Dear Editor,

We are submitting our revised manuscript, '*Processes driving the regional sensitivities of summertime $PM_{2.5}$ to temperature across the US: New insights from model simulations*', to *Atmospheric Chemistry and Physics*. In the revised version, we have expanded the discussion and conducted additional analyses to address the reviewers' comments comprehensively. Specifically, we have verified methodological robustness by including new sensitivity tests, including 10-year and 3-year rolling window analysis. Additionally, we incorporated a decomposition analysis for nitrate to provide a mechanistic explanation for its negative sensitivity, distinguishing between emission drivers and thermodynamic partitioning. We have also improved the readability of the figures, such as employing broken axes to handle outliers, and conducted a thorough language review of the full manuscript.

We thank the reviewers for the helpful comments for improving the manuscript. Responses to the individual comments are given below. Reviewer comments are in **bold**. Author responses are in plain text. Modifications to the manuscript are in *italics*. Line numbers in the responses correspond to those in the clean version of the revised manuscript.

I believe these enhancements have strengthened the scientific merit of our work, making it suitable to be considered for publication in *Atmospheric Chemistry and Physics.*

Sincerely,

Pengfei Liu

School of Earth and Atmospheric Sciences

Georgia Institute of Technology

311 Ferst Drive, Atlanta, GA 30332

RC#1

In this manuscript, the authors build upon previous literature examining relationships between air pollution and climate variability, focusing here on ways in which temperature affects PM2.5 in the United States. Using a wide ranging set of chemical transport model simulations, they attempt to identify key connections and mechanisms linking temperature to PM2.5 through a rolling window of gridded sensitivity calculations. On the whole I find many aspects of this study to be thoughtfully composed and presented, and I see many strengths and valuable elements. At the same time, I am also puzzled by several key elements and decisions, including fundamental model setup and case parameter choices. While I can see core elements of this study that should become valuable additions to the literature on publication, I have some major concerns and suggestions for improvement that I believe are necessary to tighten up the comparisons being made and resolve issues in analysis.

[RC 1.1] As one large and general criticism, I find the full set of model simulations, as described in Table 1, to be overly varied and confusingly named. The so-called BASE case has almost nothing in common with any of the other simulations. It uses a significantly older version of model code, with major differences across the board in terms of functionality and operation. Emissions, SOA scheme, and even the range of years modeled are either unique to this case or shared by only one other, which in my mind really stretches the assumed definition of a “BASE” case. MOD in this list does not represent a “modified” version of the BASE case, but rather an entirely new model version with almost entirely new model inputs. Considering the age of the code used in the BASE case (currently over 7 years old representing a gap of three major code releases and numerous smaller fixes and updates relevant to this study topic), I do not see much value in including the BASE simulation at all, unless a more systematic review of all updated components is performed. Since this study is more focused on chemistry/climate sensitivities, and less focused on cataloguing individual modeling improvements that have been made over the past decade, I suggest that the currently named BASE case be removed from the study in the sake of clarity and consistency, and all cases be renamed to better reflect their true roles and points of comparison.

[AC 1.1] We thank the reviewer for this valuable feedback regarding the clarity of our experimental design. We agree that the original naming method was confusing and may have obscured the distinct role of the different simulations.

In response, we have renamed and reorganized all simulation cases in the revised manuscript to better distinguish 1) the primary analysis cases; 2) resolution and SOA-scheme sensitivity tests, and 3) a historical reference simulation. Specifically:

- **"BASE" → "REF_2019"**: This simulation has been retained but renamed to explicitly indicate that it is a historical "Reference" case based on an older model version (v11-02c). This case is not intended to serve as a baseline for sensitivity comparison, rather, it provides continuity with prior studies that relied on this dataset, including the development of machine-learning-derived PM2.5 datasets used in many health studies. Retaining this reference case allows us to illustrate the progress in model performance compared to the historical standard used in previous literature (Silvern et al., 2019).

- **"MOD" → "MAIN_HighRes":** Our primary analysis using the updated model (v12.9.3) with GFED4 emissions and the coating effect in isoprene SOA formation.
- **"MOD_BC" → "MAIN_LowRes"**: Identical to MAIN_HighRes but conducted at a coarser resolution (4°×5°) to assess the resolution sensitivity and serve as a reference for MAIN_LowRes_SimpSOA.
- **"SIM_BC" → "MAIN_LowRes_SimpSOA"**: A low-resolution sensitivity case using the "Simple SOA" scheme.

