The manuscript describes a method for drift correction in a commercially-available closed-cell mid-infrared tunable diode laser absorption spectrometer by sampling reference gas from an on-board bag at certain time intervals. The drift correction method is demonstrated for four cases, including three cases in natural landscapes and one case for an anthropogenic agricultural source.

The manuscript rightfully identifies a challenge in UAV-based gas sensing. Drift of lightweight laser-based gas sensors for UAVs can be a real challenge, specially in UAV-based measurements, with quickly varying airflow and environmental conditions combined with short flight times not allowing the system to reach thermal equilibrium during the measurement time.
I recommend the publication of the manuscript after the following points are adequately addressed:

Demonstrating the drift correction method proposed is valuable. However, one point that is not adequately discussed is regarding possible solutions to mitigate or eliminate sensor drifts rather than adding procedures in post-processing of data, and more importantly in-flight sampling from a reference bag eats away from the precious flight measurement time. One such solution for drifts in the same line of product from the same manufacturer has been presented by Bolek et al. 2026 (See Appendix A). The authors should refer to such works and adequately discuss such solutions, their advantages and disadvantages vs. the proposed method.

What accuracy do the authors attribute to the instrument/measurements before and after the proposed correction method, and also compare with specifications from the manufacturer? For example, reading stability over a certain period of time as a percentage.

A laboratory study, with durations longer than typical flight times and with drifts larger than drifts encountered in the field, using a stable gas mole fraction (e.g. from a cylinder) is missing. Exposing the sensor to temperature swings in a climate chamber and switching on/off a blower can simulate exaggerated conditions to place bounds on the sensor performance. How stable would the sensor readings be in such an experiment before/after drift correction?

The choice of sampling frequency from the reference bag every minute assumes that the sensor's oscillatory behaviour always has a frequency that is ≤2/60 Hz. The authors have not demonstrated that they have established this for a statistically significant number of field/laboratory experiments. As seen in some instances an up/down-drift is missed.

The authors should include results of analysis showing the confidence level in detected differences from baseline after the correction method is applied, for each species. For example, enhancements >2 ppb CH4 are valid with 95% confidence, etc. and also comment on the magnitude of residuals after the correction is applied in typical flight conditions and/or worst-case scenarios for drifts/oscillations.

The math behind the drift correction needs to be included explicitly. Is the drift correction assumed to be a constant shift in the calibration curve or a scaling factor or some other relation?

It looks like the term “concentration” is used instead of mole fraction or mixing ratio. What do the sensors actually measure? ppm or ppb are unitless while concentration is not. Please consider changing concentration to the proper mole fractions or mixing rations throughout the manuscript.

Other comments:

L14: UAS acronym is not defined before. UAV is.

L75: sensor drift is said to affect detection limit, please clarify detection limit of what quantity, gas mole fractions or flux rate estimates?

L83-84: simultaneous measurement of GHGs is very general: species of interest are different in different fields, in O&G could be CH4, C2H6, CO2, H2O, in agriculture could be: N2O, CH4, CO2, NH3, H2O, for tracer release could be: C2H2.

L108: Regarding GPS drift issue. Effect of GPS altitude on flux estimate is discussed in Morales et al. 2022, please include a relevant discussion and if positional accuracy and uncertainty was included in the analyses when using GPS.

L129-131: provide more details of the relative calibration. Suggest adding a formula/equation for the procedure performed. Is the correction an addition/subtraction of the spline or a scaling factor? Or some other equation? And why? That is to say, would a detected enhancement from the baseline in the original sensor reading stay the same after the correction or scaled by a factor? In other words, have the authors assumed the sensor drift represents a constant shift in the calibration curve or a scaling factor or some other relation and what is the rational for that?

L133: “correcting the baseline…” how was it done? By subtracting the difference? Scaling by a factor? Etc.?

L142: loss of flight time because of the bag, the bag was 10L, what was the flow rates of the two sensors? Please include that information.

L145: was the frequency of the sensor drift measured in advance and was it repeatable so that it can be used to inform the frequency for sampling from the reference bag?

L152-155: It is reasonable not to sample volumes of air that are not influenced by a source. However please elaborate on how measuring at altitudes higher than the influenced altitude adds irrelevant noise? Those effects are expected to be small anyways since no enhancement from the baseline is expected and hence the flux should be near zero even though winds are stronger. Is it sensor drifts/noise that is amplified by the stronger winds, or some other effect? Please elaborate on the mechanism that adds irrelevant noise.

L167: please elaborate on how gas thermalization was ensured/estimated.

L179: add “to”: according “to” the modified mass difference method (MMD) …

L188-193: Figure 2B, the tracer seems to be detected at all transverse distances in the downwind plane, what is the wind condition like? Or is the figure just illustrative?

