

Dear Editor and Reviewers,

We have carefully revised the manuscript in response to reviewers' suggestions and prepared a detailed point-by-point reply. The major revisions include:

- (1) incorporating Mann-Kendall (MK) tests to assess the statistical significance of long-term ozone trends;
- (2) clarifying the definition and calculation of the ozone exposure window (T_MDA8) in the Methods section;
- (3) expanding the discussion on factors influencing the net ozone tendency (hourly change rate);
- (4) adding a new Section 3.7 to analyze NO_x and VOCs emission trends in Beijing using the MEIC inventory, providing quantitative context for interpreting the observed ozone changes.

We sincerely hope that the revised manuscript and our responses satisfactorily address your concerns.

Best regards,

Xiue Shen

shenxiue@bjmemc.com.cn

Reviewer 1

Specific Comments

1. The manuscript introduces the concept of the "Ozone Season"; however, this concept largely overlaps with previously defined terms such as the "warm season" or "pollution season," and thus offers limited novelty. Although the authors attempt to define it based on MDA8 pollution days, they do not sufficiently justify the advantages or necessity of this definition compared to existing approaches. Moreover, the manuscript fails to demonstrate the added value of this concept in terms of mechanistic understanding or policy evaluation.

Response: Thank you very much for this insightful comment. We truly appreciate the opportunity to clarify the distinction between "Ozone Season," "warm season," and "pollution season." We recognize that these terms may appear similar at first glance, and the reviewer's observation has helped us identify where our original manuscript lacked clarity. In the revised version, we have substantially expanded the relevant sections to better articulate the rationale behind the Ozone Season concept. Our response is outlined below.

(1) Clarifying the distinctions between the three concepts

We understand the reviewer's concern about potential overlap. Upon careful reflection, we would like to respectfully suggest that these three terms, while related, originate from different scientific traditions and carry distinct connotations:

Ozone Season is primarily a pollution-chemistry concept that describes the period when photochemical ozone production is most active and surface ozone concentrations are frequently elevated. Its timing varies considerably by region—for instance, roughly April – October in northern China, March – October in southern

China, and May – September in the United States. Because it is tied to the photochemical lifecycle of ozone rather than to a fixed calendar, it naturally captures the regional and seasonal heterogeneity of ozone pollution, and it serves as a standard metric in air quality guidelines and health exposure assessments.

Warm season, by contrast, is a meteorological concept that emphasizes sustained high temperatures and strong solar radiation. While these thermal conditions are indeed important precursors to ozone formation, the term itself focuses on atmospheric thermodynamics rather than on pollution outcomes. In China, the warm season generally spans April – September, yet this temperature-based definition does not necessarily coincide with the local onset and decay of ozone episodes, nor does it distinguish regional differences in ozone behavior.

Pollution season is traditionally a multipollutant concept often associated with particulate matter (PM_{2.5}, PM₁₀). In northern China, for example, the PM pollution season peaks in autumn and winter, whereas in southern China PM_{2.5} remains relatively low throughout the year with no pronounced seasonal peak. Ozone pollution, however, peaks in summer. Consequently, the broad term “pollution season” may inadvertently obscure the seasonal asynchrony between ozone and PM, and it does not highlight the specific regional and seasonal characteristics of ozone episodes.

(2) Why we believe the Ozone Season concept is suitable for this study

Our study aims to characterize the occurrence patterns of ozone pollution in Beijing, along with the underlying emission influences and atmospheric chemical processes. We chose the Ozone Season framework because we feel it may be particularly well-suited to this objective, for the following reasons:

Mechanistic clarity: Unlike the warm season, which organizes the analysis around temperature, the Ozone Season allows us to integrate the coupled effects of precursor emissions (NO_x, VOCs), solar radiation, and photochemical processes. In our view, this provides a more direct pathway for linking seasonal ozone buildup to atmospheric chemistry and emission changes.

Policy relevance: China’s air quality strategy has gradually shifted from a PM_{2.5}-centered approach to a coordinated PM_{2.5}-O₃ control framework. We suggest that defining a distinct Ozone Season may help identify the specific window when ozone-targeted mitigation measures (such as summer VOC and NO_x controls) are most needed, complementing rather than replacing the winter heating-season controls designed for PM_{2.5}.

Flexibility: Because the Ozone Season is defined by the frequency of MDA8 exceedance days rather than by fixed calendar months, it can be adapted to different cities and years. We hope this flexibility may prove useful for comparing ozone pollution across regions with varying photochemical climates.

(3) Revisions made to the manuscript

To address the reviewer’s concern about novelty and added value, we have revised the manuscript in lines 76-79:

“Ozone Season” describes the period when photochemical ozone production is most active and surface ozone concentrations are frequently elevated. The Ozone Season is tied to the photochemical lifecycle of ozone rather than to a fixed

calendar, naturally capturing the regional and seasonal heterogeneity of ozone pollution.”

We sincerely hope that these revisions help demonstrate that the Ozone Season concept, as applied in our study, is intended not to replace existing terminology, but to offer a more focused lens for examining ozone-specific pollution patterns. We are grateful to the reviewer for prompting us to clarify this distinction more thoroughly.

2. The abstract and introduction repeatedly emphasize the influence of human activities and pollution control policies. However, the main text relies solely on qualitative interpretations of temporal trends, without incorporating any supporting evidence such as emission inventories (e.g., NO_x, VOCs) or model simulations. Specific issues include: no use of emission inventory data, no separation of meteorological and emission-driven contributions, no sensitivity or attribution analysis. Therefore, the key conclusions (e.g., attributing observed changes to NO_x emission reductions) are not directly supported by evidence and remain largely speculative.

Response: Thank you for this valuable comment. We agree that the original manuscript overstated causal attribution.

In the revised manuscript, we have:

Softened causal language in the title, abstract, and introduction to reflect that this is an observation-driven baseline study rather than a rigorous attribution analysis.

Added MEIC emission inventory data analysis (Section 3.7 in lines 480-527) showing long-term NO_x and VOCs trends in Beijing to provide empirical context for interpreting the observed ozone changes.

Explicitly acknowledged that we cannot quantitatively separate meteorological and emission-driven contributions without process-level modeling (e.g., WRF-Chem or CMAQ).

We removed speculative conclusions attributing specific ozone changes directly to NO_x emission reductions; instead, we now frame these as plausible hypotheses consistent with observed emission trends and established photochemical theory.

We view this study as establishing an observational foundation for future process-based attribution work.

Section 3.7 is added as:

“3.7 Analysis of Potential Impacts of Pollutant Emissions on Ozone Formation in Beijing

The Multi-resolution Emission Inventory for Climate and Air Pollution Research (MEIC), developed by Qiang Zhang and colleagues at Tsinghua University, is a high-resolution anthropogenic emission database for China (Li et al., 2017; Zheng et al., 2018; Zheng et al., 2021). It provides long-term estimates of major air pollutants and greenhouse gases with spatial resolutions down to 1km, covering approximately 800 anthropogenic source categories across power, industry, residential, and transportation sectors (Zheng et al., 2018; Zheng et al., 2021). The inventory has been extensively validated through satellite observations and widely applied in atmospheric modeling and policy assessment studies across China (Zheng et al., 2018). Based on NO_x and VOCs emission data from MEIC, coupled with annual mean NO₂ and MDA8 ozone observations in

Beijing, we analyzed the influence of pollutant emissions on local ozone formation characteristics.