We have updated Table 1 and all relevant text and figures to reflect these changes. We emphasize that this study is focused on chemistry and climate sensitivities within a modern modeling framework, rather than on cataloguing incremental model development over multiple generations. We believe that the revised structure and terminology substantially improve clarity while preserving the broader context needed for comparison with prior work.

[RC 1.2] On a similar note, I think the SIM_BC case (4x5 horizontal resolution global simulation with Simple SOA) is not currently effective as a sensitivity test due to having too many changes relative to the MOD case. It shares a coarse resolution with MOD_BC, but has a unique SOA scheme and (confusingly) a shorter run length. Again, there are too many unique elements to make this an effective and meaningful comparison against the high resolution MOD case that is actually at the heart of the study. If the domain run time were lengthened it could serve as a reasonable comparison against MOD_BC, but grouped together with all cases as it often is in manuscript tables and figures it is somewhat misleading, since it is not clear whether differences against MOD are due to resolution, SOA scheme, temporal domain differences, or some combination of the three.

[AC 1.2] The MAIN_LowRes_SimpSOA (formerly SIM_BC) case was included specifically and exclusively to isolate the impact of SOA scheme complexity on temperature sensitivity. This case differs from the MAIN_LowRes (formerly MOD_BC) case only in its SOA scheme. Due to the massive computation cost, conducting multiple-year high resolution simulations for sensitivity runs were not feasible. This case is not intended to be compared directly with the high resolution MAIN_HighRes simulation, and we have revised the manuscript to make this distinction explicit.

Regarding the reviewer's concern about differing run lengths, we emphasize that all quantitative comparisons between MAIN_LowRes and MAIN_LowRes_SimpSOA are restricted to the overlapping time period, despite the longer simulation duration available for the MAIN_LowRes run. The difference in nominal run length reflects computational constraints rather than a methodological distinction and does not affect any reported sensitivity results. We also note that all other comparisons in the manuscript were conducted using identical time periods. For instance, the results in Figure 1—comprising ML-based data, observations, and the four GEOS-Chem cases—are based strictly on data from 2000–2016, as noted in the figure caption:

(Line 279-281) *Figure 1: Spatial distribution and regionally aggregated temperature sensitivity of summertime PM2.5 and its major components during 2000–2016, derived from ground-based observations, machine-learning (ML) modeled datasets, and GEOS-Chem simulations.*

Similarly, in Supplementary Figure 3, although the GEOS-Chem simulation years vary across cases, the observational results were calculated using the corresponding temporal period for each specific simulation to ensure internal consistency.

We have revised the text to clarify this methodology:

(Line 145-148) *Figure 1 compared the temperature sensitivity derived from GEOS-Chem outputs (Figure 1 a2-a5) with diagnosis from ground-based observation and ML-modeled data for 2000-2016 (Figure 1 a1). To ensure comparability, all of these results are calculated based on 2000-2016 data, although more data are available for both observations and GEOS-Chem simulations.*

(Line 218-220) *Table 1 provides a summary of all simulation cases conducted in this study. Although total simulation durations vary across cases, all comparisons with ML-based results and observations are restricted to their overlapping period to ensure temporal consistency.*

[RC 1.3] I also have concerns related to overall model performance, as shown in Supplementary Figure 4. Considering the proposed importance of OA sensitivities and the mechanisms underlying its formation and fate, the massive overprediction in later years strikes me as especially problematic. The fact that this overprediction appears to have such a strong impact on the sensitivities shown in Figure 2 is even more concerning, and deserves more analysis and discussion. An overabundance of fire emissions from GFED4 is offered as a potential reason for these effects, but this issue and its downstream impacts on resulting sensitivity analyses are far too brief, to my eye. Even assuming that PM overpredictions are being caused by overestimated primary emissions from GFED, what is the reason for the massive and regionally varying impacts on sensitivities (Figure 2, b2-b5)? What does it show about the robustness of this methodology, if one year of overestimated fire emissions can drive such wild swings in calculated sensitivities across over one third of the modeled years, for some regions? And perhaps most importantly, if 2021 fire emissions are indeed so problematic, why include that year at all? Considering how much of the sensitivity time series is apparently dominated or at least heavily impacted by this issue, I am not comfortable putting much weight on final numbers until either the issue is resolved or the impacted years are removed.

