Review comments on a manuscript entitled “Extending ozone and particulate matter pollution control from New York City to Beijing” submitted by Zhang et al.

In this study, the authors presented an observational analysis on the relationship of O3 with PM2.5, the possible factors determining the ratios of O3 to PM2.5 and their indication to development of emission control strategies in two megacities, New York City (NYC) and Beijing. This is an interesting since both cities have been experiencing different stages of emission control implementations. In other words, the emission control measures implemented by New York City may have certain reference significance for the formulation of emission control strategies for mitigating O3 and PM2.5 pollution in megacities such as Beijing in the future. The authors linked the O3-PM2.5 relationship with aerosol compositions and highlighted the importance of an inflection points (IFP) in development of effective emission control strategies on reduction of O3 and PM2.5 synchronously. These are two interesting points which are worth to study. However, there are too many problems in the current version in terms of science interpretations and English writing. I listed part of them below. Overall, I do not think the findings and quality of the manuscript in the current version are strong enough to support its publication by NPJ. Here are the major and specific comments.

Major comments

1. People used to rely on O3 formation regimes (i.e., VOCs- or NOx-limited) as a scientific evidence in support of development of effective emission control strategies on reduction of O3 ambient levels near surface, whereas this study highlighted the importance of IFP for that? This is a good try. I am curious which of these two is more meaningful for policy makers. What is(are) the possible relationship(s) between both? Any discussions or suggestions between both are very helpful.
2. IFP is meaningful and interesting. How do the IFP values vary with locations and time (e.g., seasonally or yearly)? What are the key factors in determining the value of IFP?
3. L60-61: Definition of extreme concentration is confused. What time period does the averaged top 5% value represent? Daily, monthly, seasonally or yearly? What are the typical values of O3 and PM2.5 extremes represent in NYC and BJ?
4. Line 85: It is very confused with the statement of “A steeper slope of O3 and PM2.5 is expected as the SAP mass fraction in PM2.5 decreases because more nitrate and SOA in PM2.5 as a result of higher NOx and VOCs that are capable of producing more ozone”. Which one is more important to determine the slope of O3 to PM2.5? In addition, what is the main difference between SAP and SOA?
5. L88-102: An adjusted mass concentration is proposed to demonstrate the contribution of the reduced SAP to the increase in O3-PM2.5 ratio. Such a checking is helpful but does not provide additional scientific discussion and insights. I suggest moving it to Supplement.
6. L124-128: Aerosol radiative effect has an important impact on O3 formation near surface under heavy PM2.5 polluted events (e.g., Ma et al., 2021). This could be another important mechanism that the authors need to consider.
7. Impact of SOA and SAP on O3-PM2.5 relationship: The authors pointed out that SOA is the dominant species in PM2.5 on Line 80 and then said that “sulfate … and primary organic aerosols (POA) could affect the slope of O3 versus PM2.5 (see Lines 81-82)”. This is confusing. Which one is more important in determining the slope of O3 versus PM2.5? Are there any observations to support that?
8. L130-132: The explanation is very similar to the one proposed by Li et al. (2019) while you used a regional air quality model, CMAQ and Li et al., used a global model, GEOS-Chem. Any more new thing is found beyond that?
9. L164-165: Why do you show the CMAQ simulations with reduction of SO2 emission rather than O3 precursors such as NOx emissions?
10. Method (M2, L263-266): If the initial conditions represent a clean continental condition, three-day simulations are not sufficient to represent the spin-up runs.
11. Writing issue: It is not easy to understanding the manuscript. A heavy edit is required. This includes grammar checking, the logic and clarity of sentence expression.

Specific comments

1. L45: Perhaps this is the first time I've seen DA24 being used to represent a 24-hour average in the communities of air quality and atmospheric chemistry.
2. L69: In the manuscript, the study period was separated into 4 sub-periods based on the variability of DA24 PM2.5 concentrations. What does the variability represent? Please provide specific criteria of defining these sub-periods.
3. L 74-76: What time periods do these MDA8 O3 and DA24 PM2.5 extreme concentrations represent? What are these extremes values?
4. Line 82: Why do the authors use SAP rather than POA as the abbreviation of primary organic aerosols?
5. L108: This is another example that confuses readers. I do not think this is a correct argument. Higher VOC+NOx emissions do not guarantee higher O3 production. You have to consider the dominant O3 formation regime in the region(s).
6. L111-114: Again, it’s hard to understand this sentence.
7. L116: This is not accurate sine concentrations of O3, another major air pollutant have been increasing in BJ since 2013.
8. L119: What are the criteria to define the periods?
9. L122: Again how are the DA24 PM2.5 mass extreme concentrations defined? What time periods do the values of 159 and 90 represent? The same question for O3 extreme concentrations and their values on L 123.
10. L124-125: If argument 1) is true, why it did not cause an increase in NYC during the similar time period?
11. L124: Propose -> proposed?
12. L125: The statement of “meteorological variabilities” is too general. More specific evidences are required.
13. L153-156: It is not easy to understand the sentence. Please rewrite it. Similar issues can be found in many other sentence throughout the manuscript.
14. L68, 115, 166: A period is not necessary in the subtitles.
15. L122-124: The order of figures is not consistent with the discussion presented in the text. It seems a little bit separated from each conclusion you draw in the paragraph.
16. L129: Which figure or table support this conclusion? Please specify.
17. L168: is 🡪 are
18. L196: “if the same policies lasting”?
19. L22: there are many assumptions made in the method 1. Do those assumptions affect the conclusions? Any discussion on this?
20. L484, 487: Why does SP2 cover 2016-2019 for both YRD and PRD, but 2017-2019 for BJ?