Reply to Anonymous Referee #1

We would like to thank Referee #1 for reviewing this manuscript. We found the comments and suggestions very helpful to improve this paper. Our responses to each comment are provided below in black font, and the Referee's comments are indicated in blue. Unless stated otherwise, all page and line numbers refer to the originally submitted manuscript.

This paper evaluated the contribution of fossil fuel CO2 (ffCO2) emissions in Zurich, Paris, and Munich using both the relaxed eddy accumulation (REA) method and the eddy covariance (EC) method. The contribution of winter ffCO2 was found to be about 80% in all three cities. For the summer, the measurement results in Munich demonstrated the major role of respiration, biofuels, and industrial processes that led to predominantly positive non-fossil CO2 (nfCO2) fluxes. The concentration of absolute CO2 and 14CO2 of the REA samples was used to estimate the impact of ffCO2 on local CO2 concentration, namely, excess concentrations. In the end, the authors stated that this combination of REA and EC method can be useful for quantifying ffCO2 and nfCO2 fluxes, but challenges existed due to the limits of analytical capabilities. The topic is important given the current strategy to reduce carbon emissions, especially fossil fuel consumption. The idea of combining REA and EC to separate ffCO2 and nfCO2 fluxes is novel and could significantly improve the understanding of urban carbon sources. However, the experimental setup is questionable and lacks important descriptions. For example, the instrument calibration should be included in the method section, and the application of long intake lines to deliver air samples for REA flask sampling and MGA measuring should be further evaluated. Below are more comments in detail.

We have included further information on the experimental setup as described below. We also refer to Kunz et al. (2025), where technical aspects of the REA system and the flask sampling procedure are described in detail.

1, Please double-check the references. Some references need to be updated, such as Hilland et al. (2025) and Kunz et al. (2025).

The references have been checked and updated according to the changes made during the review process of this manuscript.

2, Line 70-75. There are several timing issues that need to be clarified. Please describe the method applied to upsample 10 Hz MGA measurements to a 20 Hz dataset. Clock drift is a common issue in certain high-frequency instruments, and it appears that the MGA experienced this problem. Therefore, could you provide more information that shows your linear interpolation works great to solve this problem? How and how often did you check and sync the timestamps on the IRGASON and the MGA instrument?

Based on the questions of the Referee, we believe that some aspects of the EC data processing did not become clear from our original text. As both IRGASON and MGA^7 measured CO_2, we were able to synchronize the measurements directly on the high-frequency CO_2 time series. Consequently, clock drift and the interpolation of time lags were less relevant. To clarify, we describe the individual steps of the upsampling and handling of of time lags in more detail:

1. **Upsampling the MGA[7] measurements to 20 Hz**
   We upsampled the 10 Hz gas measurements of the MGA[7] to 20 Hz, i.e., to the time stamps of the 20 Hz IRGASON measurements, using a nearest-neighbor approach with a 50 ms search window. This essentially duplicated the MGA[7] CO_2 measurements.
2. **Determining the time lag between MGA[7] and IRGASON measurements**
   There are two reasons why the upsampled MGA[7] and the 20 Hz IRGASON CO_2 measurements may be shifted in time: clock drift between the two instruments (the MGA[7] was synced to an NTP server once per week, the IRGASON logger only periodically) and the travel time of the sample gas to the MGA[7] instrument (several seconds). However, both issues could be taken into account by synchronizing the data sets directly on the 20 Hz CO_2 concentrations. For this purpose, we determined the time lag of maximum correlation between the IRGASON and the MGA[7] high-frequency CO_2 time series.
   For individual 30 min averaging periods in which the correlation between the IRGASON and the MGA[7] CO_2 signals was poor (correlation coefficient <0.5, see Comment 3), due to low IRGASON signal strength, the time lag was linearly interpolated between adjacent blocks, assuming that time lags due to travel time and clock drift were much larger than the variance in “true” time lag caused by pressure changes or the pump. This mainly affected periods for which weather conditions were generally not suited for EC measurements.
3. **Synchronizing the MGA[7] measurements with the IRGASON measurements**
   We shifted the MGA[7] measurements according to the time lags determined in step 2.
4. **Calculating fluxes**
   Finally, fluxes were calculated with EddyPro using covariance maximization of the CO_2 with the vertical wind with a search window of ±0.5 s (increased to +/- 1 s in Munich). Due to the preceding manual synchronization in steps 1 to 3 based on the correlation of the high-frequency CO_2 concentrationss, this overwhelmingly produced a time lag of 0.0 s.

