Referee 1:

General comments:

Liu et al. tested and deployed low-cost NDIR carbon dioxide sensors. They successfully developed data processing, calibration and sensitivity correction protocols to improve instrument accuracy. This study is a critical first step towards lower cost monitoring options of marine carbon dioxide when using existing buoy platforms. Overall, the study is well-written and easy to follow. This manuscript is well suited for publication in AMT and will surely be of interest to experts and the wider readership of the journal. However, some details should be added when describing the data processing steps, see specific comments below. The calibration, cross-sensitivity corrections and data processing strategies appear to significantly improve sensor performance. However, the authors have not tracked this performance over the full lifetime of the sensors used. Re-testing would seem prudent. As this seems beyond the scope of this work, this limitation should be clearly stated in the manuscript to encourage future studies. Also, most figures could benefit from higher resolution. Specifically, the inlay text boxes seem somewhat pixelated.

We sincerely thank the referee for the positive evaluation of our study. Regarding the points raised, in the revised manuscript we have made improvements with each comment addressed individually. The limitation that the current study represents a short-term deployment and does not track sensor performance over the full lifetime has been explicitly stated, and we have highlighted that future work will evaluate long-term drift and operational robustness through extended deployments and repeated calibrations. Finally, all figures have been updated to higher resolution to ensure that inlay text boxes and labels are clear and legible. We believe these revisions fully address the referee’s suggestions.

Specific and technical comments:

Line 11: The reported numbers do not reflect the noise level of data at native temporal resolution, beyond the corrections mentioned here the data is also averaged. Should be mentioned here.

Response: We have revised the relevant sentence (Lines 8-10 in revised manuscript) in the abstract to: This approach significantly improved data accuracy: after 1-minute temporal averaging of raw data and a subsequent 1-hour moving average, the root mean square error was reduced from 8.03 ppm to 3.64 ppm in land tests and from 24.26 ppm to 1.59 ppm in marine observations.

Line 173: Clarification is needed. Raw data is collected at 2 s resolution. You refer to a 4-sigma filter here, but do not specify how much data is included and at which resolution. 4-sigma of 30 data points (1 minute of observations) can be quite different than 4-sigma of 1800 observations (1 hour of observations).

Response: We thank the referee for this important clarification. Raw data were acquired at 2 s resolution, and the 4-sigma outlier detection was applied within consecutive 1-minute windows (30 raw 2-s data points per window), rather than over an hourly timescale. We have stated Lines139-141 of the revised manuscript: After quality control, outliers deviating by more than 4σ from the mean were removed — this detection was implemented on the native 2 s resolution data within consecutive 1-minute windows (30 raw data points per window) — and temporal averaging was applied to reduce the noise level.

Line 179: You report that the optimal averaging time is 3 minutes, however later (line 183) you decide to average the data to 1-minute averages without an explanation. Why did you decide to use a suboptimal averaging time? Is there meaningful variability in the CO_2 Ocean sink (or source) below 3 minutes? Inverse modelling studies often end up using hourly averages anyways.

Response: We thank the referee for this important question. We have revised and clarified the rationale for 1-minute averaging in the revised manuscript (Lines148-153) as follows: Although the 3 minute interval yields a marginally lower Allan variance, a 1 minute averaging time was adopted in this study because the Allan variance is only slightly higher than at 3 min while allowing resolution of shorter timescale atmospheric variability (Martin et al., 2019). This choice ensures sufficiently low noise (0.4-0.7 ppm) to resolve marine CO_2 dynamics while preserving higher frequency variability associated with rapid coastal atmospheric and air sea CO_2 exchanges that would be smoothed over with 3 minute averaging, and supports detailed process analysis with the flexibility to aggregate to coarser scales as needed.