[AC 1.3] We thank the reviewer for this critical observation. We agree that the OA overprediction in the later part of the record (with particularly large biases in some years such as 2021) merits additional analysis, especially given its large influences on the temperature sensitivity shown in Figure 2. We have expanded the manuscript discussion to clarify that the late-period model bias reflects a combination of factors, including high primary emissions in the GFED4 inventory, potential under-sampling in ground-based observations near fire zones, and the inherent difficulty for CTMs to resolve extreme episodic events.

(Line 245-254) *The higher RMSE for OA simulations in certain regions can be attributed to overestimated concentrations in 2021. This discrepancy is likely driven by exceptionally high wildfire emissions within the GFED4 inventory for certain years, such as 2021 (Supplementary Figure 4). Because wildfire-driven OA is strongly temperature dependent and spatially heterogeneous, these overestimations can disproportionately affect regression-based temperature sensitivity diagnostics. Additionally, potential under-sampling by ground-based monitors near fire zones may lead to an underestimation of regional mean concentrations in observational datasets. Furthermore, current CTMs face inherent challenges in capturing the non-linear chemistry and rapid transport associated with extreme fire events (Qiu et al., 2024). Despite these uncertainties, data from high-fire years were retained in the full record, as they represent the growing influence of extreme wildfire events on* *air quality in a warming climate and reflect an important stress test for current modelling frameworks.*

Regarding the temperature sensitivity diagnosis, mathematically, the 2021 anomaly only impacts the final two 5-year windows (2017–2021 and 2018–2022) in Figure 2. The increasing sensitivity trend in the Western US and the decreasing trend in the Eastern US are clearly observable in the preceding windows (2000–2020), which are entirely independent of the 2021 fire emissions. While shortening the time window makes the sensitivity diagnosis more susceptible to single-year anomalies, it does not alter the general changing patterns. To further test the robustness of our methodology, we conducted a 10-year rolling window analysis (Supplementary Figure 13). The 10-year analysis shows smooth, stable trends that align with our decadal findings (Yin et al., 2025). The "wild swings" observed in the 5-year analysis are effectively smoothed out when the sample size is increased.

(Line 363-377) *In the western US, observed $PM_{2.5}$ sensitivity shows a substantial increase after 2014. The MAIN_HighRes case shows good consistency with observed data from 2000–2016, while overestimating $PM_{2.5}$ sensitivity in the western US from 2016–2022 (Figure 2 a4, b4), which could also be attributed to the high OA concentrations during this period (Supplementary Figure 4). These results highlight that short-term temperature sensitivity diagnoses are inherently susceptible to single-year emission or concentration anomalies, as shortening the analysis window increases the leverage of such outliers. This underscores the critical need for accurate representation of temperature-dependent processes in emission inventories—particularly for primary pollutants—within sensitivity simulations. For OA, this specifically requires a robust representation of temperature-sensitive burned area during emission estimation. Furthermore, we conducted a 10-year rolling window analysis (Supplementary Figure 13) to assess long-term stability. As expected, the longer time window smooths the interannual variability, demonstrating that the underlying signals of anthropogenic mitigation (in the East) and climate penalty intensification (in the West) are robust and persist despite the noise introduced by the 2021 fire anomaly. Consequently, the substantial fluctuations observed in the 5-year analysis reflect the high leverage that extreme events exert on short-term regressions rather than a lack of robustness in the underlying model framework. As this study aims to move beyond the long-term patterns to focus on interannual variability and its specific drivers, retaining the 2021 data can also help identify and quantify the role of wildfire emissions in temperature sensitivity simulations.*

[RC 1.4] I question the decision and justification for running Complex SOA scheme model simulations without the semivolatile POA option. In defense of this choice, the authors argue that “non-volatile POA treatment more accurately reproduces the low-troposphere POA profile compared to the semi-volatile approach”. However, the cited paper (Pai et al., 2020) appears to offer the exact opposite recommendation, saying that “a semi-volatile treatment of POA is superior to a non-volatile treatment” and “we recommend that POA be modeled as semi-volatile”. The authors here later claim that “since oxidized POA is not included in the SOA mass, disabling the semi-volatile POA scheme does not impact SOA simulation results and helps reduce computational costs”, but again I question this justification, considering the frequent use of overall PM2.5 and total OA in study figures. Turning this option off should be supported by more extensive testing to confirm that key sensitivity metrics are indeed not affected by its omission.

