General comments

This paper presents results from a low-cost CO2 sensor deployment in Seoul. The authors deploy around two dozen low-cost sensors ("HT-2000" with SenseAir S8 Co2 sensors) along with a higher-quality sensor ("LI-840a" with LI-COR Environmental NDIR sensor) for calibration. The authors discuss different calibration methods and show that a multi-point linear regression that accounts for a time lag between sensors reduces RMSE the most, while temperature and humidity do not noticeably help reduce RMSE. After calibration, the authors show results from a 1-hour deployment at an intersection and a 20-hour deployment in a tunnel. The authors observe positive correlation between CO2 concentrations and the number of cars in the tunnel.

While the monitoring results presented in this paper are interesting and align with expected patterns (e.g., higher CO2 concentrations at the traffic intersection and during rush hour in the tunnel), it is not clear how long the calibration of the low-cost sensors would last during a longer-term deployment. The authors present only 1-hour of data from the first deployment at an intersection and 20-hours of data from the second deployment in a tunnel. I think the manuscript would benefit from additional analysis evaluating the robustness of the calibration procedure for longer deployments. For example, it is very important to know how often recalibration will be required. Furthermore, I think the organization of the paper could be improved. The authors present critical details about the calibration procedure in the Discussion section that I believe should be presented earlier in the methods section. I think it's important to fully understand the calibration procedure and efficacy before showing the CO2 monitoring results.

Specific comments

- Major concern: It is implied that the multiple linear regression calibration at the intersection is done at "correction" points where all low-cost and high-quality sensors are co-located based on the Figure 1 caption. However, this is not explicitly stated. Can the authors confirm that this is indeed the case, and if so, state it directly in the manuscript. If the sensors were not co-located before they were calibrated, then this could introduce possibly large errors due to spatial variability in CO2 concentrations.
- Major concern: It is implied that the data in Figure 2 is the data used to calibrate the low-cost sensors while they were co-located with the high-quality sensor, and that the spatial maps in Figure 3 are based on the low-cost measurements at the sensor locations shown in Figure 1. Can the authors confirm if this is the case? And if so, please state this directly. It would be useful to show not just the data used to calibrate (what I believe is Figure 2) but also the time series data used to generate the spatial maps. It is a bit confusing as currently presented, as the time series from each sensor in Figure 2 look nearly identical after calibration, but the spatial maps show clear spatial patterns.

- Major concern: since the calibration and deployment periods were relatively short, it is not clear if the low-cost sensors will continue to provide accurate measurements over longer deployment periods. Do the authors believe that this sensor network can be used for longer-term monitoring? If yes, how often will the low-cost sensors require re-calibration?
- Introduction: I think some discussion of other studies that perform urban-scale CO2 monitoring using low-cost sensors is missing. How is this study different? E.g., are there new bias-correction methods that result in more accurate measurements from the low-cost sensors? Is this the first time this type of monitoring is being performed in Seoul? Or is it the first time these specific sensors have been used for continuous monitoring?
- Section 2.1: It would be interesting, if possible, to list sensor prices. Perhaps just an order of magnitude. This would provide context and motivation for why the low-cost networks are important.
- L79: How was the length of the 137s moving time window selected? How sensitive is the calibration to this length?
- L80: How was the time delay calculated for each HT-2000 sensor? In general, more detail is needed on the calibration procedure in this section.
- L84: The authors point to low-cost sensors for better temporal coverage, but they only measure for 1 hour. Would it have been possible to leave the sensors deployed for longer?
- Section 2: the discussion of sensor calibration is fragmented and hard to follow. I recommend having a subsection devoted solely to the detailed discussion of sensor calibration methods. There is also some calibration information in Section 4.1 that I think belons in Section 2 before presenting results.
- Section 2: more discussion is needed about why different calibration methods and meteorological variables were used between the two study areas. Why were temperature and humidity excluded from the tunnel study, but included in the intersection study? Why was two-point calibration and multiple linear regression used at the tunnel, but only multiple linear regression at the intersection? After reading the whole article – I think the calibration results in section 4.1 belong in section 2 (before presenting the co2 results).
- Section 2.3: This section would really benefit from a map, similar to what was provided in section 2.2
- Section 3.1: Please list the coefficient values for the multiple linear regression, either in the main text or the SI. It's not clear how much of the correction is currently coming from the meteorological variables. After reading the whole article I see that Section 4.1.3 has some information on this. See earlier note about ordering. I still think it would be useful to list the coefficient values.
- I think Section 4.1 belongs in the methods section, not the discussion section. Several of my questions were answered when reading this section, and I would have liked to see calibration details before looking at the results.

Technical corrections

- L30: This sentence seems a bit out of place, since the authors did not conduct eddy covariance measurements. Instead, I think the authors should motivate by referencing other studies that use low-cost sensor networks to monitor emissions.
- L59: "The LI-840a boasts an accuracy better than 1.5% of the reading value and an RMS noise level below 1 ppm." I think this sentence should be stated the same way as the sentence describing the HT-2000 sensor to make for an easier comparison. Something like: "according to the manufacturer, its accuracy is within +/- XX ppm or +/- YY% of the reading." It's not clear if the RMS noise level of < 1ppm stated here is the same metric as the +/- 70ppm stated for the HT-2000.
- L60-61: This claim seems unsupported by the work in this manuscript. Is there another study you can reference to support the 0.1% accuracy claim?
- Fig 1: I only see 19 blue dots. Where is the 20th sensor located?
- L76: This sentence doesn't make sense: "The LI-840a was placed at a single location near the centre of the intersection using multiple-point linear regression." Perhaps this paragraph is out of order?
- L90: Clarify that the covariates in the multiple linear regression are the fields measured by the HT-2000 sensor.
- L142: Please state which interpolation method you are using with "*scatteredinterpolant.*" It's also worth noting that this method does not leverage any information about, e.g., atmospheric transport or the presence of buildings that would block the flow of CO2 between potential sources and the sensor locations.
- Figure 4: It's unclear what the "distance from the entrance" means. Is this being measured into the tunnel, with 0m at the entrance? If so, at what distance is the exit point? It would be helpful to mark this distance on the figure. Also see earlier note about providing a map of the tunnel experiment. How many sensors were there and how far apart were they within the tunnel?
- Figure 4: Needs some discussion of how the point data was spatial interpolated. Is it the same method as the intersection case study?
- Fig 10: please state where the traffic data are coming from