Referee 2:

This is an interesting evaluation of the SenseAir K30 sensors, co-located with Picarros, over a 30 month timescale. This is a valuable study, but the manuscript needs significant revision to properly categorize sensor types and adjust all comparisons accordingly, address scalability concerns, strengthen the literature context, and better acknowledge limitations. It should be accepted for publication in AMT if the following can be addressed.

The authors fail to distinguish between "low-cost" and "mid-cost" CO2 sensors. The Vaisala CarboCap GMP 343 sensor mentioned in line 59 and used in networks like ZICOS-M (https://acp.copernicus.org/articles/25/2781/2025/) and BEACO2N is at a significantly higher price point compared to the SenseAir K30 used in this paper and the Carbocaps are now usually called a "mid-cost" sensor to distinguish them for the LCS like SENSE-IAP. Please adjust the introduction to distinguish between sensors in the $10s-$100s USD (low-cost), sensors in the mid-cost range ($1000s), and reference grade sensors (typically $10,000s).

Additionally, the accuracy and precision statistics given for other sensors should be for sensors at a similar price point to the sensors used in the paper. Vaisala CarboCap is not a comparable sensor (lines 64-68). Focus literature review on truly comparable low-cost NDIR sensors.

Response:

Thank you very much for your valuable comments. We have modified the section of introduction in lines 44-52. We added a table (Table 1) to clearly compare the cost, accuracy, and deployment status (including key cities with dense networks) of different low- and mid-cost CO_2 sensors.

In addition, we have revised the grammar and expression issues suggested by the reviewer. We added a literature review on long-term drift correction methods and added a discussion on the limitations of the long-term co-location results and the suggested correction frequency. We have provided detailed responses to the comments in the following text.

Moreover, we think that the methodologies and accuracy on mid-cost sensors and networks can provide valuable references and comparison for this study.

Recently established high-density networks use both low-cost and mid-cost sensors, and although the sensors are low- or mid- cost, the final operation and maintenance of the networks are relative high, sensor cost is not so important compared with the total cost, but the precision and accuracy of the network matter, and some networks using LCS with good correction methods can achieve comparable precision with mid-cost sensors. For the low-cost sensors, examples like SenseAir LP8 CO_2 sensor network in Switzerland, and SenseAir SENSE-JJJ in Beijing. And for the mid-cost sensor networks, there are BEACON and Paris using Vaisala GMP343 and SenseAir HPP, respectively.

Moreover, both low-cost and mid-cost sensors operate on the NDIR principle, exhibiting significant environmental sensitivity, susceptibility to jumps, and long-term drifts. Therefore, whether it is low-cost or mid-cost, their calibration principles and methods are common. Based on our results, after environmental sensitivity correction and effective long-term drift calibration, the accuracy of low-cost sensors is comparable to that of mid-cost sensors.

[1] Turnbull, J., DeCola, P., Mueller, K., Vogel, F., Karion, A., Lopez Coto, I. and Whetstone, J. (2022), IG3IS Urban Greenhouse Gas Emission Observation and Monitoring Best Research Practices, World Meteorological Organization Integrated Greenhouse Gas Information System, [online], https://ig3is.wmo.int/ (Accessed June 11, 2025)

What is China's dual carbon goal (line 72)? Please add context for the international reader.

Response: Thank you. It is that China aims at peaking carbon emissions before 2030 and achieving carbon neutrality before 2060. We have provided additional clarification regarding the China's Dual Carbon Goals.

Line 95: is "homology" the intended word? Homogeneity maybe?

Response：Yes, here "homology"is the intended word, and in environmental sciences, 'homology' refers to identical emission sources of air pollutants and CO_2 resulted from fossil fuel combustion.

What exactly is meant by background noise level (line 113)?

Response： We use the standard deviation of the signal to quantify the level of background noise or white noise (e.g. the amplitude of raw signal in Fig. 3a in the main text, blue dots). So, we used "standard deviation" to substitute "background noise level" according to your question. We have refined the text to enhance clarity.

Figure 4 map labels are too small to be legible and is difficult to locate for readers unfamiliar with Beijing city. Also the latitude labels are cut off on the left side.

Response: Thank you for this question. We added the 5th and 6th Ring in the map to show the location. But the base map provided by ESRI map server cannot display street and area names more clearly. We added an eagle-eye view map to show the site's location. We have addressed the issue of cut off label display in the map.

Line 189-190 is confusing and grammatically incorrect. Same with like 193-194.

