Interactive comment on “Mobile monitoring of urban air quality at high spatial resolution by low-cost sensors: Impacts of COVID-19 pandemic lockdown” by Shibao Wang et al.

Anonymous Referee #1
Received and published: 13 December 2020

This manuscript describes the deployment of low-cost air pollutant sensors for O3, NO2, and CO on taxis in Nanjing. This work is novel because it combines low-cost sensors with a distributed, quasi-random sampling platform. Overall the manuscript is appropriate for the journal, but it is not ready for publication at this time.

My main criticisms focus on the methods. As detailed in my comments below, the authors need to provide more information on the sensor package that they used. They do not even tell the readers whether these gas sensors were electrochemical, metal oxide, or something else. Additionally, the way that the data are assigned to points in space is not described in sufficient detail.

We are grateful to reviewer #1 for his/her effort reviewing our paper and his/her positive feedback. Here below we address the questions and suggestions raised by the reviewer #1. We provide more information on the sensor package used in this study and the way that the data are assigned to points in space.

1. Line 26-27 It’s unclear what is meant by global air pollution deteriorating by 8%. Is this for a specific pollutant?

Re: The sentence in line 27-28 was modified as “The global urban air pollution (measured by PM_{10} or $PM_{2.5}$) also deteriorated by 8%”.

2. Line 45 - you might need to capitalize Street View

Re: We modified that as suggested.

3. Line 83 - What does SORPES stand for? Also, the link in this line returned a 404 error.

Re: We clarified this in line 94-96: “The instrument is placed at the Station for Observing Regional Processes of the EarthSystem (SORPES) in the Xianlin Campus of Nanjing University (https://as.nju.edu.cn/as_en/obsplatform/list.htm) for at least seven days before the taxi began sampling”. Also, we replaced the link in line 96 with this: https://as.nju.edu.cn/as_en/obsplatform/list.htm.

4. Line 92 suggests that the sensors were not calibrated until June 2020, however the measurements started in 2019. I am confused about the calibration schedule - hope-fully the sensors were calibrated before the sampling on the taxis started. Please clarify.

Re: Thanks for your query. The XHAQSN-508 was calibrated once a month starting from September, 2019. The period June 1-17, 2019 was selected to do the calibration-validation (i.e. two-phase) experiment, but the one-phase calibration was conducted every month.

To clarify this, the sentence in line 94 was modified as: “The XHAQSN-508 is calibrated every month starting from September, 2019.”

5. I am not familiar with the XHAQSN-508. What kinds of sensors are these? Electro-chemical? Metal oxide? More detail on the specific gas sensors is needed. Also, is the sample refreshed by pumping air push the

sensors, or do you rely on the airflow generated by the moving vehicle? If it's the latter, does it impact the performance to have the sensors stationary during calibration experiencing wind during sampling?

Re: Thanks for pointing it out. The sentence in line 67-69 was modified as: "The instrument is equipped with internal gas sensors for CO, NO_2, and O_3 (dimensions: 290×81×55 mm; weight: 1.0 kg) as well as two small in-built sensors for temperature and relative humidity, and is fixed in the top lamp support pole (~1.5 m above ground) of two Nanjing taxis (Figure 1)".

And we also added the following sentences after that:

"All three sensors are electrochemical-based sensors that can detect gaseous pollutants at levels as low as ppb (Maag et al., 2018). It is continuously powered by an external DC 12V power supply provided by a taxi battery. The sample is refreshed by pumping air to the sensors. There is an air inlet at the bottom of the instrument, which is also checked periodically to avoid blockage. Because it is fixed in the taxi top lamp, it can reduce the impact of different wind direction airflow".

Then we added the relevant instrument description in line 75-77:

"The monitoring data is automatically uploaded to a database in the cloud via the 4G telecommunications network. The monitoring system of CO, NO_2, and O_3 are configured to continuous measure at a frequency of once per 10 seconds, and their limit of detection (LOD) are 0.01 μmol/mol, 0.1 nmol/mol, and 0.1 nmol/mol, respectively".