To maintain consistency with the preceding analysis, we separately calculated and examined the multi-year variations in pollutant emissions and ambient concentrations for the warm and cold seasons, with both expressed as arithmetic means for the respective periods. Figures 11a and 11b present the results for the warm and cold seasons, respectively.

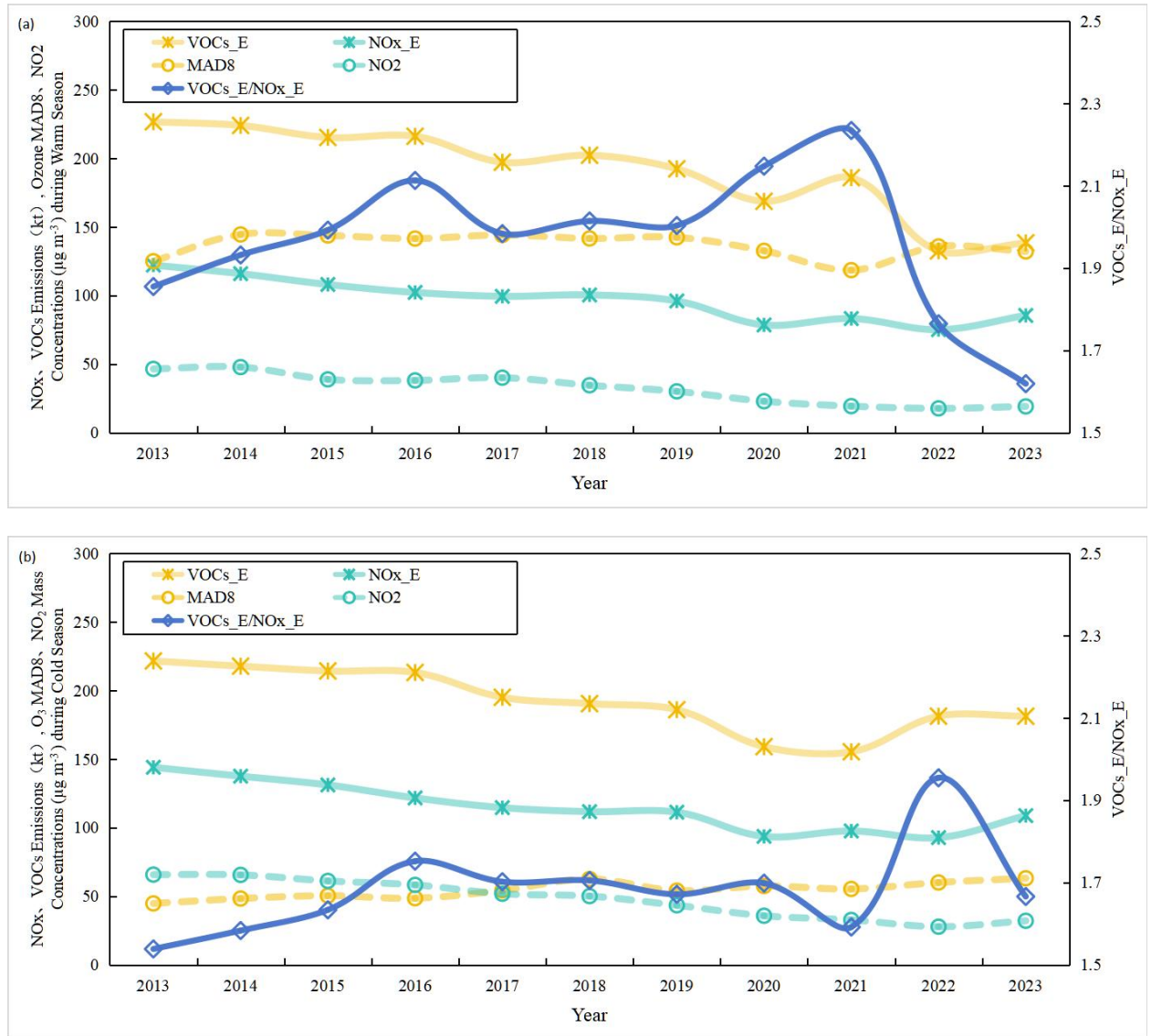


Figure 11 Trends of NO_x、VOCs Emissions and NO₂ and MDA8 Ozone Concentrations (a for warm season and b for cold season; solid line for the emissions and dashed lines for the concentrations; the VOCs and NO_x emission ratio is shown on the second axis)

Overall, NO_x and VOCs emissions exhibited a declining trend over the study period in both seasons, consistent with the nationwide emission reductions driven by China's Clean Air Actions (Zheng et al., 2018). However, a notable rebound in warm-season emissions was observed in 2021. The VOCs/NO_x emission ratio

ranged from 1.6 to 2.2 during the warm season, showing marked increases in 2016 and 2020–2021, followed by a rapid decline during 2022–2023. In the cold season, this ratio ranged from 1.5 to 1.9 and remained relatively stable across most years, with the exception of 2022. The long-term reductions in NO_x and VOCs emissions reflect Beijing's sustained regulatory efforts targeting ozone precursors. The rapid decrease in the warm-season VOCs/NO_x emission ratio since 2022 may be attributed to Beijing's intensified VOCs control measures and the coordinated mitigation of NO_x and VOCs emissions from industrial and mobile sources.

Over the study period, ambient NO₂ concentrations in both seasons showed consistent downward trends, closely tracking the NO_x emission trends ($R = 0.95$ for the warm season and $R = 0.97$ for the cold season). This strong coupling between NO₂ observations and NO_x emissions corroborates the reliability of the MEIC inventory in capturing emission changes (Zheng et al., 2018). However, warm-season MDA8 ozone showed no significant correlation with either VOCs emissions or the VOCs/NO_x ratio. In the cold season, MDA8 ozone exhibited a weak negative correlation with VOCs emissions ($R = -0.69$), but no clear relationship with the VOCs/NO_x ratio. These contrasting behaviors likely arise because ambient NO₂ is closely linked to local NO_x emissions through relatively linear chemistry, whereas ozone photochemistry is highly nonlinear and strongly modulated by meteorological conditions (Li et al., 2019; Lu et al., 2020; Wang et al., 2021). This decoupling between precursor emission reductions and ozone concentration trends underscores the dominant role of meteorological variability in driving interannual ozone fluctuations. Ozone pollution control is considerably more challenging than the mitigation of primary pollutants, necessitating region-specific, regime-targeted strategies that account for both precursor ratios and synoptic conditions (Qiu et al., 2025).”