We specified and expanded the description in Line 70 ff (Line 79 in the resubmitted manuscript) and added further information, e.g., the median time lags (see also Comment 12) in an additional section of the Appendix (Appendix D in the resubmitted manuscript).

3, Line 74. The correlation coefficient threshold (0.5) for determining poor correlation between the CO2 time series seems too low. Were the IRGASON and MGA instruments calibrated using a standard CO2 gas during the measurement? If calibration was conducted periodically, the CO2 measured by IRGASON and MGA should be in much better agreement.

The correlation coefficient threshold of 0.5 was only set to reject obviously erroneous time lags obtained when determining the maximum correlation between the 20 Hz CO_2 time series (see Comment 2 step 2). This mainly affected periods with low IRGASON signal strength, during which weather conditions were generally poor for calculating EC fluxes (quality control flag 2 according to Mauder and Foken, 2004). Both instruments were calibrated prior to the measurement campaigns. Regular on-site calibration was not directly possible due to the difficulty and cost of accessing and removing/reinstalling the IRGASON, the associated loss of data, and because the instruments were only intended for flux calculations with a 30 min averaging period, for which instrument drift occurring over much longer timescales can be considered negligible. For quality control purposes, the instruments were compared with the REA flask samples (see the next answer below). Even without regular calibration, the average correlation of the time lag corrected signals was 0.83 in Zurich, 0.76 in Paris, and 0.83 in Munich, with higher values for high and medium quality data.

We added the relevant information to an additional section in the Appendix (Appendix D in the resubmitted manuscript) and refer to it in Line 75.

4, Line 93-94. Without the information about instrument calibration, the better agreement between MGA and flask measurements does not necessarily indicate that their CO2 concentrations were accurate. Could you explain why the CO2 concentration measured by IRGASON did not agree with the data measured by the other two instruments?

As suspected by the referee, a comparison of absolute concentration measurements of IRGASON and MGA[7] with the REA flask samples was not possible due to a sparse calibration of both MGA[7] and IRGASON (see Comment 3). Instead, we compared the measured CO_2 concentration differences between updraft and downdraft flask samples (ΔCO_2) with the concentration differences calculated from the 20 Hz CO_2 time series, considering the recorded individual opening times of the updraft and the downdraft sampling valves. This comparison is more meaningful for the REA methodology than the absolute concentration accuracy.

We specified this in Line 93-94 (Line 102 ff of the resubmitted manuscript) and adjusted the reference to Appendix D, where we describe our observations in more detail. For a detailed description of calculating the concentration differences based on EC measurements, we refer to Kunz et al. (2025).

Regarding the better agreement between measured flask concentration differences with the MGA[7]-based ΔCO_2 estimates than with the IRGASON-based ΔCO_2 estimates, it is important to note that the flask concentrations measured by the ICOS Flask and Calibration Laboratory in Jena are of much higher quality (0.02 ppm uncertainty with respect to the WMO calibration scale) than the IRGASON and MGA[7] CO_2 measurements (precisions of about 0.15 ppm and 0.1 ppm stated by the manufacturers). As an open-path instrument, the IRGASON is much more susceptible to weather, obstructions in the path, etc.. In Zurich, part of the differences between the IRGASON and the other two instruments could be attributed to the fact that the IRGASON CO_2 dry molar fractions were derived from a CO_2 density output that did not properly account for high-frequency fluctuations in air temperature in the sensing path, because the ambient temperature measured by an EC100 slow-response temperature probe was used in the conversion of the absorption measurements to CO_2 density (Kunz et al., 2025). Since 13 April 2024 (end of Paris measurements), an updated logger program recorded the CO_2 measurements using a fast-response temperature of the ultrasonic anemometer. Overall, the IRGASON and MGA[7] fluxes agreed well, with correlations of the qc0 and qc1 fluxes > 0.89 and RMSE < 2.45 μmol m^{-2} s^{-1} in all three cities (see Comment 10, added in the Appendix).