[AC 1.4] We thank the reviewer for carefully examining our model configuration and the interpretation of (Pai et al., 2020). Pai et al. (2020) indeed recommend using semi-volatile treatment for POA modelling. At the same time, Pai et al. (2020) also document that the semi-volatile POA treatment in GEOS-Chem leads to longer atmospheric residence times and enhanced POA concentrations in the lower troposphere compared to the non-volatile treatment: "When compared in aggregate, the simple scheme is less biased in the lower troposphere, while the complex scheme is less biased in the upper troposphere (Figs. 8, S3)." Using the non-volatile POA scheme align with our focus on surface concentrations.

In this study, our interest is in understanding the temperature sensitivity of PM2.5 and its components. Our process-based diagnostics demonstrate that POA temperature sensitivity is overwhelmingly controlled by the temperature dependence of emissions, primarily wildfire emissions (Supplementary Figure 18). Because oxidized POA is not included in the SOA mass in GEOS-Chem, enabling or disabling the semi-volatile POA option does not directly affect SOA formation pathways, which dominate the modeled OA temperature sensitivity in the eastern US.

We acknowledge the reviewer's concern that disabling semi-volatile POA could influence total OA and PM2.5 sensitivity. However, our results indicate that the dominant drivers of PM2.5 temperature sensitivity—namely, isoprene-derived SOA production in the eastern US and temperature-sensitive primary emissions in the western US—are robust to this modeling choice. In particular, the relative contributions of POA, SOA, and sulfate to PM2.5 sensitivity, as well as the spatial and temporal patterns discussed in Sections 3.2–3.4, are governed by emission–temperature sensitivity and chemical production rather than by POA volatility treatment.

We have added more discussion on the choice of POA scheme:

(Line 787-797) *An additional source of uncertainty concerns the treatment of POA volatility. Although a semi-volatile treatment of POA is generally recommended for applications focused on reproducing absolute OA burdens (Pai et al., 2020a), it has also been shown to introduce longer atmospheric residence times and enhanced low-tropospheric POA concentrations, reflecting uncertainties in volatility distributions, aging processes, and removal pathways. We acknowledge that disabling the semi-volatile POA option may affect simulated OA mass. However, process-based diagnostics indicate that the temperature sensitivity of POA is dominated by the temperature dependence of primary emissions, particularly wildfire emissions (Supplementary Figure 19). Consequently, the diagnosed POA temperature sensitivity is largely insensitive to the choice of POA volatility treatment. Moreover, oxidized POA is not included in the SOA mass in GEOS-Chem, such that enabling the semi-volatile POA option does not directly affect SOA formation pathways. The dominant drivers of OA and $PM_{2.5}$ temperature sensitivity in the eastern US—namely isoprene-derived SOA production and sulfate-mediated chemical processes—are therefore not expected to be significantly affected by this modeling choice.*

[RC 1.5] In section 3.4, model sensitivities are further explored through a breakdown by model process, with a focus on vertical levels within the PBL. I can understand the rationale for zeroing in on near-surface processes, but I would like to at least see what things look like higher up as well, considering the potential for important chemistry and transport at higher altitudes that can influence surface concentrations. Meaningful chemistry impacts aloft should be categorized as such, not folded into mixing or transport.

[AC 1.5] We appreciate the reviewer's comment regarding the role of chemistry and transport above the PBL and their potential influence on surface PM2.5.

GEOS-Chem budget diagnostics provide mass tendency terms integrated over three predefined vertical domains: the PBL, the troposphere as a whole, and the full atmospheric column. While these diagnostics allow separation of surface-relevant processes (PBL) from column-integrated behavior, they do not provide diagnostics for the free troposphere or upper troposphere alone. As a result, it is not possible within the standard GEOS-Chem framework to explicitly isolate chemical production occurring aloft and separate it from transport or mixing processes that subsequently affect surface concentrations.

