The below text contains the reviewer comments in black and author answers (AA) in blue

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:

Comments

C1. 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.

Thanks. We have now added this new reference to the manuscript in the discussions and in Table 2 listing recent UAV method development.

C2. 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.

This is a very interesting point. The drift correction method improves measurement stability and therefore also precision, especially if combined with calibration against high-quality gas standard in the lab which also corrects for offset in addition to drift. The drift is not linear for all the gases and can shift direction over time, so quantifying the stability could be done over a time period such as 5 minutes using 100x(Cmax-Cmin)/Cavg for that time period to get a percentage of maximum variation. In the manuscript we have now clarified the level of improvement yielded by the drift correction.

C3. 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?
This is an interesting question, but it falls outside the scope of the study. The aim of this work was not to evaluate sensor performance under laboratory conditions, but rather to correct for drift while testing approaches to measure fluxes in situ.

C4. 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.
We note that the Nyquist frequency corresponding to a sampling interval of 60 s is 1/120 Hz, rather than 2/60 Hz as stated in the comment. This means that we can fit drifts that changes direction ~every 60 s which is the case for many of our flights. If this is frequent enough will depend on the gas (different complexity of drift curves) and the sampling frequency of each flight (e.g. about 50 s in Fig. A1). Successful drift correction fits can be seen in Figures A1 and A2 for most of the gases and environments. The barn flight in Fig. A3 is an example of having a too long sampling interval for C_2H_6 (about 74 s) which affects mainly the drift correction in the first part of the flight where the drift is faster – it does however not affect the fluxes as C_2H_6 was only used as a tracer gas. Overall, it is a good point that the sampling frequency is important which we also try to highlight (e.g. Fig. 5).

C5. 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.
We have now included confidence levels for detected differences of the gases in the appendix, as well as the magnitude of residuals in different scenarios.

C6. 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?
We agree and thank the reviewer for the suggestion. We have now added more detailed method descriptions and equations in the supplementary information (a new Appendix B) that includes drift correction, the wind-speed correction procedure, segment-level

and total flux integration for the different flight paths, and the corrections for the UAV motion and attitude (roll, pitch, and yaw).

C7. 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.
We have now changed the word concentration throughout the manuscript to mole fraction where measurement values and units are provided, and that the Aeris Strato measures dry mole fraction (mentioned the first time that mole fraction is used). We kept concentrations in the general conceptual discussions that were not specifically related with the measurements.

Other comments:
O1. L14: UAS acronym is not defined before. UAV is.
Changed as suggested.

O2. L75: sensor drift is said to affect detection limit, please clarify detection limit of what quantity, gas mole fractions or flux rate estimates?
We have revised the sentence to avoid the detection limit concept. For the mole fractions, the drift lowers the ppm resolution, which in turn affects how low fluxes can be detected. It is not possible to specify the flux detection limit, because other factors such as wind conditions also affect the flux assessments.

O3. 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.
We have now clarified which gases are meant (CH_4, CO_2, N_2O)

O4. 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.
The positional accuracy was included in the Monte-Carlo uncertainty calculations. However, as the drift in altitude using GPS could be several meters during a flight we instead used air pressure to calculate the altitude (as we did not use an RTK).
We have now added the reference. and that we used air pressure to obtain altitudes instead of GPS.

O5. 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?
This has now been clarified in the new Appendix B; see the answer to C6

O6. L133: "correcting the baseline…" how was it done? By subtracting the difference? Scaling by a factor? Etc.?
See the answer to C6

O7. 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.
The loss of flight time is due to hovering while sampling reference gas air, this is independent of the volume of the bag.
The flow rates of the sensors are the same as they are similar models and use the same pump speed of 4.95 V. We have now added the flow rates of the two gas analyzers in the manuscript.

O8. 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?
The frequency of the sensor drift had been estimated to about 1/120 Hz from previous test flights and tested more thoroughly using different gas bag sampling frequencies in 3.1 (Figs. 4 and 5) to see that it worked in practice and to aid in the selection of sampling frequency. As can be seen, especially in Fig. 4, the needed sampling frequency vary between the gases (CO_2 requiring less samples and CH_4, N_2O, C_2H_6 requiring one sample every 78 s in the test; but to have some margin the use of 60 s seemed like a useful balance). This frequency 1/120 Hz was repeatable and sufficient for all flights.

O9. 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.
It is a combination of the remaining sensor drift, fewer flight altitudes used at higher altitudes giving more uncertain average concentration and wind maps, and amplification by stronger winds. This effect is small for relatively high-emission sources (e.g. from livestock in the barn) but can be substantial for low-emission extended sources such as the lake shore or the rewetted peatland.

O10. L167: please elaborate on how gas thermalization was ensured/estimated.
It was not estimated but thought of in relative terms compared to the alternative of direct release as purging through the water should, among other things, make the gas temperature change towards the water temperature. The aim was not perfect thermalization, but to make the release of gas simulate behaviors of naturally emitted gas. The change of the purge water temperature during the release period was not measured, but based on skin exposure, no dramatic changes in temperature was noted.

