RC 1:

1) This is a generally very well written and very solid documentation of the setup and calibration procedure for flux measurements from a unique low flying aircraft. A necessary basis for any follow-on studies using observations from this particular system, but as such there is little novelty in concepts or methods or data. In highlights + associated comments in the attached document I suggest a number of mere technical/editorial improvements. In addition, a few issues require a bit more attention, some preferably in the form of an explicit discussion.

Thanks a lot for reviewing this manuscript and providing valuable comments/ suggestions for our manuscript. We have responded to the technical/ editorial improvements in the attached .pdf document. Additionally, we have responded to the other issues that require more attention below.

2) The integration of the Picarro trace gas analyser is poorly documented. E.g. there must a significantly long inlet tube between instrument in cabin and inlet (I assume) in wing pod. What flow rate/pump was used? How were unavoidable significant time delays dealt with (documented a bit late in the text)? Why are lag times of H2O treated differently from CO2 and CH4 if all 3 signals come from the same instrument/setup?

Thanks for the suggestions. We have incorporated more information about the integration of the Picarro gas analyzer (tube length, response time of the gas caused by the length of the tube) in the manuscript as follows:

"The distance between the inlet of the tube and the gas analyzer and the five hole probe was small (< 0.5 m). The tube was ca 6 m long, had a flow rate of ca. 5.8 sL/minute and an inner diameter of ca. 0.04 m. Based on these characteristics, the transport time of the gas between the inlet tube and the G2311-f gas analyzer was ca. 0.8 seconds"

We have also added a table with an explanation of the lags that were calculated between the gas measurements and the vertical wind w' (new Table 4, see below). Additionally, we have written some text in the section "Temporal and Spatial Alignment Wingpod Data" to redirect to section 2.5 (and Figure 2) in the manuscript where this cross-correlation was performed - to make it easier to find for the reader. This extra sentence reads now as follows:

"Temporal alignment of the wingpod data and the Picarro data is performed at a later stage in the data processing using Eddy4R (see Figure 2; for details see section 2.5)."

The lag times of H_2O are treated differently from CO_2 and CH_4, as we observed that the water vapor has more delay, when looking at the cross-correlations – H_2O can adhere to the tube surface and therefore will have more delay compared to the other measured gasses. As the adherence depends on the humidity within the tube, we assume that changes in lags are possible between the legs, which is why we use leg-defined lags to correct for time offsets between H_20 and w.

3) A bit more discussion and/or analysis of the differences/pros/cons of wavelet vs reynolds flux calculation would be welcome. Same for use of 2km integration windows with only 200m stepsize, which artificially reduces random errors and increases (sometime unrealistically) autocorrelations between data points. Sharp transitions in surface fluxes (e.g. lake) might be unnecessary convoluted.

The use of 2 km integration windows with 200 m step size may indeed introduce some autocorrelation due to 90% overlapping samples, which could artificially reduce ensemble random errors:

$$Error = \frac{\sigma}{\sqrt{N}}$$

where:

- *σ is the standard deviation of the random error in individual samples.*
- *N is the number of independent samples.*

We have accounted for this by recognizing that overlapping samples are autocorrelated, and the effective sample size N_{eff} is reduced accordingly, following the formula:

$$N_{eff} = \frac{N}{1 + 2\sum_{k=1}^{K} \rho(k)}$$

Here:

- *$\rho(k)$ is the autocorrelation function of the sample with lag k.*
- *K is the maximum lag where significant autocorrelation exists.*

Using N_{eff} in place of N corrects the ensemble random error to reflect the increased autocorrelation between samples. The individual overlapping samples are still valuable because they preserve high spatial resolution, which is critical for capturing sharp transitions in fluxes (e.g., from land to lake) and for reducing random noise in turbulent atmospheric conditions. Additionally, wavelet-based flux calculation benefits from this approach, as it allows for multi-scale analysis and better characterization of spatial heterogeneity, compared to traditional Reynolds-averaging methods that smooth out small-scale variations.