Line 217: Do you expect this calibration to be valid for the whole lifetime of the sensor? If not, it would seem appropriate to add a term that accounts for temporal drifts, e.g. due to the aging of the light source or decreasing sensitivity of the detector. This was included in previous studies, e.g. AMT - Characterization of a commercial lower-cost medium-precision non-dispersive infrared sensor for atmospheric CO_2 monitoring in urban areas

Response: We thank the referee for this insightful comment on sensor temporal drift and long-term calibration validity. We explicitly address why a time-dependent term (t) was not included in our multivariate linear regression model, based on our land-based validation results (Figure R1). Prior to environmental correction, a negative correlation between ΔCO_2 and time is observed for all sensors. However, this apparent temporal trend is not an intrinsic linear drift of the sensors: as confirmed by the temperature time series in Figure 7, the test period featured a gradual daily increase in ambient temperature, and raw sensor error is strongly inversely correlated with temperature. The observed correlation with time is therefore a direct artifact of temperature variability, not a linear aging effect. After applying our environmental correction model, no significant linear correlation between residual error and time remains (r = -0.01 to -0.02 for all CMs). This lack of linear correlation between the corrected residuals and time confirms that no meaningful linear relationship exists between sensor error and time; therefore, including a linear time term (t) in the regression model is not statistically justified. This confirms that for short-term deployments, intrinsic linear temporal drift is negligible. Critically, the lack of a consistent linear relationship between error and time means that adding a standalone time parameter t to the linear regression model would be statistically invalid and ineffective for correcting drift in this dataset.

We fully acknowledge the referee's critical point that for long-term (multi-month to multi-year) monitoring, sensor drift due to hardware aging (e.g., light source degradation, detector sensitivity reduction) must be addressed. As demonstrated in a 30-month field evaluation of low-cost CO_2 sensors (Cai et al., 2025), linear time terms cannot reliably capture complex long-term drift, and the most robust solution is periodic re-calibration.

For our coastal marine deployment, we will implement a 3-6 months periodic calibration protocol in future work. During periods of stable atmospheric conditions (minimal terrestrial influence and consistent air-sea exchange), we will use the MLO atmospheric CO_2 dataset as a stable reference to re-calibrate the CMs. This approach avoids the limitations of linear time-dependent correction terms and ensures the long-term reliability of sensor data, aligned with established best practices for long-term low-cost sensor monitoring.

We have added a discussion on temporal drift correction in Lines 189-196: It should be specifically noted that temporal drift of the sensor was evaluated (Figure S3). Previous studies have incorporated linear time terms into calibration models to account for sensor aging and drift (Arzoumanian et al., 2019). In the present study, however, the apparent temporal trend in raw error is associated with the gradual increase in temperature during the measurement period. After correction, no consistent linear relationship was observed between sensor error and time. Therefore, a linear time term (t) was not included in the regression model. For long-term monitoring, periodic re-calibration every 3-6 months using stable atmospheric CO_2 observations from background reference sites (e.g., MLO) during quiescent atmospheric conditions will be required to address non-linear temporal drift. Figure R1 is included in the SI for reference.