We expanded the discussion in Appendix D and noted in Line 92 that any additional uncertainties arising from the instrument or the processing options used are not considered in this manuscript.

5, Line 109-110. I can see that the deadband is quite important parameter in the REA method. Since the dynamic deadbank was more suitable, it would be great for other researchers interested in this method if you could provide more details on this method, such as how to determine the scaling factor δ.

We agree that the description of the REA method omits important details. However, we refer the interested readers to Kunz et al. (2025), in which the REA system and the various technical parameters are described in detail. We added a corresponding note in Line 110.

To answer the question about the scaling factor δ: The scaling factor δ was chosen as the maximum value such that the optimum amount of air required for laboratory analysis (v_{sample} = 75 l) would likely be collected within the intended measurement period (t_{tot} = 60 min). The absolute time needed to collect sufficient air (t_{sample}) is determined by the flow rates into the reservoirs q, which are controlled by two mass flow controllers and limited by the flow rates through the pumps:

$$t_{sample} = v_{sample} / q \, .$$

The fraction of time during which air is expected to be collected was estimated from test measurements and/or simulations for different deadband widths. The results agreed well with the theoretical values assuming a Gaussian distribution. For example, with δ = 1, it can be expected that both up- and downdraft conditions will be met for approximately 34.1 % of the time, i.e., each sampling valve will be open for about 20.5 min. However, it should be noted that with this statistical approach, there is still a possibility that not sufficient air will be collected during the measurement period in one or both reservoirs. In this case, the sample is rejected and not transferred into the glass flasks. Table 1 shows the required sampling time durations depending on the flow rates and the respective scaling factors used during the REA campaigns in Zurich, Paris, and Munich.

Table 1. Overview of the sampling time durations required to collect sufficient sample air and the corresponding maximum deadband widths used in each city. q is the flow rate into the updraft and downdraft reservoirs; t_{sample} is the amount of time during which up- and downdraft valve must be open to collect 75 l of sample air; t_{tot} is the chosen total measurement period (60 min), and δ is the chosen deadband width. For hyperbolic REA (HREA), the value refers to the hole size *H*.

	q [l min^{-1}]	t_{sample} [min]	t_{sample} / t_{tot} [-]	δ [-]
Zurich	4.7	14.9	0.25	0.7 (REA)
Paris	4.7 6.5	14.9 10.8	0.25 0.18	0.7 (REA) 0.9 (REA)
Munich	9.0	7.8	0.13	1.1 (REA) 0.8 (HREA)

6, Line 114-116. It looks like the REA flask samples were analyzed by different laboratories. Please describe how consistency in flask sample results between laboratories was ensured?

The REA flask samples were analyzed for CO_2 in the ICOS Flask and Calibration Laboratory in Jena, and for $^{14}CO_2$ in the ICOS Central Radiocarbon Laboratory in Heidelberg - the two Central Analytical Laboratories (CAL) of the Integrated Carbon Observation System (ICOS), as described in Levin et al. (2020), for example. As they measure different gases, their procedures are not designed to be consistent with each other, but rather to optimize their analysis of the gas in question. For CO_2, the measurements meet the compatibility goal of the World Meteorological Organization of 0.1 ppm (WMO, 2020). Further information on the two laboratories, including quality control reports, can be found at ICOS - CAL (2026).

We clarified this in Lines 114 ff.