Given this constraint, our analysis in Section 3.4 intentionally focuses on the PBL budget, as this domain is most directly relevant to the temperature sensitivity of surface PM2.5, which is the central focus of this study. Although the upward transport and dilution effect is dominant, chemical production occurring above the PBL can influence surface concentrations through downward transport and mixing, and these effects are captured implicitly in the transport and mixing tendency terms diagnosed for the PBL. In this context, chemistry aloft is not neglected, but rather manifests through the pathways by which mass is transferred into the near-surface layer.

To avoid potential confusion, we have clarified the terminology and interpretation in the revised manuscript. Specifically, we now refer to the “chemistry budget” as the production budget and explicitly state that it represents net chemical production within the PBL, while aloft chemical production influences surface PM2.5 indirectly through transport processes rather than being folded into boundary-layer chemistry.

We agree that free-tropospheric chemistry may play a more explicit role for certain species. A more complete vertical decomposition of chemistry and transport would require additional model diagnostics or targeted sensitivity simulations beyond the scope of the present study, which is focused on near-surface PM2.5 temperature sensitivity. We have clarified this point in Section 3.4 to better explain how aloft chemistry is represented indirectly and to avoid the impression that such processes are being explicitly categorized as boundary-layer chemistry:

(Line 425-430) *Additionally, for secondary pollutants, emissions are inherently accounted for in the chemistry budget through precursor concentrations. To avoid misunderstanding, the chemistry budget is referred to as the production budget in the following text. Chemical production occurring above the PBL may influence surface concentrations through downward transport and mixing. Although GEOS-Chem does not provide diagnostics that allow explicit separation of free-tropospheric chemistry from these transport pathways, these effects are captured implicitly in the transport and mixing tendency terms diagnosed for the PBL.*

[RC 1.6] Likewise, while isoprene may show up as the strongest SOA signal in the US (Figure 3), it would be worthwhile to see sensitivities of other SOA species as well. Figures 3 and 4 together appear to leave major gaps in understanding temperature sensitivities outside of the sulfate and isoprene dominated regions to the East, and getting a better handle on how SOA appears to behave elsewhere could be useful.

[AC 1.6] We thank the reviewer for this helpful comment and agree that understanding the temperature sensitivity of SOA species beyond isoprene-derived SOA is important, particularly outside sulfate- and isoprene-dominated regions in the eastern US. We also analyzed the temperature sensitivity of POA and monoterpene-derived SOA (TSOA) and its controlling processes. These results are presented in Figure 6 and Supplementary Figures 18 and 22 and discussed in Sections 3.4 and 3.5. The relevant text reads as follows:

(Line 543-550) *The temperature sensitivity of major processes contributing to POA concentration is shown in Supplementary Figure 19 a1-a5. Unlike the other species considered here, POA is directly emitted from fire events, making the emission process the dominant factor driving POA sensitivity. Additionally, the temperature sensitivity of transport significantly influences the interannual variation in POA sensitivity, particularly in the Southeast and Northeast US. TSOA, formed from monoterpene oxidation, is an important component of biogenic SOA. We find that the temperature sensitivity of TSOA is approximately an order of magnitude lower than that of isoprene SOA, typically ranging from 0 to 0.2* $\mu g\ m^{-3}\ {}^{\circ}C^{-1}$ *in most regions in the CONUS (Supplementary Figure 14). As with ISOAAQ, chemical production dominates the variability in TSOA sensitivity, whereas transport sensitivity contributes to the variations in certain years (Supplementary Figure 19 b1–b5).*

(Line 681-693) *The temperature sensitivity of SOA formed by monoterpene oxidation makes a non-negligible contribution to the overall SOA sensitivity, especially in the Southeast US (Figure 1). To better understand this contribution, we decomposed the temperature sensitivity of TSOA into components mediated by TSOG (semi-volatile oxidation in the gas phase), which reflects the gas-particle phase partitioning, and by concentrations of monoterpenes,* NO_x*, and ·OH (Figure 6 c1–c5 and Supplementary Figure 23). Our results reveal that phase partitioning plays a significant role in determining the magnitude of TSOA temperature sensitivity, particularly in the Southeast and Northeast US. Additionally, the temperature sensitivities of ·OH and* NO_x *contribute to both the temporal pattern and interannual variations in TSOA sensitivity. This phenomenon is especially pronounced in the Western US, where* NO_x *concentrations are heavily influenced by wildfire emissions (Campbell et al., 2022), which have shown a consistent increase in recent years. The rising* NO_x *temperature sensitivity, combined with the temperature sensitivity of ·OH, has driven a substantial increase in TSOA sensitivity in the West. The influence of* NO_x *is less significant in other regions. Zheng et al. (2023) suggested that the impact of anthropogenic* NO_x *on monoterpene SOA formation is missing from current models, highlighting the need for future studies to evaluate how this omission may affect simulations of temperature sensitivity.*