Why can you say that hanging on an open window is the same a field-deployment? If the instruments are even partially indoors surely they are more temperature controlled than a true field deployment?

What is the recommended length of co-location with reference instrument for determining the correction coefficients? It would be helpful to include this in addition to the recommendation of a 3-month calibration interval.

Response: Thank you for this question. Due to the limitation of experimental conditions, we set the instruments at the edge of the open window to directly take in the outside air. We compared the observed environmental conditions of indoor, our set-up instruments, and out door field in the following figure 1 (Also added in Fig. S8). We show that the instruments shared almost the same changing environments as the out door air conditions. The temperature mainly ranged from 0-40 ℃, and the RH mainly ranged from 0-60%. And both if them are quite different from the indoor conditions. We added these discussions in lines 222-232.

We have added the discussion of the recommended length of co-location with reference instrument for determining the correction coefficients in lines 131-141. It was generally one week period of co-location with reference instrument for controlled environmental experiments to determine the environmental correction coefficients of the sensors, and another week co-location for post-quality check. And to fully understand the characteristics of long-term drift, at least one year is recommended to include the seasonal cycle (both summer and winter).

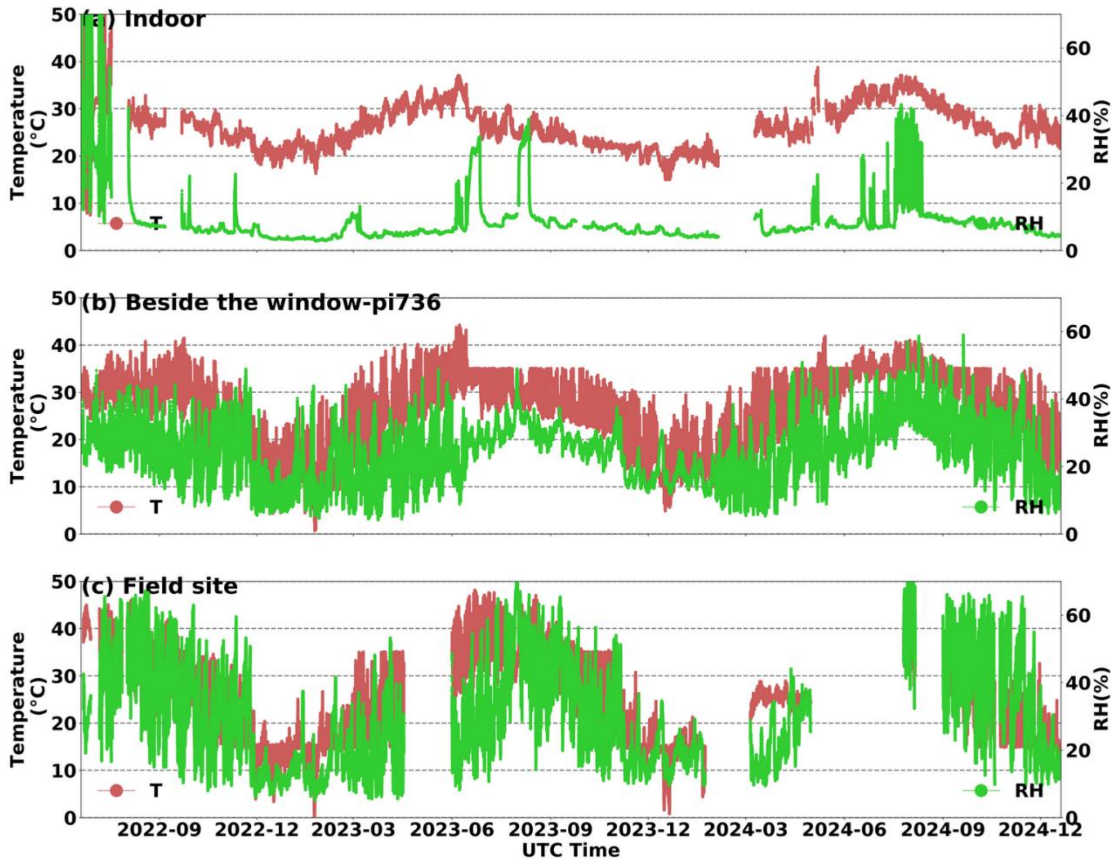

Figure 1. The temperature and humidity variation of LCS instruments under (a) indoor, (b) beside the window, and (c) field conditions during the experimental period.