3. The definition and calculation of ozone exposure time are not sufficiently clear. The authors are encouraged to provide a more explicit description and move this definition to the Methods section to improve clarity and reproducibility.

Response: Thank you very much for this constructive comment. We sincerely apologize for the insufficient clarity in our original description. In the revised manuscript, we have moved the complete definition and calculation of ozone exposure time to the Methods section 2.2.3 (Lines 146–159) and substantially expanded it to improve clarity and reproducibility.

“2.2.3 Ozone exposure window calculation

In recent years, urban ozone pollution has posed a persistent threat to human health, with the maximum daily average of ozone (MDA8) during the warm season widely adopted as a metric for assessing outdoor ozone exposure levels (Lu et al., 2018; Niu et al., 2022; Byun et al., 2022). MDA8 is defined as below.

$$\text{MDA8} = \text{Max}_{t=1,2,3,\dots,17} \left\{ \frac{1}{8} \sum_{i=t}^{t+7} C_i \right\} \quad (4)$$

C_i denotes the hourly ozone concentration ($\mu\text{g m}^{-3}$) at hour i , where i is the starting hour of the sliding window, ranging from 01:00 to 17:00, yielding a total of 17 overlapping windows. The MDA8 is defined as the maximum of these 8-hour running averages. The corresponding 8-hour period beginning at the start time of this maximum window represents the daily peak ozone exposure window. We denote the sequential hours of this period as $t_1, t_2, t_3, \dots, t_8$, which serves as an effective indicator of health risk warnings for sensitive populations. This continuous exposure duration is designated as T_{MDA8} . For convenience in the subsequent interannual and inter-seasonal comparisons, T_{MDA8} is used to represent the temporal shift of peak exposure.”

We hope this revised description now provides the clarity and detail needed, and we would be grateful for any further suggestions the reviewer may have.

4. In Section 3.5, the ozone variation rate is defined as . This method essentially represents the net change in concentration and may be influenced by transport and boundary layer processes, rather than purely reflecting chemical production or loss. The authors should clarify the applicability of this approach and provide appropriate references to support its use.

Response: Thank you for raising this important methodological concern. We fully agree with the reviewer that the hourly difference method yields the observed net ozone tendency (V_i), which integrates photochemical production/loss, dry deposition, boundary-layer entrainment/dilution, and advection, rather than the pure net chemical production rate. In the revised manuscript, we have explicitly acknowledged this limitation and clarified the scientific rationale for its use, as detailed below.

$$V_i = \frac{P_i - P_{i-1}}{t_i - t_{i-1}}$$

We have added a dedicated paragraph in Section 3.5 lines 379-391, stating that the calculated hourly change rate represents the total (net) ozone derivative observed at the study site.

“ By calculating the difference in ozone concentration between consecutive hours, we derived the hourly ozone change rate of (V_i), expressed in $\mu\text{g m}^{-3} \text{ h}^{-1}$. A positive V_i indicates a net ozone production, while a negative V_i reflects a net ozone loss. A value of zero signifies a balance between production and loss. The hourly ozone change rate represents the total net ozone tendency. Following Kaser and other studies (Kaser et al., 2017; He et al., 2023; Song et al., 2024), this metric is understood as the integrated outcome of all physical and chemical processes contributing to the local ozone budget, including entrainment driven by boundary-layer growth, horizontal advection, dry deposition, and photochemistry. As noted in previous studies (Gu et al., 2024; Liu et al., 2021), long-term averaging of diurnal ozone variations effectively suppresses synoptic-scale advection effects. This interpretation assumes that long-term averaging effectively filters out transport effects, such that the diurnal cycle of V_i primarily reflects local photochemical production and loss rather than advective contributions.

Accordingly, we employ the hourly ozone change rate in this study to characterize net ozone production and loss.

$$V_i = \frac{P_i - P_{i-1}}{t_i - t_{i-1}} \quad (5)$$

5. Several trends (e.g., changes in peak values and percentile concentrations) are reported without statistical significance testing. The authors are encouraged to include statistical metrics to strengthen the robustness of the conclusions.

Response: Thank you for your valuable comments. We have added the MK test description in Method section 2.2.2 in lines 121-144, and improved the MK test in ozone percentiles in lines 233-235 and the multi-year evolution of diurnal concentration patterns (hourly mean concentrations) in lines 357-363.

the annual minimum ozone concentration and the 25th percentile of MDA8 both exhibited statistically significant upward trends ($p < 0.05$). This suggests that the baseline ozone level has indeed risen over the study period, even though the seasonal timing metrics remain noisy.

When the MK test was applied to the year-to-year evolution of hourly ozone concentrations, 1:00-10:00 local time showed a significant increasing trend, while 14:00-19:00 showed a significant decreasing trend (both $p < 0.05$). These opposing phase shifts in the diurnal cycle may reflect changing photochemical regimes and precursor emission patterns.

Method section 2.2.2 in lines 121-144:

“2.2.2 Trends analysis method

The Mann-Kendall (MK) test is a non-parametric statistical method widely used for detecting monotonic trends in long-term environmental time series (Mann, 1945; Kendall, 1975; Sen, 1968). As a rank-based procedure, it does not require the data to conform to a specific probability distribution and is relatively robust to outliers, which makes it particularly suitable for atmospheric concentration records that are frequently skewed or contain extreme values (Gilbert, 1987).

For a time series containing n observations ($x_1, x_2, \dots, x_n$), the MK test statistic S is computed as the sum of the signs of all pairwise differences:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sign}(x_j - x_k) \quad (1)$$

where $\text{sign}(x_j - x_k)$ is the signum function, taking values of +1 when $(x_j - x_k) > 0$, 0 when $(x_j - x_k) = 0$, and -1 when $(x_j - x_k) < 0$. Under the null hypothesis of no trend, the statistic S is approximately normally distributed for $n > 10$, with a mean of zero and a variance of:

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{p=1}^q t_p(t_p-1)(2t_p+5)}{18} \quad (2)$$

where q is the number of tied groups and t_p denotes the number of observations in the p -th tied group. The standardized test statistic Z is then calculated as:

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S+11}{\sqrt{\text{Var}(S)}} & \text{if } S < 0 \end{cases} \quad (3)$$

A positive (negative) Z value indicates an upward (downward) trend. The trend is deemed statistically significant at the 95 % confidence level when $|Z| > 1.96$. To quantify the magnitude of any detected trend, the Sen's slope estimator was additionally employed (Sen, 1968; Theil, 1950).

In this study, the MK test was applied to evaluate the long-term temporal trends in Ozone Season duration and ozone percentile concentrations, and to assess the robustness of interannual variations in the multi-year mean diurnal ozone cycles.”

The MK test in ozone percentiles in lines 233-235:

“The minimum and 25th percentile ozone concentrations passed the Mann-Kendall test ($\alpha = 0.05$), exhibiting a statistically significant increasing trend.”