Detailed Rationale for the Utility of Overlapping Samples

1. ***Reducing Random Noise**: Even with overlapping measurements, retaining these individual samples is useful because averaging over autocorrelated samples still helps in reducing random noise. This is especially useful in turbulent atmospheric measurements, where sharp transitions in surface fluxes (e.g., from a lake to land) can cause significant noise. Reducing random noise improves the overall signal-to-noise ratio.*
2. ***Preserving Spatial Resolution**: Aggregating all samples into larger blocks (e.g., non-overlapping 2 km windows) would reduce spatial resolution and may obscure important features of the surface fluxes. For instance, in heterogeneous landscapes, preserving the high resolution (e.g., 200 m step size) helps capture localized changes in fluxes that would otherwise be lost.*
3. ***Wavelet Analysis vs Reynolds-Averaging**: In wavelet-based flux calculations, the overlapping samples allow for a finer, multi-scale analysis, where the goal is not just to compute a single flux value but to understand how fluxes vary across spatial scales. This contrasts with Reynolds-averaging, which smooths out small-scale variations. Retaining*

overlapping samples in wavelet-based analyses ensures that transitions in fluxes (e.g., from a lake to land) are resolved.

4) Since your turbulence probe is mounted in a wingpod, ie off the symmetry-axis of the fuselage, may be you can discuss a bit the implications this has (implicitly or explicitly) for the calibration procedures/parameterisations for the true wind vector.

Thanks for discussion this point. Indeed, similar to Mallaun et al. (2015) and Neininger et al. (2001), the wind measurements are taken off the symmetry axis of the fuselage.

This would, however only be problematic if the heading of the aircraft would be very different from the horizontal orientation of the aircraft – if the angle between the aircraft's forward direction and the direction in which the aircraft is moving would be very large (see Figure RC1.1).

However, the actual sideslip angles are much smaller than the angle that would be required to create the situation that the probe would be shaded by the fuselage (Figure RC1.1) For example, during a measurement flight with start and landing on the 21st of August, 2019 the beta angle was max. 10° (more than a factor 5 smaller).

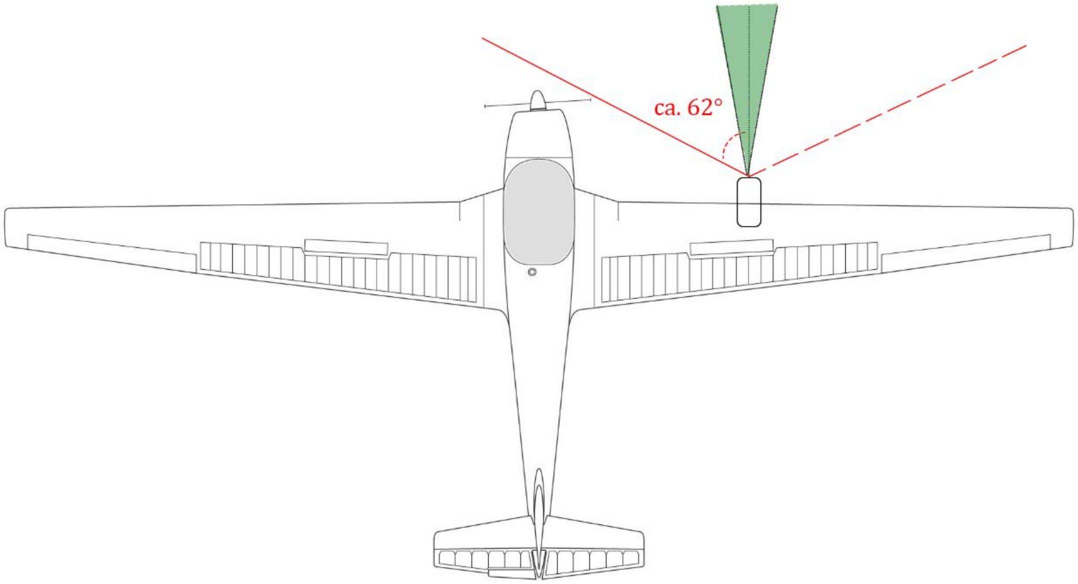
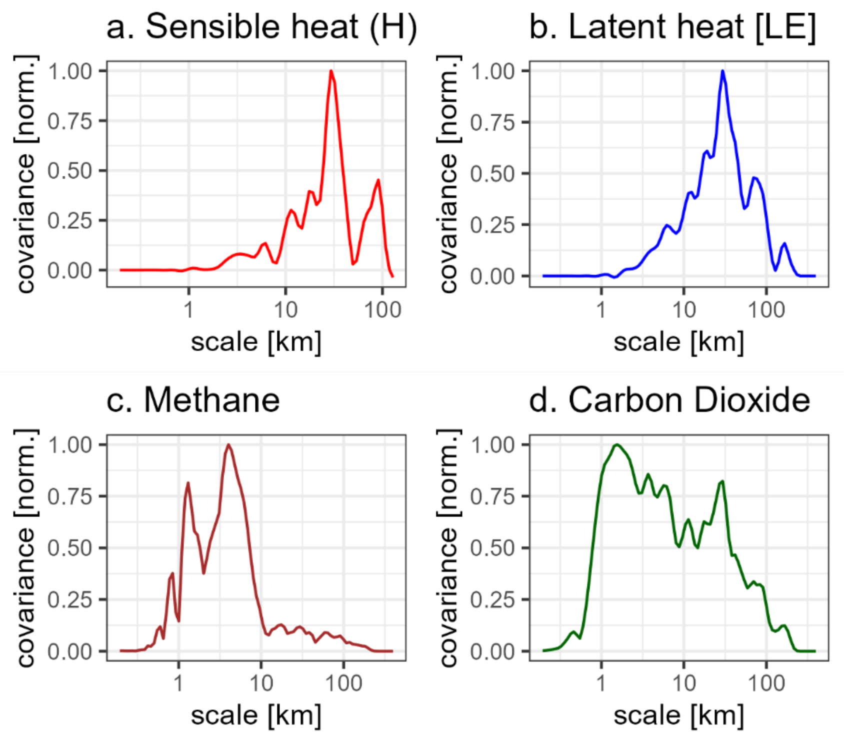


