The subject manuscript by Zhang et al. presents an analysis of ambient measurements from the Yangtze River Delta assessing the observations of increased ozone (O_3) levels during the COVID-19 lockdown period. Results are presented for three periods: pre lockdown, full lockdown, and partial lockdown and are compared with the same time periods in 2019. The authors seek to understand the relative importance of precursor volatile organic compounds (largely grouped by compound class), nitrogen oxides (NOx), ambient reactivity/oxidation capacity, and meteorology in determining ozone levels. The authors motivate the study well, and adequately convey the importance of measurements and analysis in this region. However, the methodology is not sufficiently described to support the results and conclusions, and the results and conclusions are somewhat difficult to follow as written. Specific questions, comments and suggestions on these points follow below. It is recommended that this manuscript be reconsidered for publication after major revisions.

Response: We thank the reviewer for the helpful comments and suggestions. We have carefully addressed all the comments and suggestions. Below is our point-by-point response to each comment, marked in blue. Changes made to the main text are presented in green.

Technical Comments:

line 34, lines 173-176: Using “supposed to” in this sentence does not necessarily reflect the complexity of O_3 formation. The prior sentence suggests that during the full lockdown period, the region shifted to a NOx-limited regime. Thus, it may be expected that a greater decrease in NOx relative to VOCs would lead to a decrease in O_3. However, there is also the role of NOx titration, in which decreasing NOx can lead to an increase in O_3. Later in the manuscript, this phrasing is repeated and it suggests that the authors are not necessarily referring to the NOx regime in the abstract, but the influence of meteorology (specifically T, RH). Again, this is not clear in the abstract, and oversimplified as written. What is the mechanism by which RH affects O_3 formation? How sensitive is O_3 to RH?

Response: (1) We agree that using "supposed to" in this sentence does not necessarily reflect the complexity of O_3 formation. The obvious increase of O_3 during the Full-lockdown period was caused by the joint effect of meteorology, changes in emissions and chemistry. Therefore, we have revised "supposed to" to "in favor of". From Pre-lockdown to Full-lockdown period, the average NO_x and VOCs concentrations decreased by 62.6% (20.0 ppbv) and 32.2% (10.5 ppbv), respectively, while O_3 concentration increased by 67% (12.4 ppbv). This is attributed to the fact that the O_3 formation was VOC-limited during Pre-lockdown period, and more abatement of NO_x than VOCs would lead to increase of O_3. If the over 80% of NO_x and over 45% of VOCs were eliminated, like the case of 16th Feb and 17th Feb, the O_3 formation would switch to the NO_x-limited regime and lead to decrease of O_3.

Therefore, we have revised relevant description in the abstract and discussion, please refer to Page 2, Line 32-34:The NO_x/VOCs ratio dropped dramatically from 1.84 during Pre-lockdown to 0.79 in Full-lockdown period, which switched O_3 formation from VOCs-limited regime to the boundary of NO_x- and VOC-limited regime.".

(2) The mechanism by which RH affects O_3 formation is very complicated. Higher humidity is conducive to OH production and thus likely increase O_3 production. We have added some references about RH influence here and simulated the influence of RH on ozone by box model. Sensitivity analysis has been performed to reveal this effect by increasing or decreasing RH by 10%, and the results are exhibited in Figure 1. On average, decreasing RH by10% leads to 0.28% increase of the simulated O_3, and this influence dropped to -0.35% when RH was 5% lower than the base case. When increasing RH by 5% or 10%, positive influence on the simulated O_3 was found.

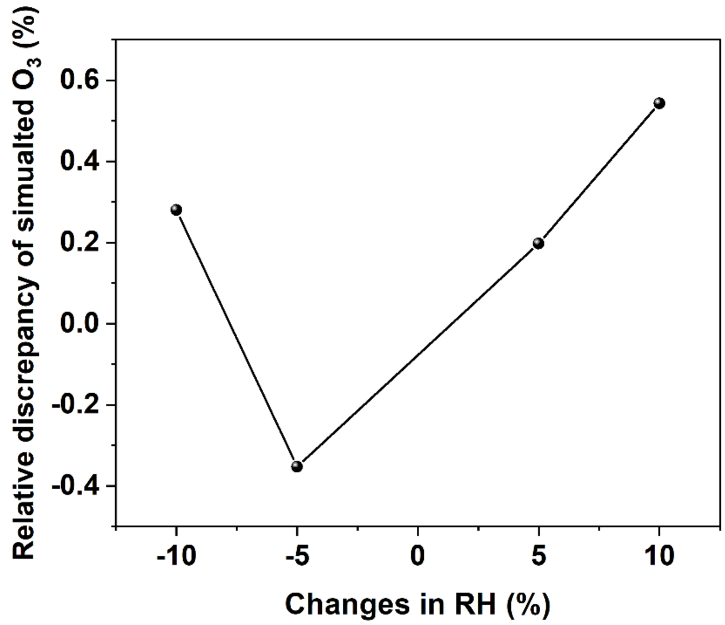

Figure 1. Sensitivity analysis of the influence of RH on simulated O_3

We have revised the manuscript accordingly, please refer to Page 9, Line 198-201:

The relatively higher T was in favor of O_3 formation during the Full-lockdown period in 2020. As for RH, the influence on O_3 is nonlinear (Zhang et al., 2020), and based on our sensitivity test, lower RH could lead to decrease or increase of O_3 concentration (Figure S2).

Similar to the use of "supposed to", the authors need to clarify "improper" decline (line 202) and "abnormal" increase of O3 (line 203).

Response: Thanks for the helpful suggestion. The "abnormal" represent the obvious higher O_3 concentration in the Full-lockdown period in 2020 than that during the same time in 2019. To avoid misunderstanding, we have replaced "abnormal" with "obvious", and revised relative description in the manuscript, please refer to Page 9, Line 192-195:

It should be noted that, compared to Full-lockdown period in 2019, the mean O_3 concentration in 2020 is obviously higher (5.5 ppbv, **Error! Reference source not found.**). Meanwhile, the average O_3 concentrations in Full-lockdown period in 2020 was 67% higher than that during Pre-lockdown period in 2020.

PTR-TOF-MS measurements (p. 5): Does the Jensen et al. companion paper address losses in the inlet and to the filter? How might these losses affect the results of the analysis presented here? The authors do not need to provide all of the details presented in Jensen et al., but should summarize the main findings, including any limitations, that are relevant to the analysis presented in this manuscript.

Response: The Jensen et al. companion paper did not address the loss in the inlet and the filter. During the observation, a 3-m long PTFE tube was used as inlet, and no strong loss was supposed. In addition, standard gases were used routinely for the calibration. Therefore, we believe the results from PTR-TOF-MS is reliable. Additionally, the main findings and limitation of Jensen et al. (2021) that are relevant to this manuscript have been added in the revised manuscript, please refer to Page 12, Line 236-238:

In addition, Jensen et al. (2021) found the VOC emissions from most industries in Changzhou share the same "U-shape" trend as our study."

Trend analysis: The authors state that the MK non-parametric test is recommended by the WMO. The authors should provide some additional detail here. What does the WMO recommend this test for? Under what conditions? What are the limitations/requirements for applicability in the context of this work? How is serial correlation applicable to the PTR-TOF-MS measurements of individual VOCs? What details of Pathakoti et al. and Alhathloul et al. are relevant here?

Response: According to Adeloye and Montaseri (2002) and WMO (1998), the Spearman Rank Order Correlation (SROC) instead of MK test is recommended to investigate the long-term trend of flow volume, and we have revised this sentence in the manuscript. We select MK test because it is a non-parametric statistical method and doesn't require any assumptions regarding the probability distribution of the data. The MK test has a limitation that the input data should have no serial correlation. The serial correlation of individual VOC has been tested by the "feasts" R package, and no serial correlation is found for each individual VOC. The paper of Pathakoti et al. and Alhathloul et al. gives the details of the calculation MK trend test. Relevant description has been revised, please refer to Page 7, Line 145-148:

By using the "feast" R package, no obvious serial correlation of individual VOC is found. Therefore, the observed VOC data is suitable for MK test. Detailed description and the calculation formula of MK trend test could be found in the study of Pathakoti et al. (2021) and Alhathloul et al. (2021).

Deweathered model: While details of the VOC measurements are somewhat lacking, and more so for the trend analysis, this section is entirely lacking of sufficient detail (and is not, as the authors note on line 191, described in section 2.4). The authors should consider that the information in the manuscript needs to be sufficient such that the results can be reproduced. Further, it is difficult to assess the robustness of the results when sufficient details about the methodology are not provided. What are the uncertainties of the approach? Are data available for all parameters over all time periods? How are missing data handled? How were the model parameters determined (number of trees, minimal node size, and number of samples)? How sensitive are the results to these model parameters?