O11. L179: add "to": according "to" the modified mass difference method (MMD) …
Changed as suggested.

O12. 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?
This Figure is illustrative but using real data (although the real walls were not as perfectly straight as illustrated). The gas concentration wall maps are of CH_4, that is the reason for the extension along the whole wall as the vegetation at the shoreline is a source for CH_4. The arrow from the tracer marks where most of the tracer hits the wall (illustration).

O13. L202: please define the range for low and medium wind speed.
Approximate ranges added

O14. L222: please define ranges for medium-to-strong emission sources.
This concept is not associated with specific flux ranges but with types of flux sources such as wastewater treatment, farm emissions etc. We have now revised the sentence in the manuscript to clarify this.

O15. L226: suggest changing "land use" to "land features" or similar.
Changed as suggested.

O16. L234: suggest changing "stable" to "stagnant".
Changed as suggested.

O17. L237: please improve the grammar in "we demonstrate a first simplified version of such mapping, …"
Changed as suggested.

O18. 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.

We have clarified this: “Under nocturnal inversion conditions and with little or no advection, gas emitted from the target area, representing a low point in the landscape, accumulates near the ground”.

O19. 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?
For both the box flights and the two-wall flights, sampling is done at different times as the flight time is ~25 min and the wind fields change during flight. This is why we need the wind speed normalization, as the wind fields will otherwise cause the mass balance to violate conservation of mass, and as there is no pressure buildup, airflow should be conserved. The two walls in the two-wall case are flown in rapid succession using two flights under stable, close to steady-state wind conditions with the assumption that the wind fields average values do not change significantly during the flight.

O20. L280: is Xg(x,y) in Eq. (1) mole fractions or mixing ratio? As they are defined differently.
Mole fraction is used in Eq. (1). This has now been clarified in the manuscript.

O21. 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.
The drone is moving at the same speed for all sides and altitudes of the box, however, when getting close to a corner the drone quickly stops, hovers both to change heading and to do a bag air drift calibration (about 18 s). During the first 9 s only bag air is used, then 9 s with outside air that will not be used (as these measurements still contain some bag air; marked in the logger) while at the end of this period starting to move the drone to be ready for measurements back at the same speed again as before a corner. This means that we do not get more points in the corners of the box.

O22. L299: Please provide range for estimated uncertainty for drift correction here. Or clarify if that error component is already included in the stated uncertainties?
The uncertainty of the drift correction is different between the gases and for every flight, so it is difficult to give a single range for all gases. We have however now given examples of typical drift correction uncertainties in the manuscript. See also the answer to question C2 for typical uncertainties for the different gases before and after drift correction.

O23. 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.
Longitude and latitude: GPS was used and uncertainties found from the drone logger, an uncertainty that depends on the number of satellites used (often 32, resulting in ±0.5m for both longitude and latitude).
For altitude - see the answer to question O4

O24. L301: why the maximum altitude has been assigned a 10% uncertainty? Please clarify.
We use the ethane tracer to estimate the vertical dispersion of emitted gas at that distance, with the tracer placed in an area of interest to be measured (at the most upstream part). The upper ethane detection on the flight wall was used to determine the altitude for flux integration. However, this is a diffuse limit with some uncertainty as some gas (changing wind speeds) will reach higher or lower on the wall. Based on ethane plume appearances a 10% uncertainty in the top altitude for flux integration was approximated and used in the Monte-Carlo calculations. This has now been clarified in the manuscript.

O25. 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).
Changed as suggested.

O26. 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.
We have now added more detailed method descriptions and formulas, including more details on the uncertainty calculations, in the supplementary information (a new Appendix B) that also includes the wind-speed correction procedure, segment-level and total flux integration for the different flight paths, and the corrections for the UAV motion and attitude (roll, pitch, and yaw). Equations for the drift correction are also included.

O27. L329: Change "Aeris" to "Aeris Strato"
Changed as suggested.

O28. L346: Please clarify why a higher CO2 level of 600 ppm was chosen? Is that representative of expected CO2 levels in natural landscapes? If not, how does that affect the accuracy of the drift correction method?
It was not chosen to be 600 ppm, but on purpose we had somewhat higher levels than the typical 400 ppm background to make it more clear in method development that we

were actually sampling from the bag or from outside. The method works if having ambient levels in the gas bag as well, but having somewhat higher levels do not affect the accuracy of the drift correction as sufficient time is used for reaching almost 100% bag air (9 s) and to go back to outside air (9 s; outside air data during this transition period are not used) it just makes the method more graphically easy to follow.

O29. 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?
It would have been a good idea, but we had not found this out yet when doing the measurements. It is however evident from the plots (e.g. Fig. 4A) that the frequency of the drift variations would be well sampled with a 60 s sampling rate. See Fig. A2 where successful drift correction were made throughout the flights.