7, Line 120. Eventually, the β was determined based on the co-located high-frequency EC data; The statement in Lines 120–121 may cause confusion. I recommend either removing this sentence or comparing this 0.627 with the β calculated based on EC data, and showing the advantage of your method.

Agreed. We removed the sentence in Line 120.

8, Line 130. Please elaborate on the limitations and uncertainties introduced by assuming scalar similarity between CO2 and 14CO2 and using their ratio to derive ffCO2 fluxes?

Scalar similarity between the scalar of interest (^{14}C-based $ffCO_2$) and the proxy used to determine the beta coefficient (CO_2) is an essential requirement for accurate calculation of fluxes from REA measurements. Evaluation of scalar similarity, and thus proper quantification of the resulting uncertainties from this assumption, requires comparison of both scalar time series and the shape of their frequency distributions or spectra (e.g., Ruppert et al., 2006). Unfortunately, for $^{14}CO_2$ and CO_2, this comparison is hampered by the lack of fast response gas analyzers for $^{14}CO_2$. However, based on analyses of scalar similarity between CO_2, water vapor, and temperature (e.g., Pearson et al., 1998; Ruppert et al., 2006), as well as between CO_2 and the stable isotopologues $^{13}CO_2$ and $C^{18}OO$ (Ruppert et al., 2008), and based on our $^{14}CO_2$ flask measurements (see below), we argue that the assumption of scalar similarity between CO_2 and $^{14}CO_2$ is justified, and that uncertainties in the calculated $ffCO_2$ fluxes are small compared to the measurement uncertainties.

We added a note in Line 138 stating that uncertainties due to scalar similarity or EC processing options, for example, were considered less relevant and not taken into account. As no other sources of uncertainty were named, we removed the explicit mention of the assumption of scalar similarity in Line 570. For completeness, we provide more details on the assumption of scalar similarity below.

Based on simulations, Ruppert et al. (2006) concluded that scalar similarity is usually good for short-duration events (< 60 s). Differences in events over longer time scales have a much greater impact on the (lack of) scalar similarity, presumably due to changes in the source/sink strength. While the short time scales cannot be analyzed for $^{14}CO_2$ or any other scalar of interest for REA measurements, slow response gas analyzers or flask measurements can provide information on the scalar similarity over the longer, more relevant time scales. Linear relationship between bulk CO_2 and $^{13}CO_2$ of REA samples reported by Bowling et al. (1999), Bowling et al. (2001), and Ruppert et al. (2008), indicate good scalar similarity between the isotopologues. Figure R1 shows a similar comparison and similarly good correlation between the CO_2 and $^{14}CO_2$ (in Δ-notation) of our REA flask samples. While the slope of the linear regression is less important, it is evident that some measurements, particularly in Munich, have a systematically different composition. Many of these were low-turbulence or storage-affected measurements. This could indicate the accumulation of respiratory CO_2 during the night, which is less depleted in $^{14}CO_2$ than during the day. This is consistent with our discussion of low-turbulence morning measurements in the manuscript. Ruppert et al. (2008) also observed differences in the isotopic composition of $^{13}CO_2$ REA samples collected during the morning transition period above different types of vegetation. They explained these observations by air above the canopy being strongly depleted in $^{13}CO_2$, resulting from respiratory built up during the night, in combination with high photosynthetic activity in the top canopy in the morning.