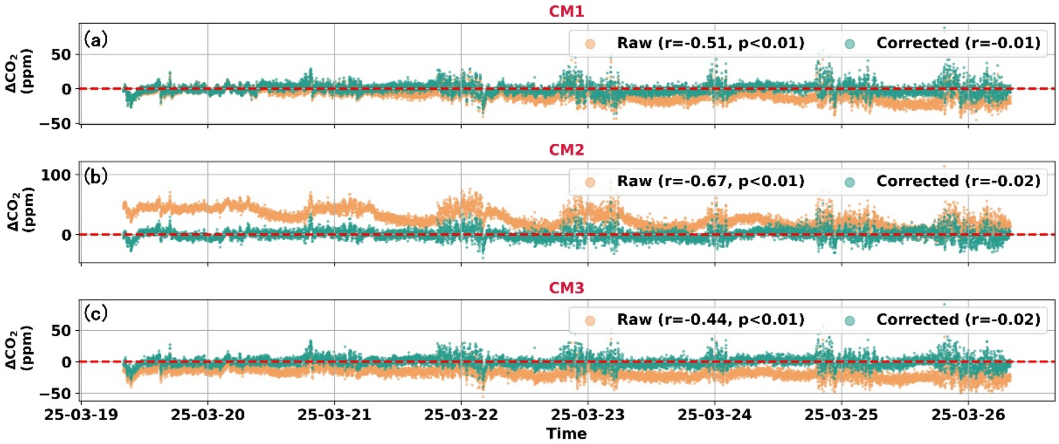

Figure R1: Variations of ΔCO_2 (ΔCO_2=CM-Picarro) for CM1 (a), CM2 (b), and CM3 (c) during land-based field observations, as functions of time. Orange dots represent data before environmental correction, while green dots represent data after environmental correction.