[RC 1.7] Early descriptions of the climate penalty with respect to PM2.5 appear to be overly simplistic relative to the cited literature. The authors state that “Higher temperatures are generally associated with exacerbated PM2.5 pollution”, but multiple studies (including some of those immediately cited) have offered a far more mixed message on observed and modeled PM2.5 temperature sensitivities, especially relative to those of tropospheric ozone. Later discussions towards the end of the introduction do go into more depth and detail on some of the disagreements seen in the literature, but I think this nuance and complexity should at least be mentioned earlier on as well, especially if papers showing differing sensitivity results are being cited.

[AC 1.7] We thank the reviewer for this helpful comment. While elevated temperatures are often associated with increased PM2.5 under certain conditions, numerous observational and modeling studies demonstrate that PM2.5 temperature sensitivities vary widely in sign and magnitude across regions, time periods, and chemical components, and are substantially less consistent than those of tropospheric ozone. We agree that this simplification may give a misleading impression.

We have therefore revised the early Introduction to explicitly acknowledge the mixed and sometimes contradictory findings in the literature, emphasizing the strong regional dependence of PM2.5 temperature sensitivity and the contrast with the more robustly positive ozone–temperature relationship:

(Line 43-49) *Higher temperatures have often been associated with increased* $PM_{2.5}$ *pollution, a phenomenon termed the "climate penalty," which reflects the potential deterioration of air quality due to warming in the absence of changes in anthropogenic activities (Bloomer et al., 2009; Duffy et al., 2019; Jacob and Winner, 2009; Schnell and Prather, 2017; Tai et al., 2010; Wu et al., 2008). This deterioration, in turn, adversely impacts human health and contributes to climate feedbacks via aerosol radiative effects. It is worth noting, however, that observational and modelling studies show that the sign and magnitude of this relationship vary substantially across regions, time periods, and* $PM_{2.5}$ *components, and are less consistent than those observed for tropospheric ozone.*

RC#2

This manuscript presents a rigorous analysis of the regional and temporal trends in the sensitivity of PM2.5 to temperature across CONUS. The improved model modifications (e.g. using GFED4, including the SOA coating effects, and the scaling of SO2 emissions) yield outputs that agree sufficiently well with observations, for the most part. Moreover, the breakdown quantifying the contributions of individual processes within the model is significant in advancing the study of model evaluations and design beyond simple observational correlations. However, the analysis reveals a few structural and methodological issues that, if addressed, would greatly increase the paper's overall impact and credibility.

[RC 2.1] Addressing discrepancies in the representation of nitrate with temperature:

The model's strong negative correlation for nitrate and the resulting bias in ammonium sensitivity represent the most persistent and significant discrepancy between simulations and observations in the Eastern and Central US. I suggest using the decomposition approach to try and quantify the relative contribution of competing effects in driving the negative nitrate bias. For example, discerning between the sensitivity of the gas-particle partitioning with temperature and the sensitivity of NOx emissions with temperature.

[AC 2.1] We thank the reviewer for this insightful suggestion. We have applied a decomposition analysis to disentangle the drivers of nitrate temperature sensitivity, specifically separating the contributions from precursor availability (total nitrate) and thermodynamic gas-particle partitioning. The relevant text reads as follows:

(Line 703-715) *Additionally, we performed a decomposition analysis for nitrate to investigate the mechanisms driving the model's predicted negative temperature sensitivity. We expressed particulate nitrate NO_3^-) as the product of total nitrate ($TNO_3 = HNO_3 + NO_3^-$) and the particle-phase fraction (f_p).*

$$[NO_3^-] = [TNO_3] \times f_p$$

The temperature sensitivity was then decomposed into two terms:

$$\frac{d[NO_3^-]}{dT} \approx f_p \times \frac{d[TNO_3]}{dT} + [TNO_3]\frac{df_p}{dT} \qquad (4)$$

