We thank the reviewer for taking the time to review our manuscript and for some very useful and constructive comments and suggestions. We have responded (red text) to each comment (black italic text) in turn below. We have used these comments and suggestions to update and improve a revised version of the manuscript.

Review

Testing the near-field Gaussian plume inversion flux quantification technique using unmanned aerial vehicle sampling

Adil Shah et al.

The authors use a near-field Gaussian plume inversion (NGI) technique to determine the strength of methane emission sources. In that context, they describe, characterize and deploy two commercial near-IR off-axis integrated cavity output spectroscopy (ICOS) spectrometers from ABB/Los Gatos Research. One instrument is ground-based and samples through a tube from the UAV, while the other, more lightweight instrument is deployed onboard the UAV. The authors rightly claim that more reliable techniques are needed to determine emissions from diffuse methane sources. The paper has two main technical sections: chapter 2 "Methane instrumentation and calibration" and chapter 3 "Method Testing". After struggling hard with chapter 2, I decided not to continue the review, and I suggest that the paper must be fundamentally rewritten, with respect to both language and content, before publication in AMT.

We regret that the reviewer has identified a number of issues with section 2. We have worked hard to address all of the specific points raised by the reviewer by making significant modifications to section 2 of the revised manuscript We hope that this improves its readability.

Chapter 2 has four subsections:

2.1 Instrument inter-comparison (20 lines) Jonas, please check # lines

2.2 Water vapour correction (61 lines)

2.3 Calibration (28 lines)

2.4 Methane enhancement and uncertainty (20 lines)

Section 2.1 d contains very little actual intercomparison, and thus the title is misleading The inter-comparison consists mainly of one table containing the manufacturer's specifications and the key numbers of one Allan Variance plot (based on laboratory measurements), which is given in S1. In the context of this paper, which attempts to use the instruments under harsh field conditions, much work could and should have been done to provide useful, real-world characterization under field conditions, beyond what the manufacturer reports. This may include (but should not be limited to) performance characteristics and comparison (see section title) of long-term stability, stability of calibration factors, temperature and pressure dependence, sensitivity to vibration (UAV), linearity, selectivity, (measured) response time, concentration range, etc.

We agree that the title of this section may be misleading. We have therefore renamed the section title, “Instrumental overview”, to reflect the more general scope of the comparisons made in this section. As the reviewer rightly points out, the Allan variance plots are useful and have now been moved to the main manuscript.

Unfortunately, we lacked the facilities to replicate some UAV sampling conditions, such as UAV vibration, in the laboratory. However, we can see no reason why vibrations this should affect measurements, especially for the MGGA which was stationary on the ground

We disagree that we sampled in harsh conditions for this work. We sampled in the bottom 50 m of the planetary boundary layer, where pressure and temperature changes are relatively small compared to laboratory conditions, especially in temperate summer months. To better convey this, the sampling conditions have now been plotted alongside the calibration conditions, for both instruments, in the supplement. Furthermore, as suggested, we now discuss the very small potential effect of temperature and pressure variations on measured methane mole fraction for the MGGA in the final paragraph of section 2.1, which showed no discernible impact on measured methane mole fraction with changes in the ambient range of cell temperature and cell pressure during the Allan variance test.

We believe that the temporal stability of calibration factors was already captured in section 2.4, where the standard deviation in offset and gain factor is given over a long calibration period in the laboratory. However the reviewer rightly points out that we failed to capture the uncertainty due to linearity which we have now determined and included in the revised manuscript in the final paragraph of section 2.4.

The measured response time depends on the length of the inlet, as air takes time to travel through the external and internal tubing. This was measured in the field. Therefore these details are provided in paragraph 3 of section 3.1 where it is more relevant to the specific instrumental set-up.

Section 2.2 is the core part of the instrumental section, which is reflected by its length. It covers an in-depth description of water vapor correction but sadly contains a number of fundamental flaws. Some examples are below, but this list is not exhaustive:

- The largest correction to obtain dry CH_4 values ($[X]^{dry}$), needed for flux calculations, is water vapor. While the authors do many tests and calculations (see below) to get this value, they never show how accurate the H_2O measurements of the instruments are. They likely have the information (based on the measurements with a dew point generator), but the corresponding performance is never shown. For such a detailed water vapor study, it would be key to deduce this effect first and then discuss (and correct) the remaining influencing parameters in a second step.