Response: We agree that the details of deweather analysis is insufficient. All parameters used in this model is available over all period and a small amount of missing data is replaced by linear interpolation. The number of trees (ntree), number of samples (nsample) and minimal node size were selected by the sensitive analysis (Table 1 and 2). According to Table 1 and 2, when ntree, nsample and minimal node size were chosen as 300, 300, and 5, respectively, the

R^2 of the deweather model is the highest. In addition, the R^2 of the deweather model is not sensitive to the choose of ntree, nsample, and minimal node size (Table 1 and 2). The uncertainty of the deweather model is obtained by growing 50 random forest models with the nree, nsample and minimal node size chosen as 300, 300, and 5, respectively, which is the same method as Grange and Carslaw (2019). The mean and standard error of the predicted O_3 concentrations is shown in Figure 2, and results of the model are stable during the 50 runs.

We have added the details of the deweather model in the revised manuscript, please refer to Page 8, Line 160-177:

Hourly data of Unix date (number of seconds since 1970-01-01), Julian day, weekday, hour of day, wind speed (WS), wind direction (WD), temperature (T), relative humidity (RH), and pressure (P), which are available during the whole observation, were used for the deweathered calculation of O_3. The missing data was replaced by linear interpolation. Training of the models was conducted on 80% of the input data and the other 20% was withheld from training. To avoid the disadvantage of overfitting during the training of RF, a process called bagging (or bootstrap aggregation) was adopted. Bagging results in new, sampled set called out-of-bag (OOB) data. A decision tree is then grown on the OOB data. Therefore, all the decision trees are grown on different observations and avoid the overfitting (Grange and David, 2019). To determine the value of number of trees (ntree), number of samples (nsample), and the minimal node size, a series of random forests were performed under difference choice of ntree, nsample, and minimal node size. The results suggest that the highest coefficient of determination (R^2, 0.84) was obtained when ntree, nsample and minimal node size was set as 300, 300, and 5, respectively (Table S1 and S2). More details of this model could be found in the study of Grange and David (2019). The uncertainty of the deweather model is obtained by growing 50 random forest models with the hyperparameters described above, which is the same method as

Grange and Carslaw (2019). The mean and standard error of the predicted O_3 concentrations is shown in Figure S1, and results of the model are stable during the 50 runs.”

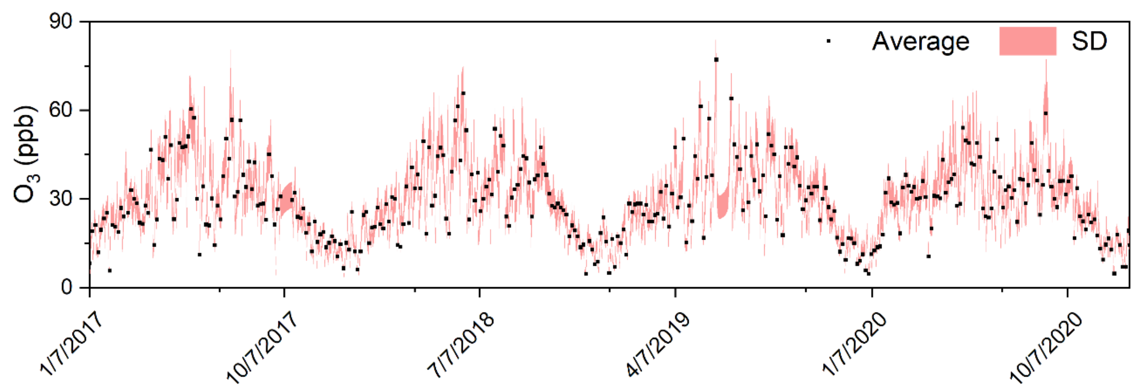

Figure 2. The mean and standard error of predicted O_3 concentrations.

Table 1 R^2 of the deweather model with different choose of ntree and nsample.

ntree\nsample	100	200	300	400	500
100	0.852	0.853	0.853	0.852	0.852
200	0.855	0.855	0.856	0.855	0.855
300	0.856	0.857	0.858	0.856	0.856
400	0.857	0.857	0.857	0.857	0.856
500	0.857	0.857	0.857	0.857	0.857

Table 1 Influence of the choose of minimum node size on R^2 of the deweather model.

minimal node size	1	2	3	4	5
R^2	0.860	0.857	0.858	0.858	0.859
minimal node size	6	7	8	9	10
R^2	0.855	0.853	0.852	0.851	0.849

In general, in the results and discussion, it is often difficult to follow whether the authors are describing results between the three periods in 2020, or between given periods in 2019 and 2020. It is recommended that the authors try to more clearly differentiate these comparisons.