Figure RC1.1: Position of the wingpod and the sideslip angle that would be required to cause conditions in which the probe would be shaded by the fuselage, green area indicates obtained angles from measurement flight on the 21st of August 2019 (as an example).

5) Unlike your near-perfect wind spectra, for the other signals they are (much) less perfect. Can you *show* that this does not affect calculated fluxes as you claim, e.g. using cospectra? How might that change if you fly lower? Somewhat related, do you do any high/low freq. corrections?

Generally, a higher level of random noise does not affect the covariance, and thus the fluxes, as the noise is not correlated with the vertical motion. This was already shown in detail by Hartmann et al. (2018).



a. Sensible heat (H)

b. Latent heat [LE]

c. Methane

d. Carbon Dioxide

Figure RC1.2: Averaged cospectra from flight legs 29th of August 2018 and 21st of August 2019 (in total 11 flight legs were available, for sensible heat only 6 legs were available as no fast Temperature sensor was installed in 2018).

The cospectra that are shown in *Figure RC1.2* clearly indicate that the noise signal that was visible in the spectral plots of the temperature, and Picarro data (Figure 10b of the manuscript) is not correlated with the vertical wind and does not cause an artificial flux signal.

References

Hartmann, J., Gehrmann, M., Kohnert, K., Metzger, S., and Sachs, T.: New calibration procedures for airborne turbulence measurements and accuracy of the methane fluxes during the AirMeth campaigns, Atmos. Meas. Tech., 11, 4567-4581, 10.5194/amt-11-4567-2018, 2018.

Mallaun, C., Giez, A., and Baumann, R.: Calibration of 3-D wind measurements on a single-engine research aircraft, Atmos. Meas. Tech., 8, 3177-3196, 10.5194/amt-8-3177-2015, 2015.

Neininger, B., Fuchs, W., Baeumle, M., Volz-Thomas, A., Prévôt, A., and Dommen, J.: A small aircraft for more than just ozone: Metair's' Dimona'after ten years of evolving development, 11th Symposium on Meteorological Observations and Instrumentation, Amer. Met. Soc., Boston, 123-128,