The reviewer rightly points out that we do not accurately derive water vapour mole fraction in the cell. As we conduct an empirical water correction, which by its nature is based on reported methane mole fraction, it is not necessary to derive water mole fraction accurately, based on the conclusions of Rella *et al* (2013). However this empirical correction requires the water vapour mole fraction measurement to be stable and independent of methane mole fraction. As it was observed that the water baseline was in fact affected by methane mole fraction, our water baseline corrected for this systematic effect. In order to emphasise the fact that our water correction is purely empirical and is not based on spectroscopic first principles, we have renamed the section “Empirical water vapour correction” and we have redefined v as an “empirical water correction factor”. To be clear, we do not attempt to characterise the suitability of these instruments for measurement of water vapour concentration, which would require a separate evaluation focussed on water vapour with a different (high accuracy) water reference instrument. This may constitute future work, but the scope of this paper is on methane measurement only, with water vapour as a “problem” to be accounted for.

- Figure S3 shows the effect of CH_4 *on apparent water, due to cross sensitivity of the analyzers. Equation 1 has no physical meaning or motivation, but it produces an exponential curve, which fits the data points nicely. If there is no clear motivation, such a weak set of data may as well be described by a linear interpolation (see S3). Furthermore, applying the same parameter (*ϖ*) for the other analyzer, which shows clearly different absolute and relative behavior (S4), is not justified.*

The reviewer is correct to point out that the exponential baseline fit was arbitrary and was used to fit the data well. On reflection, we realise that the WMO calibration scale only extends up to 5 ppm, so any water baseline correction above 5 ppm is meaningless, without an equally good test for linearity up to a higher mole fraction using certified standards. To this end, we agree with the reviewer and have now applied a linear regression to the baseline from both instruments, over the range of the WMO scale.

As we observed the water baseline to behave non-linearly at 100 ppm methane mole fraction in the MGGA, we have also applied a linear fit including the additional 100 ppm data. This allowed us to assess the change in mole fraction measurements with the 100 ppm data included in the baseline to be, on average, (0.02±0.03)% higher for all MGGA sampling recorded from UAV1. This analysis is now presented in the supplement.

- Somewhat similar is the issue in S5 and S7. It is not surprising to find a (nearly linear) dependence between v *and water vapor. It is also not surprising that the residuals become a bit smaller if using a second order polynomial. However, since there is no information on the long-term stability of this system (e.g. repeatability and reproducibility), these fit characteristics are of very little use (a larger order polynomial would give even smaller residuals).*

The choice of a second order polynomial fit was based on polynomial water corrections used in previous work, using similar spectroscopic techniques (O’Shea *et al.* (2013), for example). Furthermore, O’Shea *et al.* (2013) (who used the same off-axis ICOS technique but a different instrument) found that v remains consistent over time with negligible temporal drift. This valuable point has now been made in the final paragraph of section 2.2.

- The authors call v *the water correction factor. This, again, is misleading because it contains normalization to dry conditions, which is a standard procedure, and corrections for non-perfect spectroscopy (which is specific to their analyzers).*

We realise that this may be misleading and have therefore renamed it an “empirical water correction factor”.

- Based on these badly justified fits and corrections, the authors deduce a water uncertainty factor (σ_v*). However,* σ_v *is the standard deviation of the mean of the residuals in S6/7. This parameter carries information about the quality of the fit of the polynomial, but no information about the quality of the water correction for a specific measurement at a given* CH_4 *and* H_2O *mole fraction.*

The reviewer rightly points out that our uncertainty is based on the quality of our polynomial fit. We simply cannot derive an uncertainty on something that is already uncertain. We have to assume that the uncertainty is represented by the goodness of the fit we can produce from empirical data. We have used the same approach as Rella *et al.* (2013) and O'Shea *et al.* (2013), as guidance of optimal practice. We have therefore clarified this in the final paragraph of section 2.2.

Sections 2.3 and 2.4 describe the calibration procedure and derive measurement uncertainty.

- Calibrated enhancement in methane mass density (E) is calculated based on parameters that are obtained in the above experimental and fit procedures. The procedures themselves are correct (but see some remarks above and below). The authors then state that their approach is especially useful because it can be used to determine measurement uncertainty (equation 11). However, this uncertainty calculation is based on laboratory measurements and badly motivated fit procedures. It does not contain non-linearity nor any changes (drifts) in time, e.g. of G and v. Furthermore, it does not include any field (in)stability considerations, nor the fact that one parameter is just assumed to be the same in MGGA and pMGGA. A realistic uncertainty budget would be highly welcome, but the approach chosen here is useless.