Line 296: "manufacture" -> "manufacturer"

Response：Thank you for this comment. We revised accordingly.

Line 324-325: missing spaces around ±

Response：Thank you for this comment. We revised all such errors in the text accordingly.

Please add additional discussion of the limitations of this study and how the findings may or may not translate to other LCS and environments. Are the recommendations made only for K30s?

Response:

Thank you for the valuable suggestions. We have added the discussions for this issue in section 7 lines 365-374.

"Long-term drift is typically observed in low- and mid-cost NDIR CO_2 sensors. While our findings provide valuable references for K30 sensors, the recommended calibration frequency and observed seasonal cycle characteristics may be also applicable to other similar NDIR sensors, and the calibration method we used is universal and helpful in long-term drift corrections.

For other NDIR sensors, we recommend selecting several samples to conduct co-location with high-precision instrument under field conditions for at least one year to study their full characteristics. This extended co-location period is essential to comprehensively characterize both long-term drifts and the seasonal drift cycle. The results will provide critical guidance for remote calibration of high-density networks using this type of sensor. Furthermore, another practical substitute method is using standard gas, and the calibration frequency from the experience of this study is recommended at least one-three month."

Have the authors explored alternatives to a co-location for drift correction every 3-6 months? This may not be feasible for large deployments. Did the authors explore the performance of a remote calibration strategy at all? How would the proposed calibration scale (in cost/time) to, say, 100s of sensors deployed?

Please add additional literature review describing the existing methods for LCS calibration.

Response:

Thank you for these valuable suggestions. We added discussions in lines 65-80. Indeed, this may not be feasible for large deployments, it is useful to understand the sensor performance and characteristics in long-term drifts, and suggest us the calibration frequency after deployment. We also explored several other calibration methods for drifts. One robust and commonly used method is automatic/manual standard gas calibration. We used this standard gas method in the Beijing and Jinan networks with more than 160 instruments, the calibration frequency is one week for automatic standard gas, and 1-3 months for manual calibration, and the results showed mean biases -1.28 ~ -0.64 ppm with standard gas at one month scale (Cai et al., 2024) (below figure). The manual standard gas calibration is useful in mobile observations such as on-road observations using vehicles and vertical profile observations using tethered balloons (Liu et al., 2021; Bao et al., 2020).

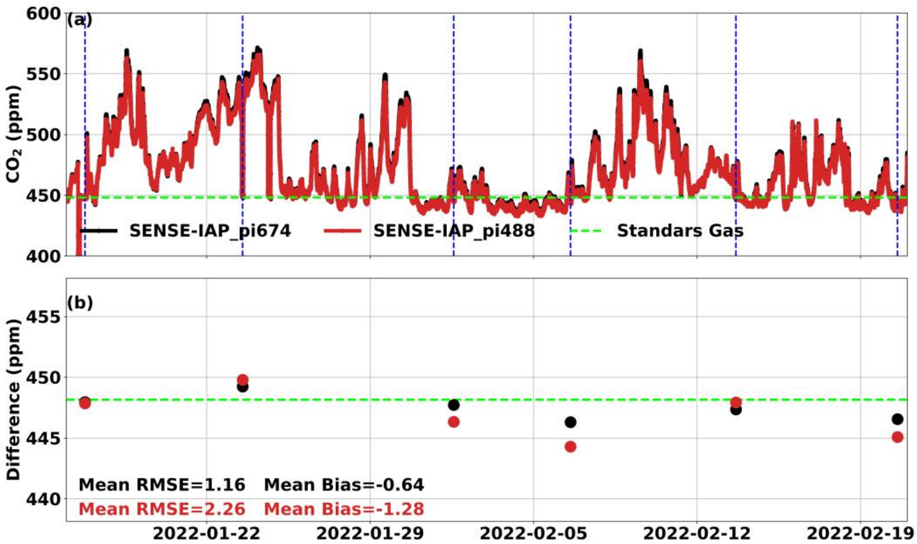

Figure 2: Comparison of CO_2 concentrations measured by SENSE-IAP and the CO_2 concentrations of standard gas at Beijing-CNEMC site in January and February 2022. (a) The time series of CO_2 per minute in the whole measurement period; the blue dashed lines mark one hour of the standard gas measurement per week, and the green dashed lines mark the concentration of standard gas. (b) The points are the hourly means of values during one hour of standard gas measurement per week, with mean biases -1.28 ~ -0.64 ppm and RMSE 1.16 ~ 2.26 ppm.