1. *Production Driver* $\frac{d[TNO_3]}{dT}$*: Sensitivity mediated by changes in total nitrate (driven by* NO_x *emissions and oxidation).*
2. *Thermodynamic Driver* $\frac{df_p}{dT}$*: Sensitivity mediated by changes in the gas-particle partitioning fraction* f_p*.*

Our analysis confirms that the thermodynamic driver is the dominant factor, as shown in Supplementary Figure 25. The model simulates a strong negative response of the partitioning fraction to temperature, which causes substantial volatilization of ammonium nitrate. This negative thermodynamic term overwhelms the production term, which is weakly positive, driven by slight increases in NO_x *emissions or oxidation rates with temperature. This quantification highlights that* *the model's negative nitrate bias is largely due to the high sensitivity of the partitioning equilibrium to temperature.*

[RC 2.2] Addressing uncertainty in the decomposition of individual processes:

The decomposition of the total sensitivity into contributions from precursors relies on a linear regression of detrended anomalies as a proxy for partial derivatives. While the results generally align with the total sensitivity, the acknowledged interdependencies among variables suggest this simplification may introduce uncertainty. To increase the robustness of this quantification method, I might include a simple uncertainty assessment, even if only for a smaller (but key) representative region (e.g. SO4 in the Southeast US). This could involve demonstrating that the derived process sensitivities are robust across different rolling temporal windows or alternative detrending methods.

[AC 2.2] We thank the reviewer for this constructive suggestion and agree that the regression-based decomposition of total temperature sensitivity into individual process contributions introduces uncertainty due to interdependencies among variables. As noted in the manuscript, the linear regression of detrended anomalies provides a first-order approximation rather than a strict physical partial derivative.

To assess the robustness of this approach, we performed an explicit sensitivity analysis using multiple rolling temporal windows. In addition to the 5-year rolling windows shown in Figure 6, we now include results derived from 3-year rolling windows for isoprene SOA and sulfate (Supplementary Figure 24). This analysis only serves as an uncertainty and robustness check, as shorter windows amplify sensitivity to interannual variability.

Across both window lengths, the relative contributions of different processes remain consistent, and the summed contributions reproduce the magnitude and variability of the total temperature sensitivity. The persistence of these patterns across different temporal windows and climatic regimes provides confidence that the decomposition method captures the dominant physical drivers despite unavoidable interdependencies among variables. The consistency of results across rolling window lengths suggests that our conclusions are not sensitive to the specific temporal framework used. We have clarified this point in the revised manuscript:

(Line 694-702) *To verify the robustness of the decomposition method, we performed a sensitivity analysis using multiple rolling temporal windows. In addition to the 5-year rolling windows shown in Figure 6, we include results derived from 3-year rolling windows for isoprene SOA and sulfate (Supplementary Figure 24). This analysis serves as an uncertainty and robustness check, as shorter windows amplify sensitivity to interannual variability and potential detrending choices. Across both window lengths, the summed contributions from individual processes consistently reproduce the magnitude and variability of the total temperature sensitivity. The persistence of these patterns across different temporal windows provides confidence that the regression-based decomposition captures the dominant physical drivers despite unavoidable interdependencies among variables. Overall, the consistency of process contributions across rolling window lengths indicates that the decomposition is robust to temporal sampling and detrending assumptions.*

[RC 2.3] Addressing outlying regarding wildfire OA sensitivity in the Western US:

The authors acknowledge that the significant OA sensitivity overestimation in the Western US in more recent years (2016-2022) could be attributed to extreme values in the GFED4 inventory. However, it might be more impactful to reframe this occurrence, rather than with a critique of the GFED4 inventory itself but with an acknowledgment that the current CTM framework does not handle these extreme events well. The modeling of the impact of future wildfires will always carry significant uncertainty, and for certain regions of the US, this inability to accurately predict wildfire contributions has a large impact on overall PM2.5 projections. However, it's clear that the wildly exaggerated modeled primary OA sensitivities (as shown in Figure 2) are obfuscating trends in SOA that are perhaps more useful to focus on in designing CTMs. Indeed, the scaling on row b of Figure 2 makes discerning trends very difficult prior to the sudden increase. (This criticism is one that I would apply to many of the figures in this manuscript, text is too small, plots are too crowded, and it is a struggle to try and make sense of things).