We thank the reviewer for highlighting the fact that we had not previously accounted for non-linearity. This has now been quantified up to 5 ppm by introducing a new non-linearity uncertainty parameter (σ_L), which has been incorporated into the enhancement uncertainty, as suggested.

With regards to drifts, the variability in G was already included in the enhancement uncertainty, represented by σ_G. As σ_G was derived from changes in G over time, we believe that it implicitly captures temporal drifts. We have now made this clear in the second paragraph of section 2.4. Additionally, we assume here that v does not drift, based on previous work using a similar instrument described by O'Shea *et al.* (2013).

Some less fundamental comments are below, illustrating that the whole paper should be strongly re-vised:

We have addressed all of the reviewer's comments and we have significantly revised the manuscript to take these comments into account.

Abstract

Page 1 – Line 12

Many parameters contribute to the uncertainty. Are they large or small, and is the contribution due to the "maybe" poorly quantified sources important or neglectable?

This sentence simply motivates the background to the paper. We are not tackling the global methane budget problem head on in the work. The choice of the words "may be" is due to the large uncertainties within the source and sink terms in global methane budget and the widely recognised difficulty in attributing the cause of these uncertainties. Section 1 expands on this and cites relevant papers that discuss the global methane problem. To make this clearer for the purposes of our manuscript and the abstract, we have added the words "potentially" and "many", so the sentence reads, "Methane emission fluxes from many facility-scale sources may be poorly quantified, potentially leading to uncertainties in the global methane budget". We hope this clarifies our point of view.

Page 1 – Line 14

This is not a NEW near-field Gaussian plume inversion (NGI) technique – it has been published before by the same authors.

The word "new" has been removed. We instead claim that this is the "first test" of the NGI method using "unbiased sampling".

Page 1 – Line 20

Modified with respect to what?

This is a good point. It was not previously clear what this was with respect to. The sentence has been rephrased to make it clear that the water correction was specially adapted for the instrument. The sentence now reads, "a water vapour correction factor, specifically calculated for the instrument, was applied and is described here in detail".

Page 1 - Line 23

Simplify sentence, e.g. the uncertainty was between ... and ...

We appreciate that this sentence can be simplified. However the quoted values are not uncertainties but rather uncertainty bounds as a fraction of the controlled emission flux. We have rephrased the sentence to simplify it, whilst maintaining the correct meaning.

Page 1 – Line 25

The term "range" implies that it is the range between 17 and 218 %, but this is not meant here.

The term "These highly conservative uncertainty ranges" has been replaced by "This range of highly conservative uncertainty bounds". This is now less ambiguous.

Page 1 – Line 27

"flux approaches": check language

The term "flux approaches" has been replaced by "flux quantification approaches".

Page 1 – Line 27

Replace "may perform well" by "may be a valuable alternative"

This replacement has been made.

Page 1 – Line 28

"applied to UAV sampling" should read "combined with UAV sampling"

This replacement has been made.

Introduction

Page 2 – Line 53

precision should read accuracy or precision and accuracy

The suggested change has been made.

Page 2 – Line 54

have not yet

This replacement has been made.

Page 2 – Line 60

This suggests that controlled release is the only acceptable method. Make statement more general, e.g. the method was not validated, and its uncertainty was not quantified (e.g. with controlled release experiments).

Although some methods can be mathematically correct on paper and conceptually sound, we stand by the idea that any method should be tested by sampling emissions of an accurate known flux rate to prove its utility. However we acknowledge the reviewer's point that a controlled release from a gas cylinder is not necessarily required and other approaches could be adopted, so long as a flux is known. We have rephrased the sentence so it reads, "However this method was not tested for UAV sampling with an accurately known (controlled) methane flux rate", to make it more general and allow for testing using any known flux.

Page 2 – Line 62

Make it clear that "Shah et al" is not any other researchers but the same group, e.g. our previous study...

The suggested change has been made.

Page 2 – Line 65

Language: the sampling does not develop the technique. e.g.: ... sampling ... used in combination with ...

The downwind sampling used during this previous study was used to develop our flux quantification technique. The technique did not exist before the sampling campaign. The measurements we acquired showed us that a new flux quantification approach was required, as previous approaches failed. We have now added this point to the sentence, to make it clearer that the new approach was a consequence of the failure of other approaches. We have also clarified that the "data-set" from the previous sampling campaign was used to develop the new method.