O30. L352: change "highest frequence" to "highest frequency"
Changed as suggested.

O31. Figure 5: Also include timeseries data for CO2 corresponding to Figure 5 before and after drift correction, in addition to colormaps.
This is very similar to Fig. 4C, and unless we misunderstand the suggestions, we hope that Fig. 4C is an adequate example. We now refer to Fig 4C from the legend of Fig. 5.

O32. L367: Please include flux estimates with uncertainties for the determination of the measurement area being source/sink, for each species, in parenthesis, for example.
We have now added a reference to Table 1 (containing fluxes for the different species with uncertainties) and also changed the section to be more clear when we talk about concentrations and fluxes:
"In this example, a rewetted peatland was targeted using a box flight up to 30 m altitude (Fig. 6). The concentration patterns across the walls are shown in panels A–D. The flux estimates in Table 1 indicate that the rewetted peatland itself was a net CO_2 sink and a source of CH_4 during the measurement period. The estimated N_2O flux was positive, but its uncertainty interval crossed zero, preventing a definitive source–sink classification. The spatial distribution of the major CH_4 concentration features, accounting for the wind directions at low altitudes, correspond well with streams and the wettest areas, consistent with these areas contributing CH_4 to the measured air (panel A). For example, the C_2H_6 tracer-release, close to the upwind wall (marked in Fig. 6C) near a small waterfall, yielded elevated concentrations (panel C; northwest wall) that were co-located with elevated CH_4 concentrations (panel A). Trees 3-4 m tall were present in the upper left part of Fig. 6 and may have channeled near-ground airflow along the edges of that area. Figure A1 shows the in-flight instrument drift correction for the rewetted

peatland area, and Fig. A5 shows the interpolated wind speeds perpendicular to the walls."

O33. L368: Please also comment on the CO2 hotspot in the west corner of the measurement area.
Thanks for pointing this out. This is an artefact from the Aeris gas analyzer and has now been removed in the Figure.

O34. L368: Table 1 does not strongly support the peatland being a source of N2O with a flux estimate of 0.3 mg/s with an uncertainty of 0.9 mg/s. Suggest rewording to reflect this large uncertainty.
See the answer to O32.

O35. Figure 6: Please add a North indicator.
This has now been added to the Figure (north is up).

O36. Figure 6: Please clarify that the interpolation used for the visualizations and if that is the same as the Kriging used for flux estimation?
Yes, it is the same as the Kriging used for flux estimation. This has now been clarified in the text.

O37. 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.
See the answer to question C7

O38. 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.
Error bars have now been added to the Figures. The lowest altitude of zero is the starting altitude of the drone (ground level). We do have measurements at less than 1 meter altitude that are included in the 1x1 m wall grids and the calculations. This has now been clarified in the manuscript

O39. Table 1. For rewetted peatland, it is understandable that C2H6 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.

Unfortunately we lack quantitative C_2H_6 release rates for the rewetted peatland. This has now been clarified in Table 1.

O40. L421: is the 1ppb mentioned here even statistically significant given the high amplitude of oscillations in the sensor drift?
There are some peaks in Fig. A2, lower right panel (N_2O). that are seem to be above the residual drift background (e.g. at 700 s). But we cannot be sure of these patterns as they, at least some of them, are close to the residual drift in amplitude. We have therefore added the following:
"A very slight potential increase in N_2O-levels (about 1 ppb) can also be seen in the same area of the outgoing wall that corresponds to the CH_4 from the littoral zone."

O41. 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.
No, this does not imply that color scales below a certain threshold are not significant. Is it just that the figure uses a different color scale that works well for higher emissions with clear plumes.

O42. 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.
It is unknown why the Aeris Strato sensors had problems with the fitting routine. We did not use the parts with bad fits for any wall plots or flux calculations.
Yes, the bag was filled with ambient air between flights which can cause a shift in bag concentration (which is compensated for as described in the text). We have now added more info on the bag change being made in Fig. A2

O43. Figure A3: sampling from the bag does not seem to have high enough frequency for C2H6 indicating that tests for optimum reference sampling have not adequately captured the behaviour of the sensor's drift for different conditions.
It is correct that the frequency was too low in this case, but it does not indicate the optimum reference sampling frequency was too low. This flight used a longer time between calibrations at some altitudes than the suggested 60 s as this area was flown before we did the test for the minimum frequency. However, it is only C_2H_6 that is affected, and that one is only used as an air tracer and a clear peak can still be seen.

O44. Figure A4: CH4 panel: the same comment as above. The sampling frequency from the reference bag does not seem to be enough <400s.

This flight had transects taking a bit longer than expected because of the proximity to bushes and low altitude flying so some drift correction points are further apart. This is in the morning after a wind-still night with elevated CH_4 levels (peaks seen in Fig. A4, panel CH_4) that are real as they also overlap with structures (stream hotspots) and agrees with adjacent transects (Fig. 10).