6. As shown in Figure 2, it seems that the calibration approach was to use the "forward" method - e.g., calibration models were built on one week of data, and then that calibration was used going forward. Other low-cost sensor studies use k-fold cross validation. In this approach, the data are divided into k chunks, and models are built on k-1 chunk sand tested against the holdout. Does a k-fold cross validation of your data result in different (or perhaps better performing) calibration models?

Re: We clarify this by adding this sentence in line 101-103: "GBRT, an ensemble learning method, is a decision tree-based regression model that implements boosting to improve model performance using both parameter selection and k-fold cross validation".

7. The authors should specify what parameters were used in the calibration models. Is it just sensor raw signal, or are variables like temperature and humidity also included? Including a humidity term may improve performance of the NO2 model, as the authors note in lines 100-101 that the NO2 model may have a humidity bias.

Re: We acknowledged this point by adding this sentence in line 100-102: "GBRT needs to be trained by a dataset with target labels (Yang et al., 2017). It takes input variables including raw signals of sensors, other air pollutants concentrations, temperature and humidity. The stationary instrument data are taken as training targets".

We also added a sentence in line 119-120: "To improve performance of the NO_2 model, temperature and humidity are also involved in the training algorithm".

8. Section 2.3 needs a better explanation of how the data are assigned to points in space. Data are logged every 10 seconds. Under many driving conditions (speed > 18 km/hr), the vehicle will cover more than 50 m in 10 sec. How is the resulting data assigned in space? Is it the location of the vehicle when the data point is logged?

Re: We clarified this by adding some sentences in line 134-138: "The driving condition is highly variable and the taxi can travel more than 50 m in 10 seconds if the vehicle speed is over 18 km/hr. However, given the complexity of the driving conditions, we ignore the vehicle trajectory in the past 10 seconds but assign the measured values to the location of the vehicle at the time of data uploading. Then, combined with GIS technology, we calculate the average of all the data points over one year that fall in the same grid."

9. Is the final output the mean concentration in each grid? Since grid cells can be sampled unevenly across different days, other studies have first internally averaged the data by day. E.g., Apte et al 2017 compute the grid cell median for each sampling day, and then compute the mean of all daily medians.

Re: No, we used the direct average of all points throughout the year. By this mean, we treat all points in each grid equally. We clarified this by adding the following sentence in line 137-138: "Then, combined with GIS technology, we calculate the average of all the data points over one year that fall in the same grid."

There are large minute-to-minute, hour-to-hour and day-to-day variabilities in pollutants concentrations. To calculate the mean (or median) of each day and then the mean of all daily mean (or median) is thus quite arbitrary. For example, why not using an eight-hour or weekly mean as the intermediate step? We argue that our method (i.e. direct mean of all points) is simpler and also robust if we have a large sample size.

10. How do large concentration spikes impact sensor performance? In our laboratory tests of electrochemical sensors, we observed that concentration spikes can cause the raw signal to remain high for several minutes. Presumably there are many spikes encountered during mobile sampling. Have the authors considered the potential impacts of these spikes? More broadly, have the authors considered that the sensors may not be able to reliably report at 10-sec resolution?

Re: Thanks for pointing it out. We indeed noticed the same phenomenon, and that is a drawback of our study. We acknowledge it by adding the following sentences in line 138-140:

"One drawback of our study is the impact of spike concentrations on sensor performance. The sensors keep reporting high concentrations in an approximate one-minute period after exposure to large environmental concentration spikes. This effect would reduce the effective resolution of our gridded concentration map."

11. I'm confused by what is shown in Figure 4. I think that the standard error of the mean was calculated for each grid cell, and then averaged over all grid cells, but that is unclear. Were grid cells excluded if they did not meet a data threshold (e.g., if they did not have "enough" data)?

Re: The review is correct and that's exactly what we did. We clarified it with relevant explanations in lines 176-177: "We calculate the standard error of the means of samples in each grid (SEM), and then averaged the SEM over all grid cells". We did not exclude any grid cells if they have more than two data points.

12. Figure 6 is hard to read. The lines indicating the roadways (or grid) are very thin, and it's hard to see the variation in the color scale with such thin lines.