Methane Instrumentation and Calibration

Page 3 – Line 97

σ_{AV} is a very unusual parameter. Contains very little information because it depends on averaging time. "Allan deviation uncertainty factor" is confusing.

We realise that the Allan deviation at the maximum sampling frequency (σ_{AV}) is not so useful for direct instrument comparison, which is why we also provide the Allan deviation at the same frequency (both 1 Hz and 0.1 Hz) in Table 1. For our purposes, as both instruments sampled at their maximum sampling frequency (or minimum averaging time) and these were different for each instrument, σ_{AV} is used to quantify instrumental noise (for the instrumental averaging time used during sampling) and serves as a component of the total mole fraction enhancement uncertainty. We have renamed σ_{AV} as "sampling noise uncertainty" to make this less ambiguous.

Page 3 – Line 108

"These three effects have a net effect of decreasing $[X]_0$ in both instruments". You probably mean "decreasing [X]". Furthermore, this is only useful information if you state before which effects have a positive/negative impact. Whether the sum is a net negative effect depends on the conditions (usually dilution dominates, so the statement is not wrong).

We feel there may be some misunderstanding here in terms of the effect of water vapour on $[X]$. $[X]$ is a true mole fraction measurement (*i.e.* the true dry mole fraction in the atmosphere), independent of the instrument, whereas $[X]_0$ is an instrumental reading which can be different to the true value because it may be affected by conditions such as atmospheric water vapour and is not calibrated.

Regarding isolating the individual water vapour effects on $[X]_0$, we do not deem it empirically possible as what the instrument observes is decreasing $[X]_0$ with increasing water mole fraction. We have described all three effects in the manuscript to provide a background and motivation for our correction procedure. We believe it is unnecessary to fully characterise each of the three individual effects if the combined effect can be adequately characterised empirically. We have now clarified that we observe consistent decrease in $[X]_0$ at a range of methane and water mole fractions.

Page 3 – Line 110

If the pressure broadening coefficient should account for effects two AND three, then state so explicitly.

We have now clarified this in the first paragraph of section 2.2 that spectral overlap is expected to be an issue. Therefore we have now made it clear that the algorithm only corrects for pressure broadening in paragraph 2 of section 2.2.

Page 4 – Line 122

Has a calibration for water vapor been performed, and what are the results?

This is addressed in our response to the comment regarding Section 2.2 above. In summary, we do not characterise this instrument for water vapour measurement.

Page 4 – Line 132

"more points" means: 1.9, 2.1, 5.0 and ca. 104 ppm. It does not make sense to derive parameters of a (complicated) exponential function if there is no physical concept that supports this function.

We agree and have replaced it with a linear regression, over the range of the WMO scale (up to 5 ppm).

This H_2O baseline is very large; your data in S3 suggests that 100 ppm (dry) CH_4 results in ca. -5500 apparent ppm H_2O. Is this analyzer malfunctioning and should be returned to the manufacturer, or is it within the specifications for water vapor measurements? Using the MGGA H_2O baseline function "as-is" for another analyzer does not make sense, especially since the experimental values for the pMGGA (S4) are significantly different from those of the MGGA, indicating that the two functions may well not be the same. Alternatively, one may just acknowledge this fact and do a corresponding uncertainty budget. In the context of the large uncertainty of flux estimation with NGI, this may be appropriate. However, it would fundamentally change the arguments and line of thought in this paper.

We do not evaluate the performance of this instrument for water vapour measurement, which is beyond the scope of the manuscript. In this work, we are interested in the effect of water vapour on our target gas, methane. Water vapour influence is therefore something we simply seek to account for. However, the instrument specifications suggest that the instrument is designed to sample up to 0.03 $mol_{water}\ mol^{-1}$, with a manufacturer specified accuracy of 0.2 $mmol_{water}\ mole^{-1}$ at 1 Hz. The objective of characterising a water baseline was to ensure that the subsequent water correction was independent of the effects of methane in the cavity.

Regarding our choice of baseline, we agree with the reviewer and now use fully independent linear baselines for each individual instrument.

Page 4 – Line 135

"Having established a well characterised water baseline…"

This baseline may be well characterized. However, it corresponds to a very large correction, which can-not easily be redone in the field. Therefore, it is necessary to measure the longterm stability of the baseline and instrument-to-instrument variability. Otherwise the "well characterized" is not helpful.

