First of all, we thank reviewer 2 for his effort in carefully reviewing our manuscript and his constructive comments.

Point-by-point answers to the comments of reviewer 2

Specific comments

Reviewer 2: *Generally, all figure labels are too small. Figures should be redrawn before final publication. Some figures (e.g. Fig.4) would benefit from rescaling or splitting.*
Authors: Done. We enlarged the font size of almost the entire text shown in the figures. The abscissa in Fig. 3 and 4 has been re-scaled.

Reviewer 2: *The anemometer calibration (Section 4) looks reasonable. With a small, slow aircraft that even can hold a fixed position, a comparison with a proven reference at a mast is possible, a big advantage to faster, manned planes. For future applications that characteristic should be exploited more in calibration and measurement pattern design.*
Authors: Our comparison of the calibrated wind measurements with mast-based reference measurements presented in Sec. 5 (Validation using ICOS measurements) indeed shows that the UAS derived wind agrees well with reference data. However, given that the uncertainty of emission estimates are often driven by the uncertainty of the wind measurements, we intend to continue our efforts to monitor and ideally also to improve the quality of our calibration in the future, if opportunities arise.

Reviewer 2: *Section 5, validation with ICOS: Why do you use level 1 data for comparison ? At low level there is more horizontal variation on a small scale in the flow and in the scalar fields than there is at higher level,where a horizontal distance to the reference has less effect.*
Authors: We intended to capture as much vertical atmospheric variability (potential gradients in temperature, wind, humidity, and CO_2) as possible. However, we agree that the deviations between our measurements and those performed at the mast may be larger for the lowermost measurement level compared to the higher levels. Most likely, this is the reason for the reduced agreement between our and the ICOS wind measurements in sequence S12 (see Fig. 6). We discuss this point in Sec. 5 (Validation using ICOS measurements) as follows: "The relative large scatter of the difference to ICOS is mainly driven by a poor agreement in S12. As visible in Fig. 5 (a and b), the tower is located close to a small piece of forest while the flight track is above a relatively free field which can lead to significant differences of the measurements at 32m height." This is also the reason, why we excluded S12 from the computation of the scatter of the wind components ("The average of the scatter of the north and east component is $0.62\,\mathrm{m\,s^{-1}}$ (excluding S12 ...).").

Reviewer 2: *p14, l240: The discrepancy in the wind data between UAV and mast around 700s (S12) looks like a problem in the directional data of the UAV. Can you comment on that ?*
Authors: As discussed above, we expect, that the poor agreement of the wind measurements in S12 (at 32m) comes from the influence of the surface.

Reviewer 2: *Has S12 been flown at 32m ? The height of the legs should be given.*
Authors: Yes, S12 is at 32m. We mention this in L241. Additionally, we describe the flight pattern in Sec. 5 (Validation using ICOS measurements): "During the first flight, the UAV twice ascended to 187m and descended level by level to 32m ... The flight pattern is shown in detail in Fig. A1.". The referenced figure A1 shows that S12 corresponds to 32m.

Reviewer 2: *You could even try to get the mean wind by having the UAV drift with it by keeping only a constant height and horizontal leveling active and deactivate position holding. Then, the drift speed should be the mean wind speed, like with a radiosonde. A tilt error should then be checked by repeating with the UAV turned by 180deg around the vertical axis.*
Authors: We agree, this could indeed be an interesting additional analysis for future calibration and validation flights.

Minor specific point, typos and such

Reviewer 2: *p3, l85: ... via an RS-232, an RS-485 ...*
Authors: Done.

Reviewer 2: *p4, caption Fig.1: the labels in the left photography of are hard to read*
Authors: We experimented with different colors (red/white), with or without outlines, bold or normal fonts but did not find a better combination than the one currently selected.

Reviewer 2: *p6, Fig.2: labels far too small, caption too brief*
Authors: Done. The caption now reads: " **a)** CO_2 concentration measured with the Vaisala GMP343 CO_2 sonde with and without linear correction (Eq. 1) as well as highly accurate reference measurements performed with an ABB LGR-ICOS ultra-portable greenhouse gas analyzer. **b)** Difference between the Vaisala GMP343 CO_2 (with and without linear correction) and the reference measurements. Pale colors represent instantaneous differences and intense colors 1h running averages. **c)** Temperature measured with the Vaisala GMP343. **d**) Relative humidity measured with a Sensirion SHT31-DIS sensor. **e)** Pressure measured with a Bosch BMP388 sensor. "

Reviewer 2: *p9, Fig.3: labels too small*
Authors: Done.