L202: please define the range for low and medium wind speed.

L222: please define ranges for medium-to-strong emission sources.

L226: suggest changing “land use” to “land features” or similar.

L234: suggest changing “stable” to “stagnant”.

L237: please improve the grammar in “we demonstrate a first simplified version of such mapping, …”

L242: In stagnant atmospheric conditions, the vertical transport can become more important than wind speed due to solar radiation and surface properties. Please clarify the underlying assumptions for a low vertical transport given the experienced atmospheric conditions.

L270: How is the normalization of wind speed justified given that the walls are measured at different times? E.g. in the two-wall case?

L280: is Xg(x,y) in Eq. (1) mole fractions or mixing ratio? As they are defined differently.

L289-294: How is the box flight case handled when the emissions are detected at the corner of the box and within the two sides of the box meeting at that corner, as Figure 2A suggests? i.e. was the drone hovering momentarily at the corner to change heading? Presumably the drone has to slow down to make the turn, that would mean more measurements in the corner per unit time and also more effect from the propeller downwash. Please discuss and clarify.

L299: Please provide range for estimated uncertainty for drift correction here. Or clarify if that error component is already included in the stated uncertainties?

L300: Is the uncertainty for altitude assumed to be similar to GPS latitude and longitude? Or another method is used to measure altitude? Often GPS units are less accurate for altitude unless RTK or multi-band GPS units are used. Please clarify. The effect of altitude uncertainty of estimated flux is investigated in Morales et al 2022.

L301: why the maximum altitude has been assigned a 10% uncertainty? Please clarify.

L307: Suggest using "effective" in "Dividing the total flux with the effective ground area" to include cases where the wind is not exactly perpendicular to (both) wall(s).

Section 2.5 on error estimation: the method used for error estimation and the associated figure (Figure 3) need more detailed explanations, e.g. in an Appendix, or references as it may not be accessible to all readers.

L329: Change "Aeris" to "Aeris Strato"

L346: Please clarify why a higher CO_2 level of 600 ppm was chosen? Is that representative of expected CO_2 levels in natural landscapes? If not, how does that affect the accuracy of the drift correction method?

Figure 4: The more frequent calibrations are later during the test where the higher frequency oscillatory behavior of the sensor(s) is generally less. Wouldn't it have been more reasonable to perform the more frequent correction measurements initially?

L352: change "highest frequence" to "highest frequency"

Figure 5: Also include timeseries data for CO_2 corresponding to Figure 5 before and after drift correction, in addition to colormaps.

L367: Please include flux estimates with uncertainties for the determination of the measurement area being source/sink, for each species, in parenthesis, for example.

L368: Please also comment on the CO_2 hotspot in the west corner of the measurement area.

L368: Table 1 does not strongly support the peatland being a source of N_2O with a flux estimate of 0.3 mg/s with an uncertainty of 0.9 mg/s. Suggest rewording to reflect this large uncertainty.

Figure 6: Please add a North indicator.

Figure 6: Please clarify that the interpolation used for the visualizations and if that is the same as the Kriging used for flux estimation?

L378: Please consider changing concentration to mole fractions or mixing rations throughout the manuscript. Figure 6 (and other figures) show color scales with ppm or ppb, but, technically speaking, that is not the units for concentration.

L390: Figures A7, A8 (and also A9, A10) lack error bars for the estimated flux, wind, and difference from background concentration/mole fraction ratios. It is crucial to include error bars as fluxes close to zero can be easily misinterpreted as sink or source if the associated error bars crossing the zero are not considered.
In addition, the lowest altitude in these figures (and other figures) has a marker at altitude zero. Please clarify how zero height was treated in the analysis/measurements. Was an extrapolation method used? Please include an explanation.

Table 1. For rewetted peatland, it is understandable that C_2H_6 was only used to track air motion, but wasn't the release rate measured, as described earlier in line 165? Please include the release rate in the table for completeness.

L421: is the 1ppb mentioned here even statistically significant given the high amplitude of oscillations in the sensor drift?

Figure 9: Visualizations for gas mole fractions are done differently from previous visualizations with a white background. Does that imply that color scales below a certain threshold are not significant? If so, please use the same visualization method in other figures as well.

Figure A2, L553: Did the authors consider fitting the spectra where the fast variations were caused by the sensor's online fitting procedure? What had caused the sensor fitting routine to be thrown off?
It is explained that after 1350s the flight corresponds to the second wall. Was the reference bag changed between the flights? If so, please include a more thorough description.

Figure A3: sampling from the bag does not seem to have high enough frequency for C_2H_6 indicating that tests for optimum reference sampling have not adequately captured the behaviour of the sensor's drift for different conditions.

Figure A4: CH_4 panel: the same comment as above. The sampling frequency from the reference bag does not seem to be enough <400s.