This is a very useful point. We are confident that that the water baseline is relatively stable. To test this, we also conducted an Allan variance test on the dry water baseline for both instruments, at 2 ppm of methane. This revealed that the Allan variance of the MGGA and pMGGA (at the averaging time used to derive the baseline) was $\pm 16 \cdot 10^{-6}\ mol_{water}\ mol^{-1}$ and $\pm 27 \cdot 10^{-6}\ mol_{water}\ mol^{-1}$, respectively. This information is now added to paragraph 4 of section 2.2. These water baseline values are far smaller than water mole fractions sampled during the water correction procedure of up to $20\,000 \cdot 10^{-6}\ mol_{water}\ mol^{-1}$, also at 2 ppm methane.

Page 4 – Line 136

What was the mole fraction of CH_4 in this cylinder?

The methane mole fraction was 2.205 ppm for the MGGA and 2.183 ppm for the pMGGA (different cylinders were used). This is now stated in paragraph 4 of section 2.2.

The H_2O baseline has a huge dependence on CH_4. Nevertheless, this experiment was only done at one CH_4 concentration only. This may not be representative and add uncertainty to the retrieval of the "real-world" measurements.

The concentration of methane in the gas cylinder was 2.2 ppm, which is similar to the current atmospheric background of approximately 1.9 ppm (depending on the time of year). Thus the water correction was performed in the most real-word conditions possible, to minimise uncertainty.

Furthermore, pressure broadening line shape changes and dilution affect the instrument methane gain factor, with increasing water mole fraction. There is no physical (instrumental) reason for these effects to manifest themselves as an effect on the instrumental methane offset. Therefore, a water correction derived at one methane mole fraction should be equally valid at other methane mole fractions.

Page 4 – Line 150

The difference between wet and dry corresponds to humidity, thus typically 1-3 %. This is not "almost equal". If it has little impact on the result, then this is because the retrieved apparent H_2O does not very strongly depend on CH_4 under the conditions used to test the simplification.

We accept that the phrase “almost equal to” may be misleading. We have therefore rephrased this as “close to”. We have also now clarified that the water correction has little overall impact on dry mole fractions in the penultimate paragraph of section 2.2, in our simplified test, where it is referred to as a “simple example”.

Page 5 – Line 156-157

The uncertainty of water measurements cannot be derived using "the water correction residual (R) from Eq. (2)". The calculated value is not the uncertainty of water measurements, nor the uncertainty after water correction. It is the standard deviation of the mean of the residuals. This parameter carries information about the quality of the fit of the polynomial, but no information about the quality of the water correction for a specific measurement at a given CH4 and HO mole fraction.

The reviewer correctly points out that we do not present an uncertainty in the water correction factor, but instead an uncertainty in our fit. This was also addressed in our response above, where this was raised. We now make this clear in the final paragraph of section 2.2. Furthermore, the water “uncertainty factor” has been renamed the water “fitting uncertainty factor” to reflect this.

Page 5 – Line 160

"For example our correction was used to increase 160 the MGGA measurement accuracy of $[X]_0$ (at 2 ppm) by +0.27%, at a humidity of 0.001 mol water mol^{-1}, and by +1.8%, at a humidity of 0.01 mol water mol^{-1}". This is a meaningless statement. It does not consider that any correct measurement includes the correction of the dilution effect. You compare apples with rotten pears.

This sentence may have been poorly written and misleading. The purpose of this sentence is to convey that we can improve the measured mole fraction, *i.e.* it more closely agrees with the true dry mole fraction. We have replaced the phrase “increase the MGGA measurement accuracy of $[X]_0$” with “increase $[X]_0$”. This reduces ambiguity, to make it clear what the correction achieves.

<u>*Calibration*</u>

Page 5 – General Comments on Section/Calibration

Useful:

The alternating measurement of two standards. This gives the repeatability of measurements under laboratory conditions. A simple average and standard deviation would characterize this. Alternatively (x - average x) vs. N (in bins) as a figure would illustrate that the values show a normal distribution.

Not useful:

The concept of uncertainty in C and G because it neglects the fact that the linearity in response is most frequently the key factor determining the uncertainty of the measurement.

Needed:

Determination of linearity.

