We would like to thank the two anonymous reviewers and the Editor for their careful reviews of our manuscript. Below we present our point-by-point responses to all of the comments. The original comments from the reviewers are shown in black, while our responses are presented in blue.

Reviewer #1

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.
  - The sensors were colocated to obtain the data for correction; now we have explicitly mentioned it in the manuscript and figure S2.
- 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.
  - Figure S2 has been newly added to show the time series used for when low-cost sensors and LICOR were co-located for correction (i), (in transit (ii)), and HT-2000 sensors were deployed for measuring data to create spatial maps (iii).

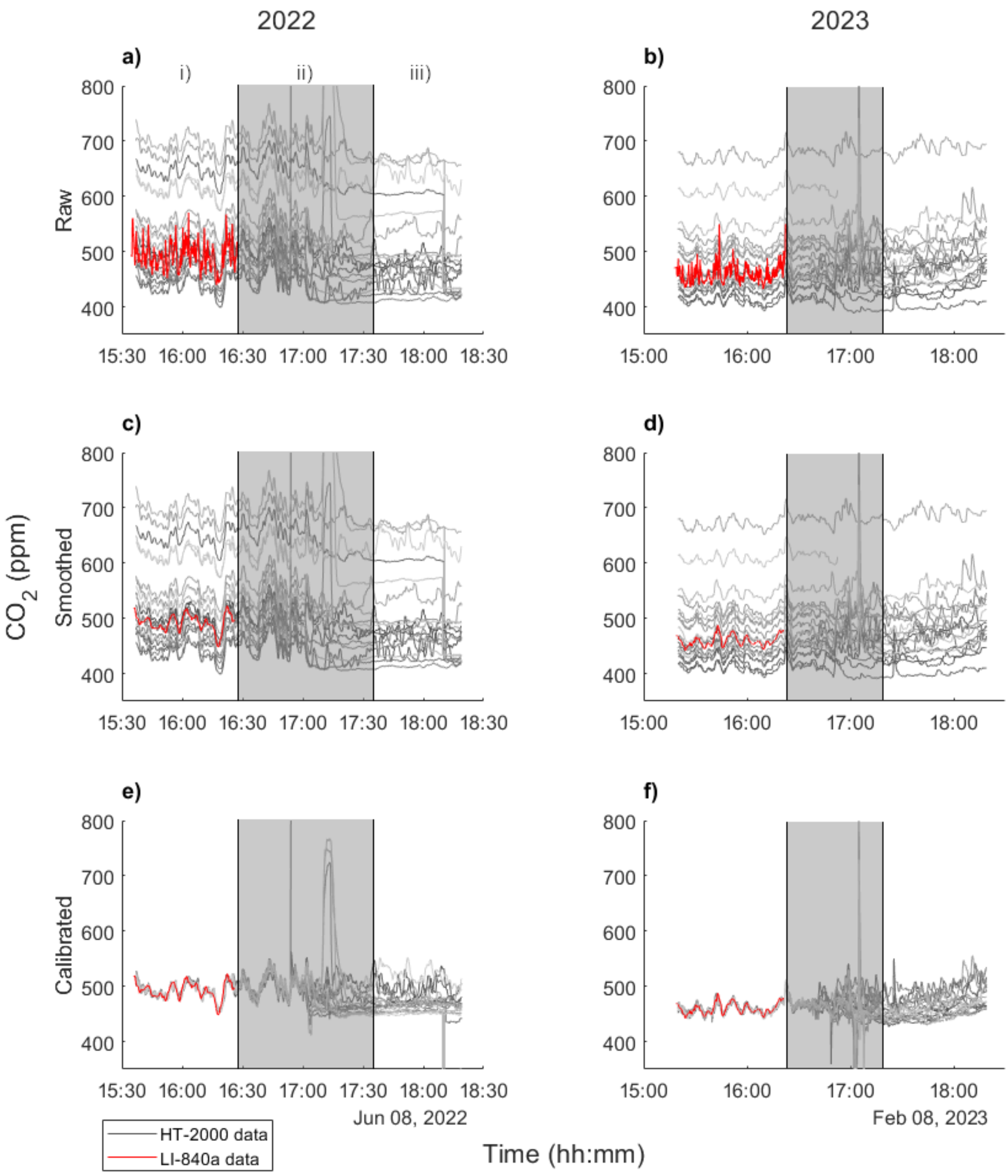

Figure S2: CO_2 concentrations measured in Bongcheon Intersection, before and after correction. a) and b) Plots of CO_2 values measured by HT-2000 (grayscale) and LI-840a (red), before any correction was applied. c) and d) The LI-840a values smoothed with a 137-second window. Note that the HT-2000 sensor values closely follow the smoothed LI-840 values, though with fixed offsets. e) and f) Results after linear correction of the HT-2000 meters. i)-iii) sections (labelled in a)) each represents: co-located measurement for obtaining data for correction of HT-2000 (i), moving HT-2000 sensors to its assigned location (ii), HT-2000 measurement in its assigned location (iii).

- 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?

- During measurements at the Guryong Tunnel, we found that the sensors remained stable for 24 hours. We expect the sensors to remain stable for a longer period, but the HT-2000's battery life is insufficient to operate them for more than 48 hours. We note this limitation in Section 2.1.

* 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?

  - We used very low-cost kits (85 USD for each) compared with other low-cost kits (~1000s USD). This enables a large number of sensors to be used at the same time to increase spatial resolution. To clarify, we added the following words in the main text:

    > "This study presents an approach using a very low-cost, pre-assembled CO_2 monitoring kit (~85 USD; including CO_2, relative humidity, and temperature sensors) to enable high-resolution CO_2 measurements in urban environments."

* 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.

  - We added the sensor price of 85 USD in the manuscript.

* L79: How was the length of the 137s moving time window selected? How sensitive is the calibration to this length?