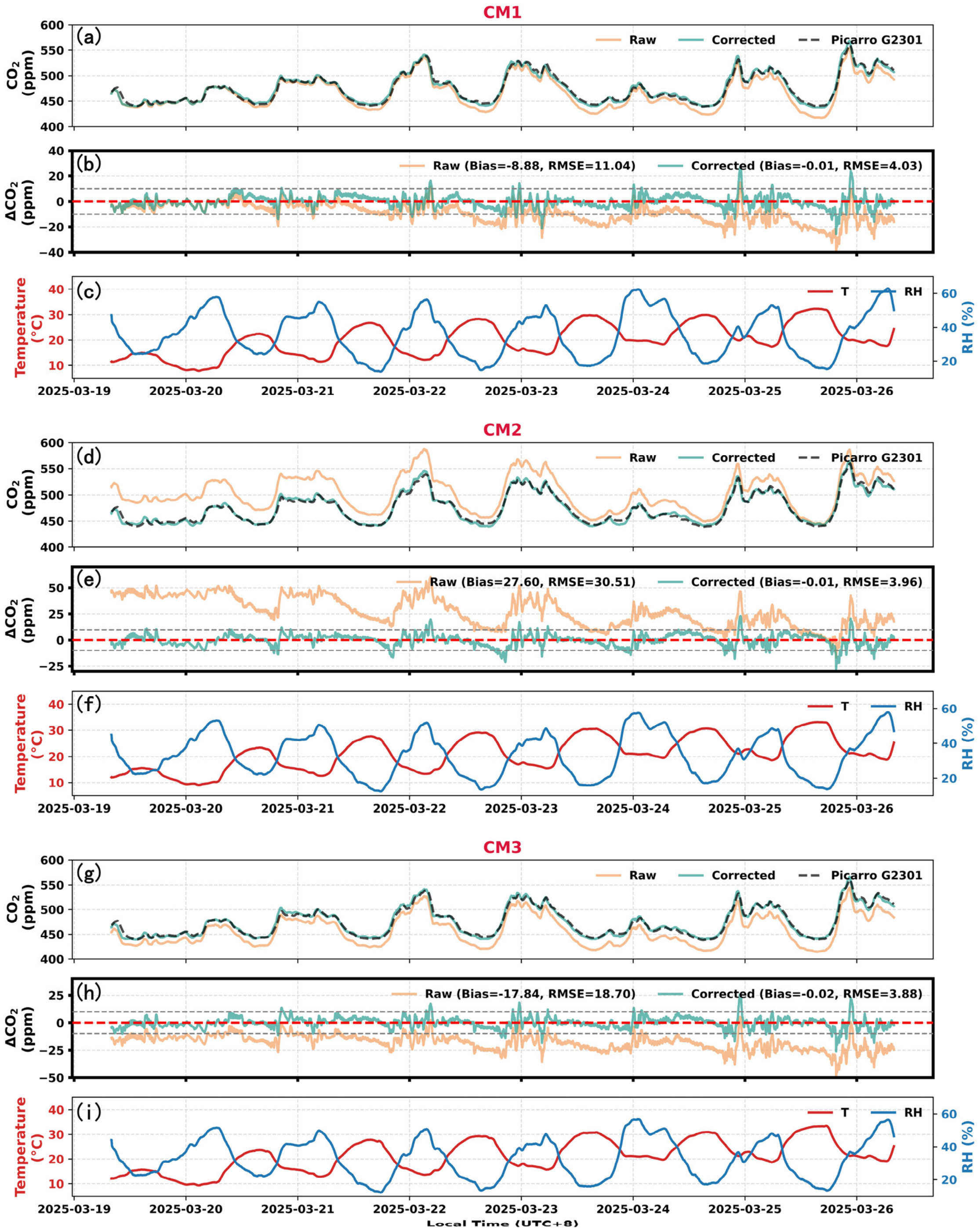

Figure 7: Hourly moving average time series of CM1 (a-c), CM2 (d-f), and CM3 (g-i) during land-based field observations: CO_2 concentration of CMs before environmental correction (orange line), after environmental correction (green line) , and CO_2 concentration from Picarro (black dashed line) (a, d, g); ΔCO_2 (ΔCO_2=CM-Picarro) of CMs before and after calibration (b, e, h); Ambient temperature (T, red line) and relative humidity (RH, blue line) (c, f, i).

Line 247: change - 0.39 to -0.39

Line 257: it seems some symbols got detached from the numerals

Response to Line 247&257: We thank the referee for these careful corrections. All unnecessary spaces between minus signs and numerical values (e.g., “- 0.49”, “- 0.39”,

"+ 0.1") have been removed, and the values are now correctly formatted as "-0.39", "-0.49" and "+0.1" in the revised manuscript.

Line 263: Have measurements been repeated after several months or years to ensure the reported sensitivities to environmental parameters are consistent over time?

Response: We thank the referee for this important question. To examine the temporal consistency of sensor sensitivities to environmental parameters, we performed post-deployment laboratory calibrations against the Picarro reference in March 2026, approximately 9 months after the initial June 2025 calibration. The results are summarized in Table 1.

Table 1: Post-deployment laboratory validation results

Data Type	Raw		Corrected		Re-Corrected	
Sensor	Bias	RMSE	Bias	RMSE	Bias	RMSE
CM1	5.90	6.54	-3.89	4.81	0.09	2.82
CM2	44.73	45.04	-14.62	15.14	0.25	3.92
CM3	1.90	2.45	-9.33	9.46	0.16	1.58
mean	17.51	17.79	-9.28	9.67	0.17	2.70

The results indicate that sensor sensitivities to temperature, humidity, and pressure remained fully consistent over the deployment period. After deployment, the mean bias and RMSE of the Corrected data was -9.28 ppm and 9.67 ppm respectively. By updating only the concentration baseline parameter a_0 (Re-corrected), the mean bias was reduced to 0.17 ppm and mean RMSE to 2.70 ppm. No changes to the environmental correction coefficients were necessary; only the baseline a_0 required adjustment to maintain accuracy. This validates the robustness of the environmental correction framework and supports periodic zero-point recalibration for long-term observations.

We have added Table 1 and the following discussion in the revised manuscript (Lines 381-383): Crucially, sensor sensitivities to temperature, humidity, and pressure remained consistent throughout the deployment. Re-correction by updating the baseline parameter a_0 effectively compensates for this temporal drift, reducing the mean bias to 0.17 ppm and mean RMSE to 2.70 ppm.

Line 292: there seems to be floating comma sign

Response: The unnecessary space between the closing parenthesis and comma in Line 253 has been removed; the punctuation is now formatted correctly as "),".

Line 316: the RMSE is significantly improved and lower than on the land-based test. However, it should be discussed how much lower the variability of ambient CO_2 values is in the marine test case compared to the land-based test. What is the standard deviation of the high-precision data for the testing periods?