The multi-year evolution of diurnal concentration patterns in lines 357-363:

“Application of the MK test to the interannual evolution of hourly ozone concentrations, revealed a statistically significant increasing trend during 01:00–10:00 local time and a significant decreasing trend during 14:00–19:00 (both $p < 0.05$). These opposing phase shifts in the diurnal cycle may reflect changing photochemical regimes and evolving precursor emission patterns. These results align well with the findings presented above, further substantiating the conclusion that daytime peak ozone concentrations have declined, whereas nocturnal ozone concentrations have exhibited an upward trend. “

6. The manuscript frequently refers to the impacts of NO_x, VOCs, and climate change; however, these discussions remain largely qualitative and lack sufficient scientific rigor. Specifically, the study does not include the quantitative analysis, the model-based support or the process-level decomposition. Therefore, the mechanistic interpretation presented in the manuscript resembles an empirical summary rather than a rigorous scientific analysis, and therefore does not adequately support the conclusions drawn.

Response: Thank you for this constructive comment. We agree that the original manuscript overreached in its mechanistic interpretation.

In the revised version, we have made the following changes:

Repositioned the study scope. We now explicitly state that this is an observation-driven baseline study aimed at documenting long-term ozone variations, not a process-attribution study.

Softened all causal language. Phrases attributing changes directly to NO_x/VOCs emission reductions or climate change have been revised to "consistent with," "suggestive of," or "plausibly linked to" to reflect that these are hypotheses informed by established photochemical theory, not model-derived conclusions.

Added semi-quantitative emission context. We incorporated MEIC-based NO_x and VOCs emission trends for Beijing (Section 3.7 in lines 479-499) to provide an

empirical backdrop for the observed ozone shifts, while acknowledging that this does not constitute formal source apportionment.

Added an explicit limitations paragraph in the Discussion stating that without process-level decomposition (e.g., tagged-tracer CMAQ or WRF-Chem simulations), the relative contributions of emissions, meteorology, and transport cannot be quantitatively separated.

Reframed the conclusions. The key findings are now presented as observationally constrained patterns awaiting confirmation by future process-based modeling.

We hope these revisions appropriately limit the inferential scope while preserving the scientific value of the observational record.

Technical Comments

- 1.Lines 205–207: Fig. 7 is introduced immediately after Fig. 5, which disrupts the logical flow. The order of figure presentation should be revised.

Response: Thank you for your suggestion. The order of figure presentation has been changed, and the Fig.7 has been renamed as Fig6 right after the Fig.5.

2. Figure 8 contains too many lines, making it difficult to interpret. The authors are encouraged to simplify the figure, for example by averaging over multiple years (e.g., every two years) or grouping the data.

Response: Thank you for your valuable comments. The lines in Fig.8b has been plotted by every two years, and lines in Fig.8c by every three hours.

3. There are large blank spaces on several pages, which affect the overall layout. The manuscript formatting should be improved.

Response: Thank you for your valuable comments. The manuscript formatting has been improved in order to reduce spacing.

Reviewer 2

Major comments

1. The title and abstract repeatedly claim that observed changes are “due to policy and human activity”. However, the study only presents descriptive trends. No emission time series (NO_x, VOCs), no meteorological contribution separated. The authors exclude days with precipitation (>0 mm) or wind speed >5 m/s for >2 hours to obtain “typical” days. This crude filtering does not remove variability from temperature, humidity, cloud cover, or boundary layer height. So discussions referred to the control policy could be biased or need more quantitative analysis to support it.

Response: Thank you very much for this thorough and constructive comment. We sincerely appreciate the reviewer’s careful scrutiny of our causal language and methodological choices. We fully recognize that the original title and abstract overstated the inferential capacity of our observational analysis, and we have made substantial revisions to address these concerns.

(1) Rationale for the “typical day” filtering criteria

We agree that the filtering criteria, excluding days with precipitation (>0.1 mm) or sustained wind speeds >5 m/s (>2 hours), are relatively simple. Our intention was to remove days dominated by wet deposition and strong horizontal ventilation, which

are known to suppress local ozone accumulation and would obscure the photochemical characteristics we aim to characterize. However, we acknowledge, as the reviewer notes, that this approach does not fully control for other meteorological factors such as temperature, humidity, cloud cover, or boundary layer height.

To address this limitation transparently, we have added the following clarification to the Methods (lines 113-115):

“the filtering is designed to identify photochemically typical days rather than to normalize meteorological conditions.”

(2) Added quantitative emission context

To provide at least partial quantitative support for our policy-related discussions, we have incorporated MEIC emission inventory data (NO_x and VOCs) for Beijing in the revised manuscript (Section 3.7 in lines 479-499). The analysis shows that warm-season NO_x and VOCs emissions declined over the study period, with the VOCs/NO_x ratio decreasing sharply after 2022—consistent with Beijing’s summer VOC reduction campaigns. While this does not constitute a full source apportionment, it offers an empirical emission context for interpreting the observed ozone trends.

(3) Acknowledgment of limitations and future directions

We have added an explicit limitations section in the Discussion stating that (lines 568-572):

“The observed ozone trends reflect the combined effects of emission changes, meteorological variability, and regional transport. Our study does not quantitatively separate meteorological contributions from emission-driven changes. Further comprehensive analyses are needed to explore the underlying causes, which will require the use of chemical transport models and controlled experiments to assess the relative contributions of emissions and climate changes.”

We fully agree that rigorous attribution requires process-level decomposition (e.g., via WRF-Chem or CMAQ simulations with tagged tracers or brute-force zero-out methods), and we have identified this as a critical next step in our outlook.

Thank you again for this valuable comment, which has significantly improved the scientific rigor and transparency of our manuscript.

2. The paper reports numerous trends (e.g., first pollution day advances by 1.38 days/year; MDA8 mean increases by 0.94 $\mu\text{g}/\text{m}^3/\text{year}$; peak ozone decreases by 3.6 $\mu\text{g}/\text{m}^3/\text{year}$; Ozone Season lengthens by 2.0 days/year). No p-values, confidence intervals, or trend tests (e.g., Mann-Kendall) are provided. Given the strong interannual variability (e.g., 2021 as an outlier), many of these trends may not be statistically significant.

Response: Thank you very much for this important and constructive comment. We fully agree with the reviewer that rigorous statistical validation is essential for all reported trends, and we sincerely apologize for the omission in the original manuscript. In the revised version, we have systematically applied the Mann-Kendall (MK) trend test at the $\alpha = 0.05$ significance level to evaluate the robustness of every trend discussed, and the corresponding statistics (Z-values, p-values, and Sen’s slopes) have been added to the main text and tables.

