Report #1:

Revised submission-- "Mobile monitoring of urban air quality at high spatial resolution by low-cost sensors: Impacts of COVID-19 pandemic lockdown"

The authors have considered responses to each of the points raised by the reviewers. However, there are key points at the base of earlier comments that are still quite unclear. These include: 1) specifics on the sensors used as basis for data source and 2) the implications of sensor choice, and 3) the implications of sensor calibration protocols on the data and its use in a traffic impacted environment.

Re: We thank the reviewer for the further comments. Please see our responses to these three points below.

A final, perhaps less important general concern is with the title of this paper. The title clearly shows this was a study of the impacts of COVID lockdowns. This event was a "nice" add-on, but it is not at all a major part of the effort. I suggest a title that better reflects the nature of the study.

Re: Our observation covers different stages of the COVID-19 pandemic and also reflects the impact of the COVID-19 pandemic on the air pollutants variations. This is also one of the main objectives of our research, so we think it is suitable to keep the title as is.

The overall concern is that while the modelling and allocation of pollutants to roadways are interesting demonstration of mobile sensor-based monitoring it is unclear that actual and accurate pollutant data represents air quality on roadways. And while the calibration data from a Nanjing University campus show good performance it is not clear that this translates to accurate data from on road measurements. And without further information on sensors and the data they produced for this study it is not possible to evaluate or accept the modelling results.

Re: Please see our response below.

Key points

1. Specifics on sensors

The current version now adds a few words to identify the sensors as electrochemical cells, but there are many possible sources of cells of this type, and they differ. Please provide the make and models of the sensors. And fully describe any special characteristics that apply to the sensor calibration and validation in this study. For example, do they include any chemical filtration?

Re: We checked with our instrument vendor, but they only released the internal model number of these sensors and kept the ultimate make and models in confidence. However, we think this is acceptable as we conducted regular calibration and validation processes and the results indicate the sensors worked reasonably well (Section 2.2).

We clarified this by adding the internal model numbers of the sensors in line 66-67:

"The instrument is equipped with internal gas sensors for **CO (model XH-CO-50-7)**, **NO_2 (XH-NO2-5AOF-7), and O_3 (XH-O3-1-7)** (dimensions: 290 × 81 × 55 mm; weight: 1.0 kg) …".

and some sentences to describe the sensor calibration were added in line 94-99:

"For example, environmental conditions are known to cause nonlinear behavior as well (Popoola et al., 2016). Due to aging and impurity effects over a long time, low-cost sensors are prone to signal drift and low sensitivity (Kizel et al., 2018). In addition, cross-sensitivities differ largely according to the ambient temperature and level of gas the sensor is being exposed to (Lösch et al., 2008). So, multi-parameter joint calibration training is utilized to reduce the interference issue between air pollutants, including air pollutants concentrations, temperature and relative humidity".

2. Implications of sensor choice

Electrochemical sensors are not fully chemically specific in response. This complicates their use in complex atmospheres—such as near roadways due to established interactions between ozone and nitrogen dioxide. Most "ozone" electrochemical sensors are "oxidant" sensors since they respond to oxidants (especially ozone and NO_2). Data from them should not be simply viewed as ozone, as is the case in this study, since it is performed on NO_2-rich roadways where ozone is likely to be disproportionately lower than in ambient air. NO_2 electrochemical sensors may also have interference issues. Without clear identification of the sensors and a description of any features or data correction steps taken, it is not possible to understand the nature of data from either the "ozone" or NO_2 sensors. Use of data from these sensors in subsequent plotting and modeling are open to considerable uncertainty. The reviewer is concerned of the use of sensor data as an off-the-shelf product without sufficient quality control and attention to details.

Re: We agree with the reviewer that the sensor data cannot be used as off-the-shelf product. We therefore have a regular sensor calibration and validation process during our study. We think this is an acceptable practice as the results indicate **the sensors work reasonably well (section 2.2)** and **the uncertainty is explicitly acknowledged**.

We acknowledged the cross-interference between NO_2 and O_3 in line 131-132:

"…, the interaction between O_3 and NO_2 may influence the detection accuracy of these two chemicals, especially for NO_2 (Ivanovskaya et al., 2001)".

We acknowledged the uncertainty of our measurements in variability analysis in line 140-141: "Overall, the sensor results have substantial uncertainty compared to reference methods, we thus focus on the relative temporal and spatial distributions rather than the absolute concentrations".