Reviewer 2: *p10, Fig.4: labels too small, abscissa could be reduced to the range of 350s–1200s, e,f,g: labelling/text mismatch.*
Authors: Done.

Reviewer 2: *p12, Fig.5: labels too small*
Authors: Done.

Reviewer 2: *p13, Fig.6: labels too small.*
Authors: Done.

Reviewer 2: *p15, l266: A ?possible? explanation*
Authors: In L269ff we discuss: "A potential explanation of the general overestimation could be the heating of the GMP343 CO_2 sonde, which is intended to reduce the possibility of condensation on the optical components but which could also slightly warm the temperature sensor."

Reviewer 2: *p15, l276: 40k west of Bremen, dito in captions Fig.7 and 8., p20,l382*
Authors: Done.

Reviewer 2: *p16, l305: while this method seems acceptable for a first estimate in using a new device, shouldn't the background concentration be determined on the upwind side of the emitter ?*
Authors: According to Gauss's theorem, one would ideally have continuous measurements all around and above the facility to exactly quantify the inflow of CO_2. Of course, this is not possible so that an as good as possible choice of the undisturbed background concentration has to be made. Upwind measurements would indeed minimize the influence of potential upwind sources. However, they would have the disadvantage, that they cannot be performed during the same flight because of the larger distance. Furthermore, the flight time in the background air would considerably be extended at the expense of the flight time within the plume. For these reasons, we considered nearly simultaneous measurements in the left and right neighborhood of the plume as better choice to estimate the undisturbed background concentration.

Please also note that similar approaches are often used in the literature (e.g., Carotenuto et al., 2018; Krings et al., 2018) and that Krings et al. (2018) discuss: "The single-screen approach was chosen for practical reasons, because flying around a source means spending most of the time in background concentrations ... Some circumferential tracks ... confirmed the background concentrations found on the edges of the single screens."

Reviewer 2: *p17, l308: With a linear interpolation you assume a*

horizontal gradient in the concentration, why ? Why dont't you take the average ? Especially as you assume no upwind sources anyway.
Authors: As one can see in Fig. 8, there is no significant variability in the background concentrations. Therefore, it would practically make no difference to use the average or the linear interpolation. However, we consider the linear interpolation a good choice because theoretically, the background may vary slightly along the flight path, e.g., because of upwind variations of natural fluxes. Additionally, small sensor drifts may exist, which would also be better accounted for by a linear interpolation.

Reviewer 2: *p19, l339: ... beyond the scope of this paper ...: Well, you could at least tell us the difference in the estimate between both flights.*
Authors: As shown in Fig. 7, we shifted the flight pattern of both flights against each other, so that the sampling becomes denser in the center. The downside of this strategy is, that the western pattern likely misses some parts of the plume in the east and the eastern pattern some parts in the west (see Fig. 7). Therefore, the expected values of the emissions derived from the individual flights are not identical to the expected value of the combined flight. For this reason, we decided to omit the discussion of the individual results in the paper and would prefer to keep it as is.

Nevertheless, out of curiosity, we computed the cross-sectional fluxes per flight and found $209 \pm 17\,\mathrm{ktCO_2\,yr^{-1}}$ for the eastern and $156 \pm 17\,\mathrm{ktCO_2\,yr^{-1}}$ for the western flight pattern. The uncertainty estimates have been computed in the same way as done for the combined cross-sectional flux. As discussed in the paper, this uncertainty estimate is incomplete. Foremost it does not consider turbulence which is one (or even the main) reason why it is no contradiction that both values differ by more than their uncertainties.

References

Carotenuto, F., Gualtieri, G., Miglietta, F., Riccio, A., Toscano, P., Wohlfahrt, G., and Gioli, B.: Industrial point source CO_2 emission strength estimation with aircraft measurements and dispersion modelling, Environmental monitoring and assessment, 190, 165, 2018.

Krings, T., Neininger, B., Gerilowski, K., Krautwurst, S., Buchwitz, M., Burrows, J. P., Lindemann, C., Ruhtz, T., Schüttemeyer, D., and Bovensmann, H.: Airborne remote sensing and in situ measurements of atmospheric CO_2 to quantify point source emissions, Atmospheric Measurement Techniques, 11, 721–739, https://doi.org/10.5194/amt-11-721-2018, URL `https://www.atmos-meas-tech.net/11/721/2018/`, 2018.