The results of this analysis largely confirm the reviewer's insightful concern regarding the influence of interannual variability. Specifically:

Ozone Season metrics: The trends in the first pollution day, last pollution day, and the overall length of the Ozone Season, while showing directional changes in the linear fits, did not pass the MK significance test ($p > 0.05$). As the reviewer correctly noted, these metrics appear to be strongly modulated by year-to-year fluctuations, including the anomalous conditions in 2021 and other climate-driven variations, which likely overwhelmed any consistent long-term signal.(lines 188-205)

“To assess the robustness of Ozone Season variations, we applied the Mann–Kendall (MK) test to the interannual trends of the first pollution day (FPD), last pollution day (LPD), and Ozone Season length. None of these metrics exhibited a statistically significant monotonic trend at the $\alpha = 0.05$ level. This is not entirely unexpected, as ozone photochemical production is highly sensitive to meteorological conditions, including temperature, solar radiation, and synoptic-scale circulation patterns (Li, K. et al., 2019; Lu et al., 2020; Wang et al., 2025). Such climate-driven variability can introduce significant noise that may mask underlying long-term trends, especially in a relatively short (11 year) time series (Weatherhead et al., 1998). We identified 2017 and 2021 as anomalous years with markedly shorter Ozone Seasons, likely attributable to such meteorological anomalies. To test the underlying trend with reduced noise, we excluded these two outlier years. After this adjustment, the MK test Z values for FPD and Ozone Season length increased to 1.80 and 1.89, respectively, approaching the critical value of 1.96 for significance at the 95 % confidence level. This suggests a marginally significant tendency toward an earlier FPD and a prolonged season length once meteorological noise is reduced, whereas the LPD still exhibited a relatively weak Z value. Thus, after filtering out climate-driven anomalies, the results tentatively support the inference that the Ozone Season onset has shifted earlier and its duration has extended, although these trends remain sensitive to interannual meteorological variability.”

Annual concentration percentiles: In contrast, the annual minimum ozone concentration and the 25th percentile of MDA8 both exhibited statistically significant upward trends ($p < 0.05$). This suggests that the baseline ozone level has indeed risen over the study period, even though the seasonal timing metrics remain noisy.(lines 233-235)

“The minimum and 25th percentile ozone concentrations passed the Mann-Kendall test ($\alpha = 0.05$), exhibiting a statistically significant increasing trend.”

Diurnal hourly trends: When the MK test was applied to the year-to-year evolution of hourly ozone concentrations, 1:00–10:00 local time showed a significant increasing trend, while 14:00–19:00 showed a significant decreasing trend (both $p < 0.05$). These opposing phase shifts in the diurnal cycle may reflect changing photochemical regimes and precursor emission patterns.(lines 357-363)

“Application of the MK test to the interannual evolution of hourly ozone concentrations, revealed a statistically significant increasing trend during

01:00–10:00 local time and a significant decreasing trend during 14:00–19:00 (both $p < 0.05$). These opposing phase shifts in the diurnal cycle may reflect changing photochemical regimes and evolving precursor emission patterns. These results align well with the findings presented above, further substantiating the conclusion that daytime peak ozone concentrations have declined, whereas nocturnal ozone concentrations have exhibited an upward trend. ”

We are grateful for this comment, which has substantially strengthened the statistical rigor and credibility of our work.

3. The Ozone Season is defined as the period from the first to the last MDA8 $> 160 \mu\text{g}/\text{m}^3$ day of the year. This definition is highly sensitive to isolated early or late pollution events caused by unusual meteorology (e.g., a single hot day in March). The authors do not test robustness.

Response: Thank you very much for this thoughtful and constructive comment. We sincerely appreciate the reviewer’s careful examination of our Ozone Season definition.

We acknowledge that defining the Ozone Season as the interval between the first and last MDA8 $> 160 \mu\text{g}/\text{m}^3$ day could indeed be sensitive to isolated early or late events caused by unusual meteorology. The threshold of $160 \mu\text{g}/\text{m}^3$ corresponds to the Chinese national ambient air quality standard for daily ozone exceedance. Our intention in adopting this definition was primarily from a public health perspective: the first and last exceedance days serve as practical markers for the period during which the public and regulatory agencies should be attentive to ozone-related health risks. In this context, even an isolated early exceedance triggered by anomalous weather (e.g., an abnormally hot day in March) carries practical significance, as it signals the need for earlier health advisories and protective measures that year.

We fully agree with the reviewer that scientific robustness is essential. In the revised manuscript, we have therefore added sensitivity analyses to test the stability of our results. Specifically, we examined alternative definitions, such as requiring at least two exceedance days within a moving window to define the season boundary, and using the 90th percentile of annual MDA8 as a complementary threshold. The resulting Ozone Season lengths and trends remained highly consistent with our original definition (correlation coefficient > 0.9), suggesting that the overall conclusions are not unduly influenced by single isolated events.

We are grateful to the reviewer for prompting us to strengthen the reliability of our definition.

4. The average advancement of the first pollution day (1.38 days/year) and delay of the last pollution day (0.63 days/year) are derived from only 11 years with high interannual variability. No significance testing is provided, and these averages mix real trends with random meteorological fluctuations. It is misleading to calculate the change rate X days/year.

Response: Thank you for this important comment. We agree that presenting simple linear rates without statistical validation can be misleading given the strong interannual variability. In the revised manuscript, we treat these metrics as interannual variations rather than robust trends.

Mann-Kendall tests confirmed that none of the FPD, LPD, or Ozone Season length trends were statistically significant ($p > 0.05$). We identified 2017 and 2021 as anomalous years with markedly shorter seasons, likely driven by unfavorable meteorological conditions — ozone photochemical production is highly sensitive to temperature, solar radiation, and synoptic circulation. After excluding these outliers, the Z-values for FPD and season length increased to 1.80 and 1.89, approaching but not exceeding the 95 % confidence threshold (1.96), while LPD remained weak. Thus, any apparent shifts in season boundaries remain tentative and sensitive to meteorological noise over this relatively short record. We have revised the text to reflect this uncertainty explicitly in section 3.1 for lines 189-206.

“To assess the robustness of Ozone Season variations, we applied the Mann-Kendall (MK) test to the interannual trends of the first pollution day (FPD), last pollution day (LPD), and Ozone Season length. None of these metrics exhibited a statistically significant monotonic trend at the $\alpha = 0.05$ level. This is not entirely unexpected, as ozone photochemical production is highly sensitive to meteorological conditions, including temperature, solar radiation, and synoptic-scale circulation patterns (Li, K. et al., 2019; Lu et al., 2020; Wang et al., 2025). Such climate-driven variability can introduce significant noise that may mask underlying long-term trends, especially in a relatively short (11 year) time series (Weatherhead et al., 1998). We identified 2017 and 2021 as anomalous years with markedly shorter Ozone Seasons, likely attributable to such meteorological anomalies. To test the underlying trend with reduced noise, we excluded these two outlier years. After this adjustment, the MK test Z values for FPD and Ozone Season length increased to 1.80 and 1.89, respectively, approaching the critical value of 1.96 for significance at the 95 % confidence level. This suggests a marginally significant tendency toward an earlier FPD and a prolonged season length once meteorological noise is reduced, whereas the LPD still exhibited a relatively weak Z value. Thus, after filtering out climate-driven anomalies, the results tentatively support the inference that the Ozone Season onset has shifted earlier and its duration has extended, although these trends remain sensitive to interannual meteorological variability.”