Response: Thank for this important comment. We have added the discussion in Lines 273-275 of the revised manuscript as follows: The lower RMSE during the marine test can be partly explained by the substantially lower ambient CO_2 variability over the ocean, as reflected by the smaller standard deviation of the MLO reference data (1.81 ppm) compared to that of the Picarro in-situ measurements at the land site (29.29 ppm).

Line 403: Are you referring to additional mixing in the surface ocean or the atmospheric boundary layer?

Response: We thank the referee for this clarification. We have revised in Lines 344-346 of the revised manuscript as follows: As atmospheric temperatures rise, enhanced turbulent activity thickens the atmospheric boundary mixed layer, diluting CO_2 through vertical atmospheric mixing and reducing its concentration.

Line 407: Your finding suggests that the daily amplitude is 50% higher when using the CMs. How did you validate the assumption that this difference cannot cause a bias when data is used in atmospheric inverse models?

Response: We thank the referee for this rigorous and important question. We clarify at the outset that no high-precision in-situ CO_2 observations are available at our northern South China Sea coastal site to fully validate potential biases in an atmospheric inverse modeling framework. The MLO dataset provides only a remote open-ocean background reference rather than local ground truth, and the mean differences between the CMs and reference cannot be fully quantified under true coastal marine conditions.

Nevertheless, we can address the potential for systematic bias from two key perspectives. First, atmospheric inverse models are primarily sensitive to long-term mean values and temporal patterns (e.g., phase of diurnal cycle), rather than the absolute magnitude of diurnal amplitude. The ~50% higher amplitude observed in the CMs affects the peak-to-trough range but not the overall mean or the timing of diurnal variations, which are the primary constraints for flux inversions. Second, the larger diurnal amplitude is consistent with physical differences between coastal marine environments and remote open-ocean sites such as MLO, even after removing terrestrial influences. Coastal regions exhibit stronger diurnal evolution of the atmospheric boundary layer and air-sea exchange, which naturally enhance diurnal CO_2 variability compared to the stable open ocean.

On this basis, we have no evidence that the observed difference in diurnal amplitude introduces systematic, flux-altering bias when the data are averaged to the temporal scales typically used in inverse modeling. However, we fully acknowledge the limitation that formal quantitative validation against inverse modeling results cannot be performed without dedicated modeling experiments and long-term, high-precision in-situ constraints at the coastal site. Importantly, this limitation itself highlights a critical gap in current observational research: there remains a lack of extensive direct observations of atmospheric CO_2 concentrations over the ocean. Our study makes a key breakthrough in addressing this gap by developing and validating a feasible method for coastal marine CO_2 monitoring.

The further significance of our data for atmospheric inverse modeling—including verifying the impact of any potential small biases (e.g., the ~1 ppm uncertainty raised by the referee)—ultimately requires dedicated inverse modeling experiments, which in turn depend on a larger volume of high-quality observational data. The method developed in this study is therefore essential for acquiring such large datasets, laying the foundation for future inverse modeling work to fully validate the utility of our observations and quantify their impact on flux estimates. These will be important directions for future work, alongside extended observations to capture monthly, seasonal, and interannual variability.

Line 445: Please state clearly if this statistic applies to the native 2 seconds resolution or for 1-minute averages.

Response: Thank and this has been explained in Lines 390-392 of the revised manuscript as follows: With land-based Picarro G2301 co-located observations for comparison, CO_2 accuracy improved from 8.03 ppm to 3.64 ppm (corresponds to the 1-hour moving average of minute-by-minute data).

Line 447: Please provide a reference for the concentration range of 420-480ppm. Are you referring to daily, monthly or seasonal variations here.