The data collected while on road may be quite different than those from ambient community sites. This could complicate ozone reported. The authors should describe how ozone data was produced from the output of sensors and whether they considered interferences due to NO_2. And depending on the NO_2 sensor, they should address interferences that may occur for that sensor in the atmosphere under study. Because the accuracy of these two data streams is essential to the overall study, the details presented at adequate depth to inform the reader of what was done.

Re: We indeed considered the interferences between NO_2 and O_3. A sentence was added to clarify how the data was produced in line 76-77:

"Pollutant concentration data is generated by different voltage values sensed by gas sensors, which is automatically uploaded to a database in the cloud via the 4G telecommunications network".

We also clarified this by adding a sentence in line 96-99:

"cross sensitivities differ largely according to the ambient temperature and level of gas the sensor is being exposed to. So, multi-parameter joint calibration training was utilized to reduce the interference issue between air pollutants in our research, including air pollutants concentrations, temperature and relative humidity”.

3. Calibration issues
Calibration is a key activity to assure good data from sensors. In this case the calibrations were performed using periodic co-location at an air monitoring station located at a campus of Nanjing University that may be distant from the urban center. This is possibly an acceptable approach, however some studies have shown the calibration results may not be transferrable especially crossing different concentration ranges or different environments. Expanded consideration of environmental conditions included in the calibration and how the change of environmental conditions can affect the are needed to help ensure credibility of data when sensors are deployed in the field. Again, back to sensor selection and data issues— The mix of ozone and NO_2 are likely to be quite different between the roadways and the fixed site. Are there near-road ambient air monitoring sites in Nanjing that can show NO_2, NO an O_3 data levels that can be compared to the university site (which seems quite far from busy roadways)? How do they differ? How might these differences impact the utility of the calibrations performed?
It is important in sensor-based papers, that calibration as part of robust quality control and assurance is evident to ensure sufficient confidence for data interpretation or data fusion with modelling results. Overall, these topics are weakly considered in the manuscript.
Re: We thank the reviewer to bring this up. However, we do not have access to a regular sampling site near the road that uses reference method to measure these pollutants. All the national network sites were deliberately positioned far away from roads to represent a regional background.

We acknowledge this drawback in the revised manuscript in line 106-108:
“One drawback of our study is that the air pollutants concentrations observed at SORPES are lower than those observed in a road environment, which might cause issues for the calibration process”.

One further point is identified that is separate from the above point regarding sensors and sensor data. That is with regards to the use of data from these two taxis to represent an entire urban area road grid. It seems that in many of the observed roadway segment data were collected perhaps once or twice over a one- or two-month interval while in other cases several observations may have occurred on a single day. The reviewer is not convinced this small data can be representative enough to be compared with modeling results to draw meaning conclusion. How might the nature of the temporal (even seasonal) nature of data quality impact the representation of the city roadways? Are two mobile monitors enough for a city the size of Nanjing?
Re: We indeed notice that the trajectories of the two taxis cannot fully cover the entire city, especially over less-populated regions. The re-visiting frequency of some grid points are quite low (Fig. 5b). However, most (~ 70 %) of the grid points have a sampling frequency > 50. It indicates that the dataset may be good enough to represent a long-term average, despite that more sensors are indeed needed if we want to pursue an air quality map with both high temporal and spatial resolutions.

We clarified the re-visiting frequency on the main roads in the revised manuscript (line 207-209): "A total of 1.32 million pieces of data were obtained during the observation period, which covers 66.4 % of the major roads in Nanjing in the sampling domain with a large repeat-visit frequency [median repetition = 61 (14 and 264 as the lower and upper quartiles, respectively, the same hereinafter)] (Fig. 5a)".

and a sentence to clarify the re-visiting frequency in each grid was in line 213-215:
"As shown in Fig. 5b, the median number of repeated frequency in each grid is 66 (18, 286), with the highest value of 15449 in Nanjing South Railway Station and the lowest in some residential roads (1)".

Detailed comments
Line 77—states that taxis with natural gas and electric power were used. It is unclear how many taxis, but line 67 says that 2 were used in this study. If this is correct, may we assume this means one each?
Re: It was clarified in line 69-70:
"**Two** taxis fueled with electricity and liquefied natural gas **(one each)** are selected to reduce the impact of emissions from the sampling vehicles themselves".