[AC 2.3] We thank the reviewer for this insightful comment. We have added more discussion on the overestimation of OA in 2021. Please refer to [AC 1.3] for more details.

(Line 245-254) *The higher RMSE for OA simulations in certain regions can be attributed to overestimated concentrations in 2021. This discrepancy is likely driven by exceptionally high wildfire emissions within the GFED4 inventory for certain years, such as 2021 (Supplementary Figure 4). Because wildfire-driven OA is strongly temperature dependent and spatially heterogeneous, these overestimations can disproportionately affect regression-based temperature sensitivity diagnostics. Additionally, potential under-sampling by ground-based monitors near fire zones may lead to an underestimation of regional mean concentrations in observational datasets. Furthermore, current CTMs face inherent challenges in capturing the non-linear chemistry and rapid transport associated with extreme fire events (Qiu et al., 2024). Despite these uncertainties, data from high-fire years were retained in the full record, as they represent the growing influence of extreme wildfire events on air quality in a warming climate and reflect an important stress test for current modelling frameworks.*

Regarding figure presentation, we have redrawn row b in Figure 2 using a broken y-axis to accommodate the large range introduced by the 2021 anomaly while preserving the visibility of the trends from 2000–2016. We have also made broader improvements to figure layout and text size throughout the manuscript to enhance clarity.

[RC 2.4] There are minor grammatical inaccuracies throughout. I'll list some examples here, as shown from Lines 695-708. My suggestions are written in capitalized letters.

"We quantified the contributions from THE **temperature-dependence of isoprene and sulfate to ISOAAQ sensitivity."**

"Given the important role of gas-phase production in MODULATING **sulfate sensitivity, as indicated by our findings, we further quantified the contributions from the temperature response of precursors of** THE **gas phase reaction… and** THE **cloud fraction to sulfate temperature sensitivity."**

“The long-term temporal pattern of THE **temperature sensitivity of sulfate is mainly driven by the decreasing response of SO2 concentrations to temperature as SO2 emissions declined** OVERALL**.” (Remove “rise” to make the sentence flow better).**

“For monoterpene SOA, gas-particle phase partitioning plays a significant role in overall sensitivity, while its dependence on precursor concentrations, including monoterpeneS**, ·OH, and NO_x, collectively contributes to interannual variability.” (I think monoterpenes should be pluralized here, to indicate the full class of molecules.)**

I picked this paragraph as an example, but similar grammatical inconsistencies are found throughout the paper. I suggest doing a full round of edits looking specifically at the wording.

[AC 2.4] We thank the reviewer for the detailed proofreading. We have corrected the specific instances noted and have conducted a thorough proofread of the entire manuscript to fix similar grammatical inconsistencies.

References:

Pai, S. J., Heald, C. L., Pierce, J. R., Farina, S. C., Marais, E. A., Jimenez, J. L., Campuzano-Jost, P., Nault, B. A., Middlebrook, A. M., Coe, H., Shilling, J. E., Bahreini, R., Dingle, J. H., and Vu, K.: An evaluation of global organic aerosol schemes using airborne observations, Atmospheric Chem. Phys., 20, 2637–2665, https://doi.org/10.5194/acp-20-2637-2020, 2020.

Qiu, M., Kelp, M., Heft-Neal, S., Jin, X., Gould, C. F., Tong, D. Q., and Burke, M.: Evaluating Chemical Transport and Machine Learning Models for Wildfire Smoke $PM_{2.5}$: Implications for Assessment of Health Impacts, Environ. Sci. Technol., 58, 22880–22893, https://doi.org/10.1021/acs.est.4c05922, 2024.

Yin, L., Bai, B., Zhang, B., Zhu, Q., Di, Q., Requia, W. J., Schwartz, J. D., Shi, L., and Liu, P.: Regional-specific trends of PM2.5 and O3 temperature sensitivity in the United States, Npj Clim. Atmospheric Sci., 8, 1–15, https://doi.org/10.1038/s41612-024-00862-4, 2025.