Response: Thanks for the helpful suggestion. The comparisons in the results and discussion have been clearly differentiated as suggested.

line 214: TVOC dropped to 22.19 ppb from what mixing ratio?

Response: This sentence has been revised to:

Full-lockdown period, the TVOC dropped to 22.19 ±7.9 ppbv from 32.78 ± 13.81 ppbv, which was mainly affected by the decrease in industrial activities and traffic volume.

lines 218 and 234: The authors use “interesting” in these sentences, but it is not clear what is interesting about the observations as presented. The higher mixing ratios due to lower boundary layer heights (line 218) is a common observation, and the lower values of transportation-associated VOCs (line 234) during the lockdown is expected (and has been reported previously).

Response: Thanks for the suggestion, we have removed “interesting” in these sentences.

In line 233, is the decreasing trend based on the Z-score or Q value? Were these metrics consistent? Why or why not? It might be useful to include the Z-scores and Q values for all compounds in the SI.

Response: The decreasing trend is based on Q value or Z score, which has been mentioned in the manuscript. The Z value and Q value are consistent because both indicate the trend of a time series. But the Q value is usually used for the quantification of the rate of the trend. The Z score and Q values for all compounds are included in the SI.

line 224-226: Can the authors be more quantitative about how many of the measured VOC species shown exhibited this U-shape pattern and then explicitly list those VOCs that didn’t? It is a little contradictory to say “most” and then “except for several”.

Response: Thanks for the helpful suggestion. We have revised relative description, please refer to Page 13, Line 248-259:

Total 42 VOC species exhibited an 'U' shape trend during the whole observation, while formaldehyde (HCHO) and methanol showed an obvious increasing pattern.

line 269: Is it expected that biogenic emissions would be the dominant source of O_3 in this region?

Response: This site is in the urban area of Changzhou city, where the VOCs are dominated by anthropogenic sources. Hence, the dominant source of O_3 in this region is not expected to be biogenic emissions. To avoid misleading, we have removed this sentence.

lines 273-275: This discussion of alkenes is not clear as written. The chemical reactivities of compounds is tied to their oxidation formation potential (and is not independent of).

Response: Thanks for the comment. The discussion of alkenes has been revised, please refer to Page 17-18, Line 296-299:

However, the OFP of alkenes and OVOCs only decreased by 8.9 and 22.5 μg/m^3, respectively. During the observation, the most abundant alkenes measured by PTR-TOF-MS are 1-hexene and isoprene, with the k_{OH} of 37 and 100 $\times 10^{-12}$ cm^3 molecule^{-1} s^{-1}, respectively, which are much higher than that of the most abundant aromatics (1.22, 5.63, and 17 cm^3 molecule^{-1} s^{-1} for benzene, toluene, and xylene, respectively).

line 279: How are the MIR values calculated? What are they dependent on? Is it expected that the MIR would be reflective of the different NOx/VOC regimes? I'm not sure this is the case.

Response: The mean MIR was calculated by dividing the total OFP by the total concentration of VOC, this parameter depends on the individual MIR and concentration of each VOCs. The mean MIR could not reflect the different NO_x/VOC regimes, while it can represent the ability of VOC species composition to produce ozone. The relevant description has been added in the revised manuscript, please refer to Page 17, Line 304-307:

To compare the average reactivity of VOCs during different periods, we calculated the mean MIR, derived by dividing the total OFP by total VOC concentration, in each period, and a higher MIR means stronger capability of VOCs to produce ozone.

line 286: Was NO_x eliminated (which suggests some chemical/physical removal)? Or were the emissions reduced to a greater extent that VOC emissions?

Response: To avoid misleading, the sentence has been revised to "suggesting more NO_x was reduced than VOCs during Full-lockdown period".

line 326: What is the relevance of the stable OVOCs across the lockdown periods in the context of emission sources/anthropogenic activity? Is there any offset between emissions and chemistry during this period?

Response: Thanks for pointing this out. The original description has a mistake. After double check of the data, we found the simulated OVOC concentrations during the three periods were 26.65, 20.75, and 23.80 ppbv, respectively, which follows the same trend as k_{OH}. This sentence has been revised to:

As k_{OH} from OVOC, it shared the same trend as OVOC concentration, which reached the minimum value (5.56 s^{-1}) during the Full-lockdown period.

Added References:

Adeloye, A. J. and Montaseri, M.: Preliminary streamflow data analyses prior to water resources planning study. Hydrological Sciences Journal 2002; 47, 679–692.

WMO: Analyzing long time series of hydrological data with respect to climate variability, World Meteorological Organization, Geneva, Switzerland 1988.