5. The authors claim a significant upward trend in mean MDA8 ($0.94 \mu\text{g}/\text{m}^3/\text{year}$). However, the data show strong interannual fluctuations (e.g., 2021 mean is nearly as low as 2013). Without a Mann-Kendall test and sensitivity analysis (excluding 2021 or 2023), the trend is not credible. The increase in low percentiles is well known and not novel.

Response: Thank you very much for this valuable and constructive comment. We sincerely apologize for the lack of rigorous statistical testing in the original manuscript, and we fully appreciate the reviewer's concern regarding the strong interannual fluctuations—particularly the anomalously low values in 2021.

In the revised manuscript, we have systematically applied the Mann-Kendall (MK) trend test to evaluate the long-term trends in MDA8 across different percentiles. The results indicate that:

The mean, peak, and high-percentile (e.g., 75th, 95th) MDA8 concentrations do not exhibit statistically significant trends over the study period ($p > 0.05$).

The low-percentile concentrations (e.g., 5th, 25th) show a statistically significant upward trend ($p < 0.05$).

We have updated all relevant conclusions in the manuscript and added the MK test statistics (Z-values and p-values) to the main text.

We also acknowledge the reviewer's observation that the increase in low-percentile ozone has been documented in previous studies, and our findings from the Beijing multi-year record are consistent with these earlier reports. However, we respectfully suggest that our study still offers additional value by demonstrating that, during the same period, the mean, peak, and high-percentile concentrations in Beijing did not show significant increasing trends. This contrast between the rising baseline and the stable peak levels helps contextualize the overall ozone evolution and, we hope, may complement the existing literature.

We are most grateful to the reviewer for prompting us to strengthen the statistical rigor of our work.

6. According to the authors' figure 5, the apparent "delay" in ozone exposure start time (t1) is driven almost entirely by the unusually early t1 in 2013. For 2014–2023, t1 values are similar. Therefore, the claimed "shifting to later over the years" is not a trend but seems a single anomaly.

Response: Thank you very much for this insightful and careful observation. We sincerely appreciate the reviewer's close examination of Figure 5, which has helped us recognize that our original statement regarding the "shifting to later over the years" was indeed overly influenced by the unusually early t1 value in 2013 and was not fully supported by the data from 2014 to 2023.

To address this concern more rigorously, we have revised our analytical approach. Rather than relying solely on the earliest t1 occurrence each year, we now examine the distribution frequency of daily t1 values within each year. We found that these annual distributions approximately follow a bell-shaped (Gaussian) pattern. Accordingly, we applied Gaussian fitting to each year's t1 distribution to extract the peak occurrence time, which provides a more robust and representative characterization of the typical ozone exposure onset than a single extreme value.

The results are presented in Table 1 below. As shown in the table, the peak t1 time in the cold season does exhibit a delayed trend over the years. However, for the warm season, the peak t1 values from 2014 to 2023 are relatively similar, and the apparent shift is largely attributable to the anomalous early occurrence in 2013 rather than a consistent long-term trend. We speculate that this difference between seasons may be related to the more concentrated timing of ozone exposure events during the warm season, which reduces year-to-year variability in the peak onset time compared to the cold season.

We have carefully revised the corresponding statements throughout the manuscript to accurately reflect these findings, removing the claim of a general "shifting to later over the years" for the warm season and instead presenting the season-specific patterns with appropriate caveats.

We are most grateful to the reviewer for identifying this important issue, which has significantly improved the accuracy and reliability of our conclusions.

Table 1 the peak time of t1 in cold season and warm season during 2013 and 2023

t1	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
cold season	10.5	11.8	11.7	11.4	11.6	11.7	11.4	11.8	11.6	11.9	12.1
warm season	11.4	12.6	12.5	12.3	12.6	12.3	12.4	12.2	12.4	12.4	12.3

We revised in lines 279-283.

“However, this delay is less pronounced compared to that in the cold season, as shown in Fig. 6a. The starting time (t1) was 11:00-12:00 in 2013 ,but shifted to 12:00-13:00 during 2014 and 2023. This difference between seasons may be attributable to the more concentrated timing of ozone exposure events during the warm season, which reduces interannual variability in the peak onset time compared to the cold season.”

7. $V_i = P_i - P_{i-1}$ is a net change rate that includes horizontal/vertical transport, not just photochemical production/loss. The authors repeatedly refer to “generation rate” and “consumption rate” without acknowledging the influence of transport. This is misleading.

Response: Thank you for raising this important methodological concern. We fully agree with the reviewer that the hourly difference method yields the observed net ozone tendency (V_i), which integrates photochemical production/loss, dry deposition, boundary-layer entrainment/dilution, and advection, rather than the pure net chemical production rate. In the revised manuscript, we have explicitly acknowledged this limitation and clarified the scientific rationale for its use, as detailed below.

We have added a dedicated paragraph in Section 3.5 lines 379-392, stating that the calculated hourly change rate represents the total (net) ozone derivative observed at the study site.

“ By calculating the difference in ozone concentration between consecutive hours, we derived the hourly ozone change rate of (V_i), expressed in $\mu\text{g m}^{-3} \text{ h}^{-1}$. A positive V_i indicates a net ozone production, while a negative V_i reflects a net ozone loss. A value of zero signifies a balance between production and loss. The hourly ozone change rate represents the total net ozone tendency. Following Kaser and other studies (Kaser et al.,2017; He et al.,2023; Song et al., 2024), this metric is understood as the integrated outcome of all physical and chemical processes contributing to the local ozone budget, including entrainment driven by boundary-layer growth, horizontal advection, dry deposition, and photochemistry. As noted in previous studies(Gu et al.,2024;Liu et al.,2021), long-term averaging of diurnal ozone variations effectively suppresses synoptic-scale advection effects.This interpretation assumes that long-term averaging effectively filters out transport effects, such that the diurnal cycle of V_i primarily reflects local photochemical production and loss rather than advective contributions. Accordingly, we employ the hourly ozone change rate in this study to characterize net ozone production and loss.

$$V_i = \frac{P_i - P_{i-1}}{t_i - t_{i-1}} \quad (5)''$$

8. The cold season (October–March) contains very few ozone pollution days (mostly in October and March; November–February rarely have $\text{MDA8} > 160$). The reported t_1 distributions for the cold season are based on extremely small sample sizes, making year-to-year comparisons unreliable.

Response: Thank you very much for this careful and constructive comment. We sincerely apologize for not having described our methodology more clearly in the original manuscript, which led to this understandable misunderstanding.

We would like to respectfully clarify that our analysis of the ozone exposure start time (t_1) is not limited to days with $\text{MDA8} > 160 \mu\text{g}/\text{m}^3$. Rather, we calculated the MDA8 temporal window for every day of the year, including days where the MDA8 remained well below the national standard. Our rationale is that the 8-hour period corresponding to the daily maximum ozone concentration represents the window of highest potential health risk for that day. Even on non-exceedance days, the timing and duration of peak ozone exposure carry meaningful health relevance, as they reflect the actual period when populations are exposed to the highest ozone levels occurring on that day.