Re: We have changed the image to a higher resolution, so we can see it clearly by zooming in. Very few readers read a paper version after all.

13. Section 3.3 - This section is titled uncertainty analysis, but the discussion (especially lines 186-194) are more about spatiotemporal variability than uncertainty. This means that I am unclear on whether Fig 6 shows variability in measurement uncertainty (e.g., because of different sensitivities for different species), or if the variations in the coefficient of variation represent physical phenomena associated with emissions and chemistry.

Re: Thanks for the suggestion. We modified the title of section 3.3 as "Variability analysis".

14. Are the concentrations shown in Table 1 the mean concentration, or the concentration above background? The latter might be more informative.

Re: We added a sentence to clarify it in line 233-234: "The pollutant concentrations shown in Table 1 are the values after deducting the background concentrations, which are calculated by the annual mean concentration of stationary stations".

15. Section 3.4.2 - how are the different types of roads defined?

Re: We clarified this in line 146-147: "We divide the urban roads in Nanjing area into five types, including highways, arterial roads, secondary roads, branch roads, and residential streets (https://wiki.openstreetmap.org/wiki/Key:highway)".

16. Figure 9 shows diurnal patterns for the different pollutants. Were the data sub-selected in any way? I imagine that the locations sampled might be different across different times of day (e.g., maybe more time on highways at certain hours). It would be best if the data were somehow filtered - e.g., by only showing data collected on a certain road type, or by ensuring that data for each hour have a similar mix of road types sampled.

Re: We thank the reviewer for bringing this up. We acknowledged this point by adding this sentence in line 289-291: "The difference of the hourly variation of the mean sample of different types of roads over a year was small (Figure S6), so the data in Figure 9 is not filtered in anyway, but for each hour have a similar mix of road types sampled".

17. Figures 8 and 9 do not show a strong weekend effect. In the US, there is a strong weekend effect due to lower commercial diesel traffic (so there is lower NOx on week-ends); but gasoline passenger cars have similar activity on weekends as weekdays, so CO is similar. Do your data suggest something about traffic patterns on weekdays versus weekends?

Re: Thanks for pointing it out. We added some discussion for this effect in line 300-305:

"Wang et al. (2013) found that NOx displays weekly cycle in the Beijing–Tianjin–Hebei metropolitan area, with higher level on weekdays than weekends. Qin et al. (2004) observed a significant weekend effect in southern California, showing that in the morning traffic rush time, the concentrations of CO and NOx at weekends were about 18% and 37% lower than on weekdays. The difference between our study and other cities lies in the difference of fleet fuel structure, and the different weekly routine of human activities and the taxi driving trajectories (Xie et al., 2016)".

18. Figure 10, much like Figure 6, is hard to read. Maybe the authors could show a single panel in the main text and put the rest in the Supplement.

Re: We revised it as suggested.

19. Line 293 - is the traffic percentage of O_3 even a useful figure? As the authors note, attributing O3 is complicated because of secondary chemistry. I think they should remove the ozone estimate and focus here on CO and NO2.

Re: Yes, it's a good suggestion. We have removed the ozone estimate in the revised paper.

20. Figures 11 and 12 are too faint to be readable.

Re: We have revised it in the revised paper.

Anonymous Referee #2

Received and published: 22 December 2020

General Comments

This paper presents a mobile monitoring study of CO, NO2, and O3 concentrations in a major urban area. The research in this paper is a solid scientific study that adds to the knowledge we have of the variability in air concentrations in large urban areas. Below I detail some specific comments that should be addressed by the authors as well as some technical corrections.

We thank the reviewer for this comment and the helpful suggestion. We have carefully addressed the reviewer's concerns. Please see below our replies. We hope he/she is satisfied with our answers and the new (figure) we provided.

Specific Comments

1- Lines 85-90: please provide detailed information on the machine learning algorithm used, including the equations used to calibrate the data, what is considered a "substantial deviation" from the national network measurements, how recalibration was conducted if there was a substantial deviation, and how many times recalibration was needed.

