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 (PM2.5, PM10). In northern China, for example, the PM pollution season peaks in autumn and winter, whereas in southern China PM2.5 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 PM2.5-centered approach to a coordinated PM2.5-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 PM2.5.

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 70-74:

> "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 captureing 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., NOx, 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 NOx 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 (Section 3.7) 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).

Removed speculative conclusions attributing specific ozone changes directly to NOx 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.

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 138–152) 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 used to assess outdoor ozone exposure levels(Lu et al., 2018; Niu et al., 2022; Byun et al., 2022). MDA8 is defined as below.

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

C_i is the hourly ozone concentration (μg/m³) at hour, and i is the starting hour of the sliding window, ranging from the 1st hour (01:00) to the 16th hour (16:00), yielding a total of 16 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 t1,t2, t3…t8, 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, is used to represent ."

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(Vi) ,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 362-369, stating that the calculated hourly change rate represents the total (net) ozone derivative observed at the study site. As supported by Kaser et al. (2017), this quantity is the sum of all physical and chemical processes contributing to the local ozone budget, including entrainment through boundary-layer growth, horizontal advection, deposition, and photochemistry. The use of hourly observed differences to derive ozone tendency has been widely adopted in atmospheric observation studies, including boundary-layer process analyses (Kaser et al.,2017; Klein et al.,2013) and urban ozone diurnal characterization (Gu et al., 2024).

Furthermore, we have clarified that the effect of advection can be substantially mitigated by the application of long-term averaging for diurnal trend analysis. Specifically: (1) the Shanghai Tower observational study (Gu et al.,2024) explicitly states that "the effect of advection can be mitigated by the application of a monthly average for the diurnal trend analysis"; (2) Liu et al. (2021) demonstrated that synoptic-scale cycles (< 33 days) can be effectively removed by KZ filtering, leaving the baseline seasonal component; and (3) European long-term trend assessments employ multi-year monthly averages to minimize interannual variability. These studies collectively support our approach of using multi-year monthly mean diurnal cycles to filter out synoptic-scale transport perturbations and highlight climatological ozone behavior. All cited studies explicitly treat the quantity as an observed net tendency subject to multiple process contributions, consistent with our revised terminology.

References

Kaser, L., Patton, E. G., Pfister, G. G., Weinheimer, A. J., Montzka, D. D., Flocke, F., Thompson, A. M., Stauffer, R. M., and Halliday, H. S.. The effect of entrainment through atmospheric boundary layer growth on observed and modeled surface ozone in the Colorado Front Range. Journal of Geophysical Research: Atmospheres, 122(11), 6075–6093. https://doi.org/10.1002/2016JD026245, 2017.

Klein, A., Ravetta, F., Thomas, J. L., and Ancellet, G. A lagrangian analysis of lower tropospheric ozone variability during the ESCOMPTE campaign. Journal of Geophysical Research: Atmospheres, 118(24), 13, 776–13, 791. https://doi.org/10.1002/2013JD020146,2013

Gu, Y., Li, Y., Li, J., Li, H., Wang, Y., Liu, Y., Li, H., Liu, J., Guo, W., and Li, W. (2024). Ozone vertical profile and its formation regime at the top of the planetary

boundary layer in a coastal megacity of eastern China. Science of The Total Environment, 921, 170506. https://doi.org/10.1016/j.scitotenv.2024.170506, 2013.
Liu, Y., Wang, T., and Weng, X. Quantifying the separable meteorological and non-meteorological contributions to the 2013 – 2018 tropospheric ozone variation over China. Atmospheric Chemistry and Physics, 21(19), 14789 – 14804. https://doi.org/10.5194/acp-21-14789-2021, 2021.

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 section2.2.2, and improved the MK test in ozone percentiles in line 221-223 and the multi-year evolution of diurnal concentration patterns (hourly mean concentrations) in line 335-341.

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.

6. The manuscript frequently refers to the impacts of NOx, 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) 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 educe spacing.