Section 2.2, line 90—the authors list important characteristics to be considered with use of sensors. The list does not include specificity and interference issues or adequate details/citations on these matters. Overall, section 2.2 needs considerable editing and clarification. Calibration is a key point for this study.
Re: We clarified the interference issues by modifying the sentence in line 92-94:
"Different from traditional instruments, low-cost sensors have some limitations, such as nonlinear response, signal drift, environmental dependencies, low selectivity, and **cross-sensitivity**, so it is important that calibration procedures are applied with respect to these limitations (Maag et al, 2018; **Lösch et al., 2018**)."

And some sentences were also added in section 2.2 in line 94-99:
"For example, environmental conditions are known to cause nonlinear behavior as well (Popoola et al., 2016). Due to aging and impurity effects over a long time, low-cost sensors are prone to signal drift and low sensitivity (Kizel et al., 2018). In addition, cross sensitivities differ largely according to the ambient temperature and level of gas the sensor is being exposed to (Lösch et al., 2008). So, multi-parameter joint calibration training was utilized to reduce the interference issue between air pollutants in our research, including air pollutants concentrations, temperature and relative humidity".

this sentence in line 106-108:
"One drawback of our study is that the air pollutants concentrations observed at SORPES are lower than those observed in a road environment, which might cause issues for the calibration process."

these sentences in line 130-132:

“To improve performance of the NO_2 model, temperature and relative humidity have also been involved in the training algorithm. However, the interaction between O_3 and NO_2 may influence the detection accuracy of these two chemicals, especially for NO_2 (Ivanovskaya et al., 2001).”

and a final sentence in this section:
“Overall, the sensor results have substantial uncertainty compared to reference methods, we thus focus on the relative temporal and spatial distributions rather than the absolute concentrations.”

It is stated that NO_2 comparisons with the fixed station calibrations, which was only fair (R^2=0.67 for sensor 2), were improved by training and inclusion of t and rh. But the resulting improvements are not shown in this section.
Re: Sorry for the confusion, but R^2=0.67 is already the results with considering temperature and relative humidity. We clarified this by adding the following sentences in line 130-132:
“To improve performance of the NO_2 model, temperature and relative humidity have also been involved in the training algorithm. However, the interaction between O_3 and NO_2 may influence the detection accuracy of these two chemicals, especially for NO_2 (Ivanovskaya et al., 2001).”

Line 107—added text is unclear and perhaps is important. “The success of supervised model training with target labels (i.e. co-located with SORPES, Figure 2a) does not guarantee for its predicting power for conditions without labels (i.e. on road or co-located with SORPES but not feeding the station data to the algorithm, Figure 2b)”. Perhaps this is related to key point 2. However, it should be rewritten and clarified. If it is related to point 2 then it should be considerably expanded.
Re: We had the following sentences following this one:
“We use a calibration-validation methodology to evaluate the performance of the calibrated sensors (Chatzidiakou et al., 2019). This method includes two phases: first, the sampler was calibrated against the SORPES station for 10 days (Jun. 1-10, 2020), and the sensor data were used for sensor algorithm training as above described (Fig. 2a); second, we continued to place the sampler in the station (Jun. 11-17, 2020). However, the sensor data are not used for calibration but directly fed in the algorithm trained in the first phase. The results are compared with the station data (i.e. validation phase, Fig. 2b).”

We added a sentence in line 96-99:
“… cross sensitivities differ largely according to the ambient temperature and level of gas the sensor is being exposed to (Lösch et al., 2008). So, multi-parameter joint calibration training was utilized to reduce the interference issue between air pollutants in our research, including air pollutants concentrations, temperature and relative humidity”.

And other modifications were listed above in section 2.2 in point 2.

Line 140—it is unclear how the overall data reductions employed highly time and spatially resolved data and generated hourly average data. What was this data used for?
Re: The hourly average data is used to evaluate the diurnal cycles of air pollutants (Figure 9). We modified the sentence in line 157-158 as:

"Similarly, we calculate the hourly average concentrations by considering only the data sampled in the same hour of different days."

Line 143—what was the accuracy and data completeness for GPS in the urban areas. Others have found such data difficult to collect reliable complete files in urban canyon conditions.
Re: Our GPS works quite well except some minor spatial offsets. We clarified this by adding a sentence in line 82-83:
"A GPS device (U-blox, Switzerland) is utilized to record the spatial coordinates and the spatial offsets are corrected by Arcgis 10.2 software".