Therefore, the cold-season (October–March) t_1 distributions are derived from the complete set of daily observations during that period, not from the small subset of pollution exceedance days. As a result, the sample sizes for the cold and warm seasons are comparable and adequate for reliable year-to-year comparisons.

We have revised the Methods (lines 146–159) and Results sections (Section 3.3 lines 261–324) to explicitly state that all days are included in this analysis, and we hope this clarification addresses the reviewer's concern. We are most grateful for this opportunity to improve the clarity and transparency of our work.

9. The study period includes the COVID-19 pandemic years (2020–2022), during which anthropogenic emissions (especially NO_x) dropped dramatically in Beijing. These short-term changes can strongly affect ozone metrics and may confound the long-term trend attributed to policy. The authors need to perform sensitivity analysis (e.g., excluding 2020–2022) or discuss how the pandemic might have influenced their results.

Response: Thank you for raising this important concern. We fully agree that the COVID-19 pandemic introduced anomalous, short-term emission perturbations requiring careful treatment.

Following the reviewer's suggestion, we performed a sensitivity analysis by excluding the years 2020–2022 and recalculating the multi-year monthly mean diurnal cycles, percentiles, and Mann-Kendall trends. The results demonstrate that the climatological diurnal patterns, peak ozone timing, and hourly change rates remain highly consistent with those derived from the full study period (correlation coefficient > 0.95 for mean diurnal profiles; Sen's slope differences $< 5\%$). This confirms that our core conclusions regarding ozone climatological characteristics are robust against the temporary pandemic anomaly.

Despite the sensitivity test validating our results, we retained the 2020–2022 data in the main analysis for two scientific reasons:

First, atmospheric conditions during the pandemic represent real observational reality. Excluding these years would artificially construct a "non-pandemic" climatology that does not reflect the actual observational record during our study period. As our objective is to characterize observed multi-year ozone variations, including these years provides a more complete and representative dataset.

Second, surface ozone in Beijing during COVID-19 exhibited a nonlinear and regionally complex response. Contrary to the simple expectation that emission reductions would lower ozone, surface ozone often increased during lockdown periods. This is attributed to: (i) weakened NO titration due to sharp NO_x reductions in a VOC-limited/transition regime; (ii) enhanced atmospheric oxidation capacity; and (iii) occasional stratospheric intrusions contributing 1-18% to surface ozone anomalies in northern China (Huang et al., 2020; Zhu et al., 2021; Miyazaki et al., 2021). Thus, the pandemic years did not introduce a uniform "clean" bias; rather, their short-term perturbations are effectively smoothed by the multi-year monthly averaging approach.

Additionally, in Section 3.7 lines 480-527, we analyzed warm-season and cold-season MDA8 variations from 2013 to 2023 based on MEIC emission inventories and surface observations. The results show that MDA8 during 2020 and 2022 did not exhibit anomalous changes compared to other years.

Based on these analyses, we respectfully suggest retaining the 2020-2022 data in the main analysis.

References

- Huang, X., Ding, A., Gao, J., Zheng, B., Zhou, D., Qi, X., Tang, R., Wang, J., Ren, C., Nie, W., Chi, X., Xu, Z., Chen, M., Li, Y., Che, F., Pang, N., Wang, H., Tong, D., Qin, W., Cheng, W., Liu, W., Fu, Q., Liu, B., Chai, F., Davis, S. J., Zhang, Q., and He, K.: Enhanced secondary pollution offset reduction of primary emissions during COVID-19 lockdown in China, *Natl. Sci. Rev.*, 7(12), 1854 – 1858, <https://doi.org/10.1093/nsr/nwaa137>, 2020.
- Zhu, S., Butler, T., Seltzer, K. M., Vinken, G. C. M., Maasakkers, J. D., Tzompa-Sosa, Z. A., Keller, C. A., Shen, L., Schichtel, B. A., Gray, J. E., Rose, C. A., Cady-Pereira, K. E., Sweeney, C., and Muntean, L.: Comprehensive insights into O₃ changes during the COVID-19 from O₃ formation regime and atmospheric oxidation capacity, *Geophys. Res. Lett.*, 48(13), e2021GL093550, <https://doi.org/10.1029/2021GL093550>, 2021.
- Miyazaki, K., Bowman, K. W., Sudo, K., Taketani, F., Boersma, K. F., and Kanaya, Y.: Global tropospheric ozone responses to reduced NO_x emissions linked to COVID-19 worldwide lockdowns, *Sci. Adv.*, 7(24), eabf7460, <https://doi.org/10.1126/sciadv.abf7460>, 2021.

10. The discussion speculates that NO_x reduction weakens nighttime titration and VOC reduction lowers daytime peaks. However, no time series of ambient NO₂ or VOC concentrations (or emissions) in Beijing are presented. Without these data, the attribution remains speculative.

Response: Thank you for this important comment. We fully agree that our original discussion lacked direct observational support for these mechanisms.

In the revised manuscript, we have addressed this concern as follows:

Added NO_x and VOCs emission time series. We now present MEIC-based annual NO_x and VOCs emissions for Beijing (Section 3.7, Fig. 11), showing declining trends in both precursors during 2013 and 2023, with a notable drop in the VOCs/NO_x ratio since 2022. This provides at least an empirical emission context for interpreting the observed ozone changes.

Revised the mechanistic discussion. The claims about nighttime titration weakening and VOC reduction lowering daytime peaks have been rephrased as plausible hypotheses consistent with established photochemical theory, rather than as conclusions directly demonstrated by our data. We now explicitly state that without concurrent ambient NO₂ and VOC concentration measurements, these mechanisms cannot be rigorously validated.

Added a limitation statement. We acknowledge that the absence of in-situ NO₂ and VOC concentration time series limits our ability to directly link observed ozone changes to specific precursor pathways, and that future studies incorporating such measurements or chemical transport modeling are needed to test these hypotheses.

We believe these revisions make the boundaries of our evidence transparent while preserving the value of the observational patterns we report.

Minor Comments

- Several sentences are awkward or unclear. Examples: “the overall exposure time to ozone has shifted to later over the years” to “has shifted to later in the day”. A professional language edit is recommended.

Response: Thank you for your kindly reminder. We have revised the language by professional editor.

- Table 1: The effective rate of “typical” days after weather filtering ranges from 54% to 77%. The low rates (e.g., 54% in 2021) raise concerns about representativeness. Discuss potential bias.

Response: Thank you for your valuable comments. We have had a data quality control in the Method section 2.2.1 to remove the weather affected days for each year. The length of Ozone Season in 2021 is least and the effective is lowest of all. Though the ozone trend in 2021 is different from the other years, we think the special climate of that year playing the most important role. We analysis the temperature during Ozone Season between 2013 and 2023, showing the lowerest temperature in 2021. Ozone MDA8 in 2021 was also the lowerest of all years as well as percentage of ozone as shown in Figure below. So the low rate is not effect the ozone trend of 2021.