We also explored a remote calibration strategy under specific weather conditions by using reference concentrations from models, with a relative lower accuracy than the standard gas or reference instrument methods, whose general model-observation mismatch is 0-5ppm (below figure 3). The preliminary results were published in BAMS, showing the CO_2 concentration gradients observed in Beijing (Han et al., 2024). More results on the accuracy of the Beijing networks are in preparation and to be submitted in the following studies.

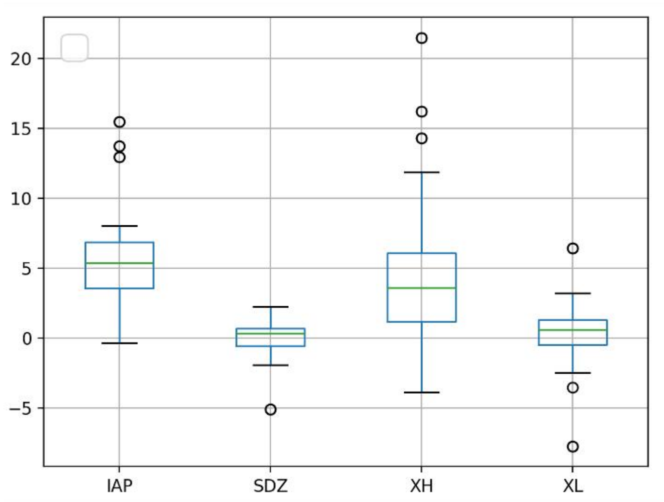

Figure 3. Model and Picarro mismatch at some favorable weather conditions and urban Beijing (IAP), suburban Beijing (Xianghe, XH), and background Beijing (Shangdianzi, SDZ, and Xinglong, XL) sites.

The calibration cost using automatic standard gas (1 week frequency) can reach \$300/station/year, which consumes 2 tanks of 8L 10 MPa gas (one tank work standard gas, and one tank target/quality-check standard gas). And the time cost is maintaining workers calibrating gas and sending the calibrated gas to the stations, generally 3-5 stations/day/worker depending on the distances of the stations from the laboratory. And thus 100s of sensors deployed will cost \$30,000/year, and 1 worker of 1-2 month time.

We added a literature review of the long-term drift calibration for LCS networks in Switzerland, Beijing, California, and Paris, lines 58-81 in section 1.

The data availability statement in my opinion does not follow the best practice for open research. Why have the authors not made their data and calibration codes readily available to all readers in an online repository?

Response: Thank you for this question. We provided all the Picarro and SENSE-IAP data used in this paper (publicly available at https://doi.org/10.6084/m9.figshare.29310890.v3). We further added more descriptions on the calibration algorithm and procedures in the main text (lines 133-141), and we disclosed our calibration method as much as we can, and several similar methods and precision were achieved in other studies (Lian et al., 2024; Shusterman et al., 2016). However, due to the commercial confidential requirements (Beijing, Jinan, Fuzhou, and several other cities' LCS networks we served), the detailed calculation formulas cannot be disclosed in this research paper.

Bao, Z., Han, P., Zeng, N., et al., 2020. Observation and modeling of vertical carbon dioxide distribution in a heavily polluted suburban environment. Atmospheric and Oceanic Science Letters, 1-9.

Han, P., Yao, B., Cai, Q., et al., 2024. Support Carbon Neutral Goal with a High-Resolution Carbon Monitoring System in Beijing. Bulletin of the American Meteorological Society 105 (12), E2461-E2481.

Lian, J., Laurent, O., Chariot, M., et al., 2024. Development and deployment of a mid-cost CO_2 sensor monitoring network to support atmospheric inverse modeling for quantifying urban CO_2 emissions in Paris. Atmos. Meas. Tech. 17 (19), 5821-5839.

Liu, D., Sun, W., Zeng, N., et al., 2021. Observed decreases in on-road CO_2 concentrations in Beijing during COVID-19 restrictions. Atmospheric Chemistry and Physics 21 (6), 4599-4614.

Martin, C.R., Zeng, N., Karion, A., et al., 2017. Evaluation and environmental correction of ambient CO_2 measurements from a low-cost NDIR sensor. Atmospheric measurement techniques 10.

Shusterman, A.A., Teige, V.E., Turner, A.J., et al., 2016. The BErkeley Atmospheric CO_2 Observation Network: initial evaluation. Atmos. Chem. Phys. 16 (21), 13449-13463.