Line 163—"background" data are generated by finding the minimum value of all the stations in the Nanjing area. Is it clear that they represent the urban "background"? Overall, this text is unclear. Are any of these remote from the city or roadside stations? If so, how might these impact any determinations of background? Roadside station, for example might have very low ozone levels due to reaction between NOx and ozone. But this might not be urban "background". Please expand discussion of background assumptions.
Re: We made an explanation in the revised manuscript in line 172-176:
"Seven state-operated air quality observation stations in Nanjing are selected in our research, including Maigaoqiao, Caochangmen, Shanxi Road, Zhonghuamen, Ruijin Road, Xuanwu Lake, and Olympic Sports Center (Zhao et a., 2015; Zou et al., 2017), which are far away from major roads and large point sources, so they are usually used as regional backgrounds in different functional areas (Zou et al., 2017; An et al., 2015). For example, Zou et al. (2017) chose the Olympic Center station (G, Figure 1) to get the background characteristics of CO and NO_2 in Nanjing".

Line 185—reference to the Apte study may or may not be valuable or applicable here. It was performed in a small area of a city—16km^2 and employed considerable resampling of roadways over a prolonged period. Does it apply well in this much larger city/region?
Re: The results obtained by the Apte study were similar to that by the background site method (Figure 10). And a sentence to clarify this was added in line 358-359:
"Although our data coverage is much larger than that of the Apte et al. (2017) study, we find that the reference method is still applicable in our research area".

Line 239/246 and table 1—how were specific sources of pollutants, such as cooking identified, as source contributors to hotspots? Only one cooking establishment is actually reported as a hotspot contributor. Basically, a structured assessment of hotspots vs. sources is valuable, but this paper does not present any robust information on what allowed the identification of contributions beyond visual sightings or general proximity.
Re: We added a sentence to acknowledge this limitation in line 261-263:
"This method has substantial uncertainties to attribute the potential sources to these "hotspots", and further source-receptor relationship and detailed chemical component analyze are required to identify the exact emission sources."

Report #2:

The authors have addressed the majority of my comments from the first round of review. I have several additional minor comments below.

Line 75 - is the data reported every 10s the average of the previous 10s, or an instantaneous measurement? If it's the former, it seems like the spatial assignment should correspond to the median of each 10s sampling interval.

Re: It's an instantaneous measurement and be modified in line 79-80:

"An **instantaneous measurement** of CO, NO_2, and O_3 concentrations is configured to continuous measure at a frequency of once per 10s sampling interval".

Line 104 - please be specific about what concentrations of other pollutants are used in the calibration models.

Re: It was clarified in line 113-114:

"It takes input variables including raw signals of sensors, air pollutants concentrations (**CO, NO_2, and O_3**), temperature and relative humidity".

Figure 2 - (1) please clarify that the scatter plots show concentrations rather than raw signals; (2) it seems like the slopes of these scatter plots are equally as important as the R^2.

Re: It was modified as suggested and the slopes of these scatter plots were added in Figure 2.

Equation 1 seems to show a normalized traffic contribution, but the accompanying text suggests that the equation shows an absolute traffic contribution. Please clarify.

Re: The equation 1 shows the normalized traffic contribution of pollutants. We clarified this by adding a sentence in line 176-178:

"… the **normalized** contribution from traffic-related emissions can be obtained by differencing the mobile measurements and the stationary ones **to minimize the influence of daily meteorological variations in the urban air quality**, following Bossche et al. (2015)".

The ozone "hotspots" in fig 7 seem like cold spots because of NOx titration. Are there any true O_3 hotspots (i.e., where O_3 is higher than the surroundings)?

Re: Thank you for pointing this out. We modified this by identifying the true O_3 hotspots with the area where the pollutant concentrations are 50 % higher than nearby grids. The results were shown in Fig 7, and some sentences was added in line 272-276:

"As shown in Figure 7, the higher O_3 concentrations in these hotspots area are mainly caused by higher NOx and VOCs emissions from the heavy traffic (W, 46.8±27.4 μg m^{-3}, Xie et al., 2016; Ding et al., 2013), cooking emissions (Q, 38.5±26.0 μg m^{-3}), and ozone precursors from industrial emissions [e.g., K (47.1±36.5 μg m^{-3}) and J (37.6±25.8 μg m^{-3})], such as VOCs. In addition, biogenic VOC emissions also have a significant impact on the formation of ozone [(U (40.4±18.3 μg m^{-3}) and V (33.5±20.4 μg m^{-3}), Liu et al., 2018)]".