Response: We thank the referee for this comment. The range 420-480 ppm is a visual estimate of the overall temporal variability observed across the entire measurement period, as shown in Figure 8a. It reflects the full range of concentrations during our observation campaign, not restricted to diurnal, monthly, or seasonal variations alone. As this range is directly derived from our own data, no external reference is provided. We have clarified this by adding “in this research” in Line 393.

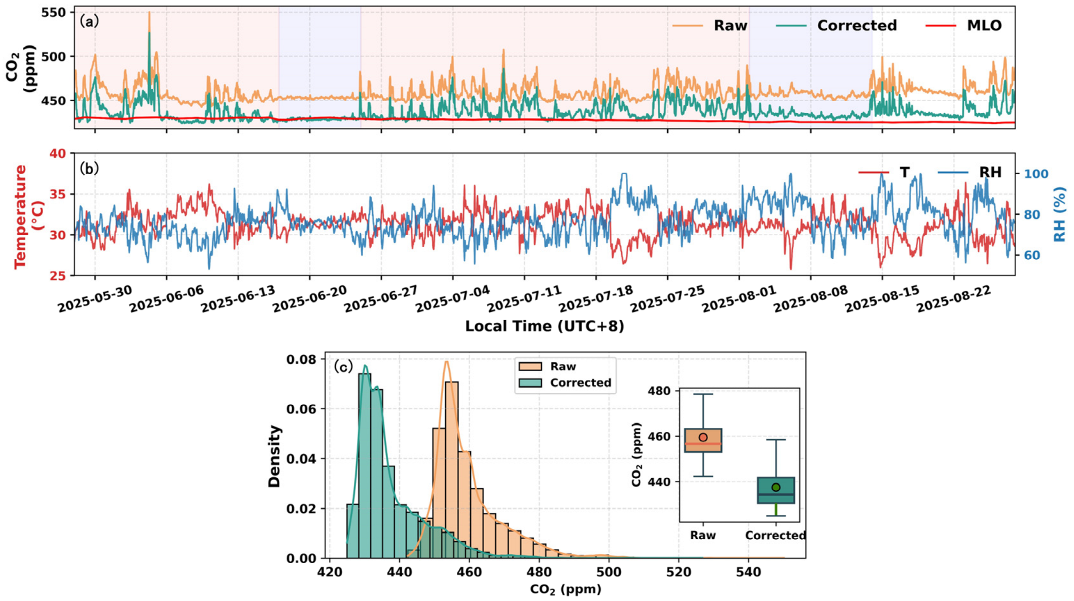

Figure 1: Offshore buoy observation results of CMs. (a) Hourly moving average time series of CO_2 concentrations from CMs before correction (orange line) and after correction (green line), together with daily mean CO_2 series from Mauna Loa Observatory (MLO, red line). The light red and light blue shaded backgrounds correspond to CO_2 fluctuation periods and stable periods, respectively. (b) Time series of ambient temperature (T, red line) and relative humidity (RH, blue line). (c) Histograms and boxplots showing the distributions of CO_2 concentrations before (orange bars) and after correction (green bars).

Line 472: What kind of long-term sensor drifts do you expect here? Aging of light source and sensor? If so, please consider adding or at least discussion the use of a coefficient for long-term drift (see comment: line 217)

Response: For long-term deployments, we do expect non-linear temporal drifts associated with light source aging and detector sensitivity degradation, as suggested by the referee. We fully recognize that addressing long-term sensor drift is critical for sustained atmospheric CO_2 monitoring over marine environments. We have therefore further discussed this issue in the outlook section of the revised manuscript (Lines 415-420): The current short-term deployment does not allow for full assessment of long-term sensor drift, which will require correction in extended observations. Long-term operation of NDIR sensors is expected to introduce non-linear temporal drift associated with light source aging and detector degradation. Since such drift exhibits no consistent linear relationship with time, a linear correction coefficient was not adopted in this study. Instead, periodic re-calibration every 3-6 months using stable background CO_2 observations under quiescent atmospheric conditions is recommended for future long-term deployments.