emissions retained their original UTC timestamps for both cities while being displayed on a local-time x-axis, producing a systematic temporal shift that made emissions appear to be far too high at local midnight on July 1. We have corrected this error (see the corrected Figure 6a in revised manuscript and below). This issue was limited to the time axis in this specific panel and, fortunately, does not affect the emissions data themselves or the conclusions of the study.
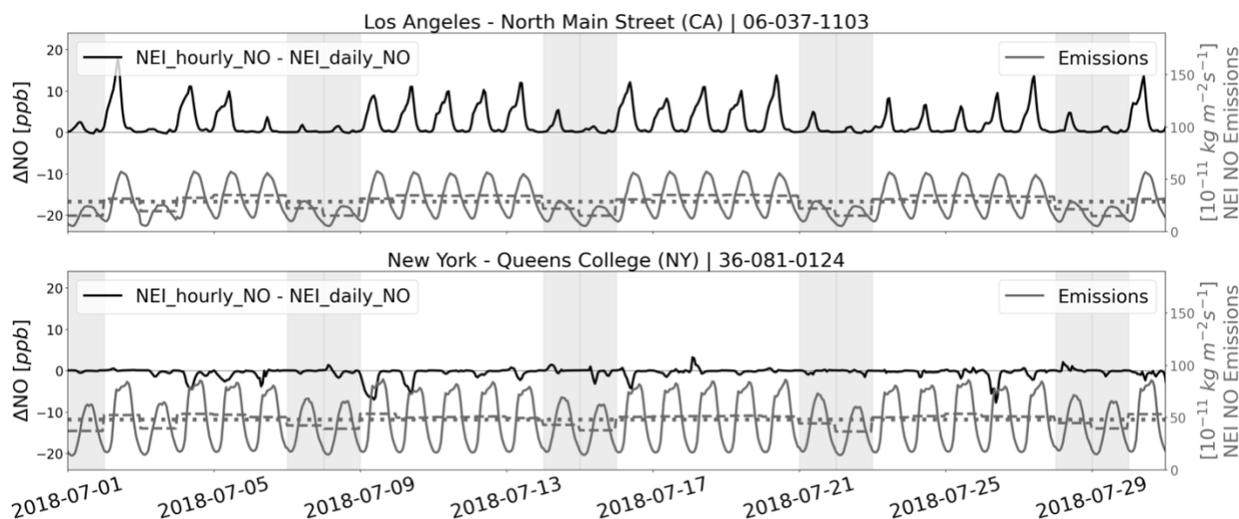

Fig. 6a. "Hourly time series of July differences in surface NO concentrations between NEI_hourly_NO and NEI_daily_NO (black solid lines), alongside hourly (solid gray), daily mean (dashed), and monthly mean (dotted) NO emissions. Weekend days are shaded in gray."

Additionally, the shift in time of the productivity of ozone for LA but not NYC in Figure 7 contributes to the idea that an offset of three hours of emissions changes may be shaping these results. These patterns are indicative of the time zone not being handled correctly when hourly emissions were implemented.

We thank the reviewer for this thoughtful concern. Fig. 7 shows July distributions of the model-diagnosed ozone production rate, $P(O_3)$ (ppb hr^{-1}), versus NO_x concentrations, with points colored by local-time windows (9-11 AM, 12-2 PM, 3-5 PM). In generating Fig. 7, we did reassign the model output time coordinate from UTC to city-specific local time prior to filtering by hour-of-day in local time. Therefore,

the time-zone handling in Figure 7 is consistent across cities and does not contain the UTC-to-local mismatch that affected the originally submitted Figure 6a.

In the previous version of Fig. 7, box widths reflected the NO_x bin ranges, which compressed boxes at low NO_x values and visually overemphasized higher-NO_x bins. We have revised Fig. 7 to use equal spacing along the NO_x axis, making low-NO_x boxes easier to see, and to communicate the sample size for each bin through box width and annotated counts along the top of the panel, with the intent of providing a clearer and more balanced visualization across bins.

I have hoped that this explanation for the temporal shifts is not the best one since the authors carefully show and explain the potential physical and chemical reasons for differences in these outcomes across the country. If the authors demonstrate the accuracy of the hourly emissions implementation, I would be excited to revisit this substantial contribution that demonstrates unexpectedly different chemical responses between the western and eastern U.S. to the incorporation of hourly emissions.

Our best interpretation of the diagnostics above is that the hourly emissions are implemented in MUSICAv0 as intended, with the emissions read in UTC, diurnal variability preserved in each region's local time, and no systematic offsets associated with time-zone boundaries. We do share and appreciate the reviewer's concern for the need to be absolutely sure that this interpretation is sound. We hope that the additional work motivated by the reviewer's comments further strengthens the study's central conclusion and underscores the value of resolving hourly emissions when interpreting regional air-quality responses.