- The smoothing window of 137 seconds is chosen using an algorithm; that is, we have tried various lengths of smoothing window using a FOR loop for both correction periods and found out that 137 seconds yielded the least RMSE difference between the LI-COR and HT-2000 instruments. To clarify, we added the following words:

  "We have tried various lengths of smoothing window using a FOR loop for both correction periods and found out that 137 seconds yielded the least RMSE difference between the LI-COR and HT-2000 instruments."

- 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.

  - The time delay was also calculated algorithmically, by finding the "best match" between the LI-COR and HT-200. i.e., by running a FOR loop to vary the time delay, and calculating the standard deviation between the difference of HT-2000 and LI-840a. We added the following words:

    - "The time delay for each HT-200 was also calculated algorithmically, by finding the "best match" between the LI-COR and HT-200."

- 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?

  - During measurements at the Guryong Tunnel, we found that the sensors remained stable for 24 hours. We expect the sensors to remain stable for a longer period, but the HT-2000's battery life is insufficient to operate them for more than 48 hours. We note this limitation in Section 2.1.

- 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.

  - We added new Sections 2.3.3 through 2.3.6 and moved the content from the former Section 4.1 into them.

- 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).

  - At the first measurements Guryong Tunnel, the distribution of CO_2 is one-dimensional, simpler than the 2-D distribution in Bongcheon Intersecrtion, so we thought the two-point method would be enough. However, doing so resulted in significant increase of RMSE. We also applied the multi-point regression for the Guryong Tunnel to compare the results from the two methods.

  - We clarified the changes in the correction methods, along with the reasons for them, in Section 2.3. The previous Section 4.1 has been moved to Section 2.

- Section 2.3: This section would really benefit from a map, similar to what was provided in section 2.2

  - A figure (see below) was newly included in the revision.

Figure 1 Aerial photograph of Guryong tunnel and the location of placed HT-2000 sensors. Map data © Airbus 2025. The tick marks indicate the location of the sensors placed; 0 indicates the entrance to the tunnel.

- 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.

- The coefficient values are now included in the Supplement.

- 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.
    - We placed the previous section 4.1 inside section 2.