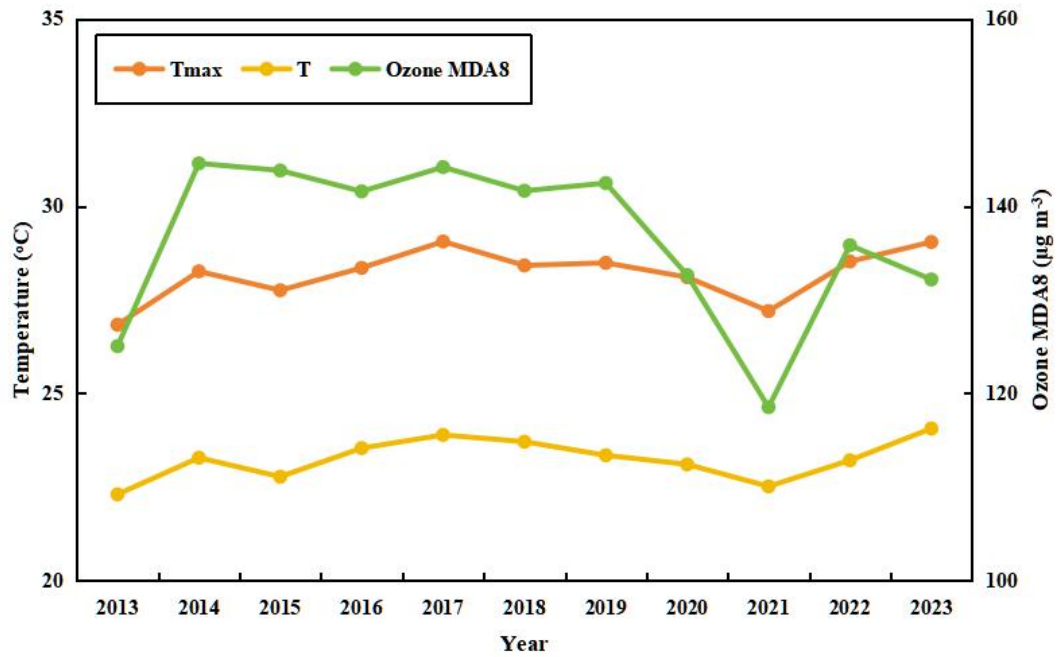
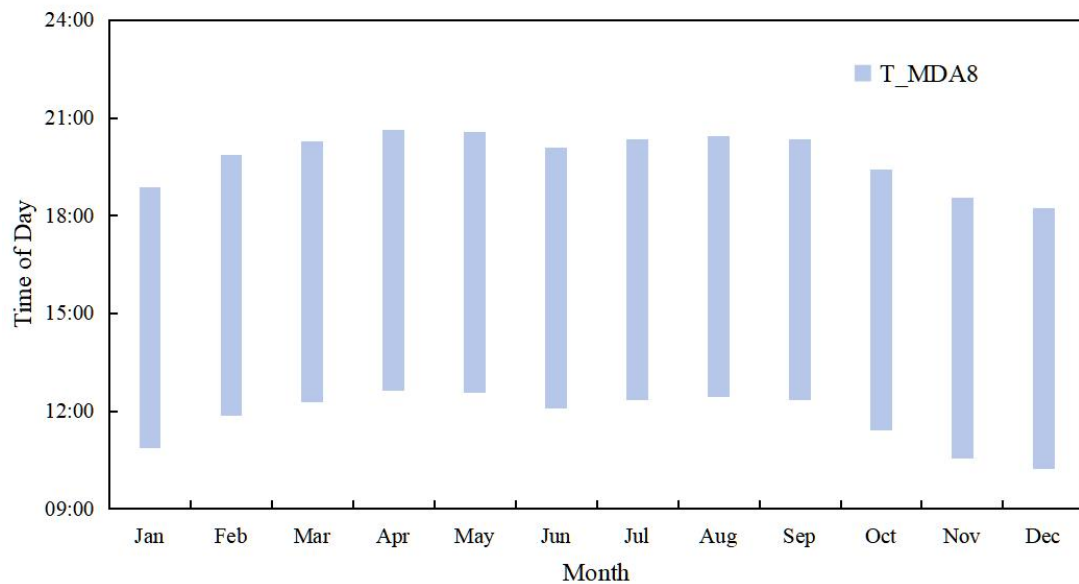


Figure Temperatures and MDA8 during 2013 to 2023 (Tmax for average daily max temperatures, T for average daily temperatures)

- Figure 4: The y-axis is time of day, but the caption is unclear. Clarify that colors represent frequency or concentration. Add units.

Response: Thank you for your valuable comments. We have revised the y-axis, and added the description of the Fig.4.



- Figure 5: why there are hot colors at 1:00 am.

Response: Thank you very much for this perceptive observation. We sincerely apologize for not having addressed this phenomenon more explicitly in the original manuscript, and we appreciate the opportunity to clarify it here.

During our data analysis, we indeed identified cases where the MDA8 window fell between 1:00 and 8:00 local time. This typically occurs under a specific combination of conditions: elevated ozone levels from the previous day persist into the nighttime hours (residual ozone), while the following day experiences improved atmospheric dispersion, such as stronger winds or more effective boundary-layer ventilation, that keeps daytime ozone concentrations comparatively low. Under these circumstances, the 8-hour running mean reaches its maximum during the early morning rather than during the typical afternoon photochemical peak.

We quantified the frequency of such occurrences to assess whether they might affect the reliability of our results. These nighttime MDA8 events account for approximately 3–10% of days in the cold season (with the highest proportion observed in 2015) and 1–7% in the warm season (peaking in 2013). Given their relatively limited overall frequency, we believe they do not substantially alter the climatological diurnal patterns or the interannual trends presented in our study. Nevertheless, we have now added a brief discussion of this phenomenon to the revised manuscript to improve transparency and completeness.

- As temperatures rise, this exposure window shifts, with the duration extending from 14:00 to 21:00 during the months of April to September. 14:00 to 21:00 seems not match the figure.

Response: Thank you very much for this careful observation. We sincerely apologize for the error in our original description of the exposure window timing. Upon re-examining the data, we confirm that during the warm season, the exposure window is indeed markedly delayed compared to the cold season, with the duration primarily occurring between 12:00 and 21:00. We have corrected this inaccuracy throughout the revised manuscript, and we are grateful to the reviewer for bringing this to our attention.

- How the data in some of figures are not clearly explained. Many figure captions are not self-explanatory.

Response: Thank you for your valuable comments. We have revised the text.

- Some citations are incomplete or inconsistent (e.g., “Beijing Municipal Bureau of Ecology and Environment, 2024” should be formatted as a proper reference). Check journal guidelines.

- In the conclusion, the statement “The decrease in peak concentrations benefits from coordinated emission reduction policies for regional VOCs and NO_x (Wang et al., 2022)” includes a citation. Citations are generally not recommended in the conclusion section, as conclusions should summarize findings without introducing new references.

Response: Thank you for your valuable comments. We have revised the citation problems.