Re: The detailed information on the machine learning algorithm was added in line 101-105: "GBRT, an ensemble learning method, is a decision tree-based regression model that implements boosting to improve model performance using both parameter selection and k-fold cross validation. GBRT needs to be trained by a dataset with target labels (Yang et al., 2017). It takes input variables including raw signals of sensors, other air pollutants concentrations, temperature and humidity. The stationary instrument data are taken as training targets".

Since we did not calculate the "substantial deviation" from the national network measurements, we deleted it in the revised manuscript.

2- Lines 91-99: explain why you are using a machine learning algorithm. My understanding from your paper is that Figure 2a shows actual measurements, while Figure 2b shows the machine learning air concentration estimates for the mobile sensors compared to actual measurements at the fixed site. The correlations in Figure 2a are much better than those in Figure 2b, which would suggest that there is no need to train an algorithm to develop better estimates of concentrations. Why can't you simply use the measurements from the low-cost sensors for your calibration/validation? Is it because the study data were collected throughout the city, and not just near fixed monitors? If so, perhaps you can do a second calibration using data near fixed monitors, without the machine learning algorithm.

Re: To clarify this, we added this sentence in line 90-94: "Different from traditional instruments, low-cost sensors have some limitations, such as dynamic boundaries, nonlinear response, signal drift, environmental dependencies and low selectivity, so it is important that calibration procedures are applied with respect to these limitations (Maag et al, 2018). The sensors are usually trained with co-located data collected by reference methods before being deployed to actual measuring campaigns (Kaivonen and Ngai, 2020; Chatzidiakou et al., 2019; Bossche et al., 2015)".

We added a sentence in line 98-100 to further clarify:

"Comparing different calibration models, we found that machine learning algorithm can improve sensor/monitor agreement with reference monitors, and many previous studies have used this method (Qin et al., 2020; Esposito et al., 2018; Vito et al., 2018)."

We also added a sentence in line 107-109: "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)".

3- Lines 128-130: this is a broad statement, and not true of all urban monitors. Can you provide citations to studies or reports that show that the stationary monitors do not have a significant impact from traffic emissions and are representative of urban background air quality?

Re: We clarified this by adding the following sentences in line 155-159: "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".

4- Lines 205-206 and Table 1: explain how you are identifying the main source contributions to the hot spots. Is it based on nearby sources and wind direction? Do different sources have different fingerprints (i.e., different relative concentrations of the measured pollutants)? Are there other studies showing that these sources had significant contributions at these locations?

Re: We clarify this in line 235-238: "To identify the main sources contributing to these hotspots, we use the different relative concentrations of the measured pollutants (Zhao et al., 2015). We also use field information around hotspots area, such as the existence of subway stations, construction sites, factories, and restaurants nearby".

Other studies had consistent results as stated in line 252-254:

"Previous studies have also found that the air pollutants "hotspots" are associated with traffic-related emissions [e.g., heavy-duty diesel vehicles (Targino et al., 2016) and vehicle congestion (Gately et al., 2017)] and high-density urban areas (Li et al., 2018)."

5- Lines 334-335: do the observations at fixed monitors support the theory that increased temperature/insolation is the cause of higher O3 concentrations in P3 as compared to P1?

Re: Yes, they do. We added several references to support it in line 375: "……(Xie et al., 2016; Fu et al., 2015; Reddy et al., 2010)".

Technical Corrections

1- Figure 2: both the x- and y- labels on the regression plots are labeled "station." Please change this to specify which station.

Re: The Figure has changed in the revised version. The x- and y- labels in Fig. 2a represents sensor-1 and sensor-2 respectively, while in Fig. 2b represents SORPES station and sensors data respectively.

2- Figure 5: the resolution isn't good on this figure. Can you re-plot with better resolution? Also, the yellow/orange colors are hard to differentiate in Figure 5b.

Re: We replace it with a high-resolution image, which can be viewed by zooming in.

3- Line 334: 'insulation' should be changed to 'insolation'

Re: We modified 'insulation' to 'insolation' in line 374.

4- Figure11: this figure is very hard to read. Can it be made a higher resolution or different color scheme?

Re: We have replaced it with a higher resolution image in revised version.