The reviewer raises a valuable point regarding quantifying uncertainty due to non-linearity. During our main calibrations, only two standards were used (2 ppm and 5 ppm), as the World Meteorological Organisation Greenhouse Gas Calibration Scale does not extend higher than 5 ppm and reliable alternative standards cannot readily be obtained. Therefore, we have now tested the linearity within the calibration range, by sampling 5 gases with the MGGA. This showed that the residual uncertainty between measured and certified $[X]$ was on average 2.3 ppb. As we no longer have access to the pMGGA, we assumed similar linearity in both instruments as they use identical spectroscopic techniques. This test is described in section S4 of the supplement. The non-linearity uncertainty has been incorporated into the overall enhancement uncertainty.

Page 5 – Line 168-169

This does not make sense. Water vapor in reference gas cylinders should be very low if filled dry (which I assume was done here). Even if filled wet, then 60 bar/20°C should be significantly below 500 ppm. Thus, H_2O from a pressurized bottle (if treated correctly) is so low that it has very little impact on the measurements. An additional trap (as used here) may well lead to water contamination if one is not very careful.

We used a water trap as these cylinders were synthetic blends, prepared to include ambient levels of argon to provide an appropriate air matrix. This may cause a small amount of water to be present in the cylinders.

Page 6 – Line 195-196

C deviates by about 12 ppb from 0. Whether that is good or bad (and under what conditions) depends on the application. However, to say that this is "almost equal to 0" for an analyzer with an Allan Minimum below 1 ppb is bizarre.

This is a valid point. 12 ppb is indeed much larger than a 3 ppb Allan deviation (at 10 Hz) for the MGGA. However when compared to the atmospheric methane background of approximately 1800 ppb, 12 ppb is very small. The sentence now includes this qualification, to clarify our point of view.

Page 6 – Line 195

v is ca. 0.98 at 1 % H2O. This is not "almost equal to 1". Furthermore, v in S6 contains all water-related corrections, including normalization to dry conditions. Stating that this is "almost equal to 1" does not make sense, because the correction to dry conditions should be done in any case.

We recognise the reviewer's point. We have therefore removed any reference to *v* from this sentence.

Additional remarks

"A lightweight wind sensor (FT205EV, FT Technologies Limited) was mounted on-board UAV1, on a carbon fibre pole 305 mm above the plane of the propellers. It recorded wind speed and direction at 4 Hz. This data was used to model the change in wind speed with height above ground level (z)."

- Was this sensor validated?

The wind sensor was used to derive a wind-height profile, for cross comparison against stationary ground-based wind measurements. The key need here was a linear response to wind speed changes. This was tested by comparing wind speed at 3.3 m from the UAV sensor to a stationary 3.3 m anemometer. A linear fit intersecting the origin agreed with all points, within uncertainty, implying linear response of the UAV wind sensor. These results are now presented in section S7 of the supplement.

- Are the measurements independent of the propellers? How was this validated?

The effect of air disturbance on the measured wind field due to rotating propellers was minimised as the wind sensor was positioned 0.3 m above the plane of the propellers. Unfortunately, we did not have required facilities (such as a wind tunnel) to fully test that this was the case. However each UAV transect was carefully planned to minimise this effect by rotating the UAV orientation at each turning point. This would effectively cancel out any potential residual net wind vector distorting the wind field due to the rotation of the propellers. This is now stated in section S7 of the supplement.

- How reliable were the direction and speed when the drone was moving? How was this determined?

The reviewer is correct to suggest that the UAV wind measurement may have been affected due to the movement of the UAV. This impacts the measured wind vector component parallel to the direction of movement of the UAV. Therefore only the wind component perpendicular to the orientation of the UAV sampling plane was used. The motion of the UAV has no impact on the wind component perpendicular to the plane of movement of the UAV. It is the perpendicular wind component that is used to construct the wind profile used to derive flux. This rationale is now explicitly stated in paragraph 1 of section S7 in the supplement.

One analyzer was using an inlet just underneath the drone. The corresponding sample is representative of the air above the drone because of downwash. Was this considered? How?

This is an astute comment. The reviewer is right to point out that mixing due to downwash could occur and that what is sampled is air from a subtly higher elevation to the position of the inlet. The inlet was 31 cm below the plane of the propellers. We have not (and cannot usefully) account for this small height difference in the calculation of flux as the air sampled is likely mixed in a thin (31 cm) layer between the plane of the rotors and the plane of sampling.

Flux of one drone --> the other just sampled further away.

We do not understand what the comment above is suggesting.

The Allan Variance plot should be in main section of the paper. Its v-shape is relevant for interpreting the measurements during flight (i.e. 10 – 30 minutes, typically).

Both plots have now been moved to the main manuscript, as suggested.