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.
    - The following words were deleted:

      "Ground-level CO_2 concentrations can be converted into atmospheric fluxes using flow measurement techniques such as eddy covariance (Burba et al., 2013; Vardag and Maiwald, 2023). Therefore, measuring ground-level CO_2 enables more accurate CO_2 flux estimation."

- 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.
    - We changed the sentence as follows:

      "According to the manufacturer, LI-840a has an accuracy better than 1.5% of the reading value and RMS noise level below 1 ppm. In addition, from our experience, the sensor's precision can be calibrated to significantly surpass the manufacturer's specifications, often yielding fluctuations of approximately 0.1ppm. Despite its superior performance, the high cost (>3000 USD) of the LI-840a unit limits the number of units that can be deployed simultaneously.

- 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?
    - In general, LI-COR is popularly used for background measurements with precision better than 0.1 ppm. We have not independently checked it, but we assumed other labs also have achieved similar precision results. Considering the uncertainty of the low-cost sensors (~ a few ppm), even if the accuracy of LICOR is significantly larger than 0.1ppm, it would not make significant change in our conclusions.
- Fig 1: I only see 19 blue dots. Where is the 20th sensor located?
    - The sensor placed in one spot ended up having problems for the both measurements (in the Bongcheon Intersection), so it is not depicted in the map. The caption was changed to convey this information.
- 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?
    - Sentence changed to read more smoothly:

      "We then corrected the 20 HT-2000 meters with multiple-point linear regression; the data for regression was obtained by placing all HT-2000 sensors and the LI-840a sensor at the same place, near the centre of the intersection."
- L90: Clarify that the covariates in the multiple linear regression are the fields measured by the HT-2000 sensor.
    - Description added as follows:

      "In this experiment, y_i corresponds to the CO_2 concentration measured at time i using the LI-840a analyser, while, x_{i1}, x_{i2}, and x_{i3} represent the CO_2 concentration, temperature, and humidity (measured by each HT-2000), respectively."
- 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.

- We have used the default interpolation method which uses Delaunay triangulation to divide the plane into triangles, then linearly interpolates within the triangle. In results session, we mentioned the issue:

  "The corrected CO_2 concentrations obtained from each HT-2000 were then interpolated temporally for each HT-2000 sensor, then interpolated spatially using the *scatteredinterpolant* function in MATLAB which uses Delaunay triangulation then performs linear interpolation on each of the triangles on default settings."

- We believe we may sidestep the limitation (of not taking into advantage the urban landscapes / air transport) by adding more sensors. It is now mentioned in the Section 4.1 as follows:

  "Our methods do not incorporate any information about the urban landscape, such as building locations or shapes, nor do they rely on air-transport data. They can also be integrated with atmospheric modeling by replacing the spatial interpolation step with a suitable modeling approach. Because the sensors are low-cost, their numbers can be increased substantially, enabling much more comprehensive monitoring."

- 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?

  - It represents the distance from the tunnel entrance. The map (currently figure 2) has been provided, with tick marks indicating the sensor locations.

- Figure 4: Needs some discussion of how the point data was spatial interpolated. Is it the same method as the intersection case study?

- Since the tunnel is linear, we have linearly interpolated each sensor in time first, then linearly interpolated spatially, using all sensors. Former figure 4 (currently figure 9) caption has been changed as follows:
  - "Figure 9. CO_2 values measured (corrected) at the Guryong Tunnel, on July 25 (a) and November 21 (b, segment) in 2024. Data was temporally resampled from 10 s between each data points to 1 s using linear interpolation, then spatially interpolated for each second using 1-dimensional linear interpolation. The y-axis is measured from the tunnel entrance, and the 1180m mark represents the exit of the tunnel."

- Fig 10: please state where the traffic data are coming from
  - The traffic data were obtained from Seoul Transport Operation and Information Service (https://topis.seoul.go.kr/refRoom/openRefRoom_1.do).
  - We added the data source in the Fig. 11 caption (note that the figure numbers were changed in the revision).