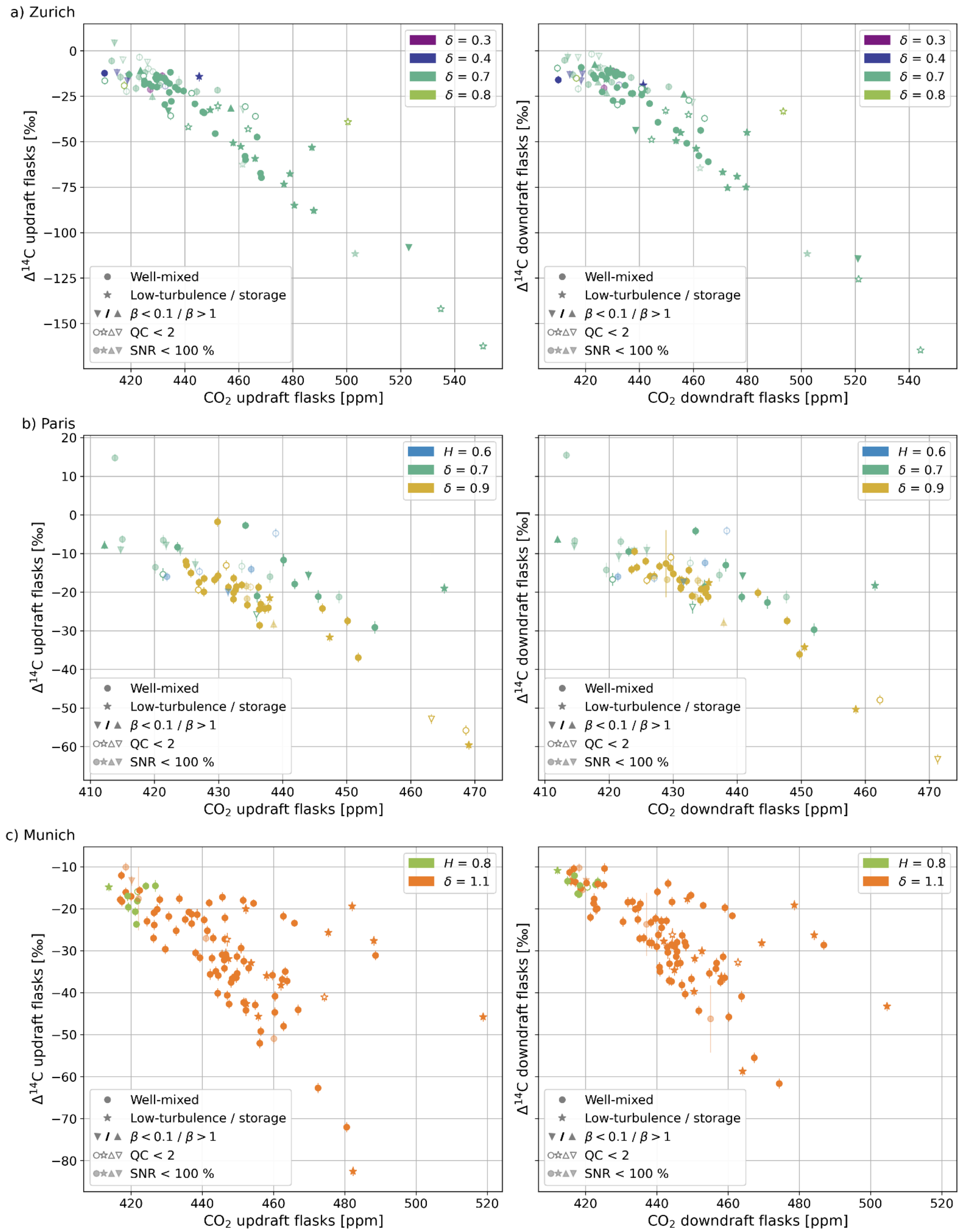

Figure R1. $\Delta^{14}CO_2$ as a function of the absolute CO_2 concentrations of the updraft (left) and downdraft (right) flask samples.

Based on these results, and given the fact that all CO_2 isotope turbulent exchange is part of the bulk CO_2 turbulent exchange, it is reasonable to assume scalar similarity between $^{14}CO_2$ and CO_2. However, given the heterogeneous distribution and temporal variability of $^{14}CO_2$ and CO_2 sources and sinks in the urban environment, some differences in the scalar exchange and the scalar similarity are expected, particularly when non-fossil CO_2 fluxes are high, i.e., during time periods of intense biospheric activity. Less scalar similarity would introduce some error in the REA flux results. However, the relative flux error resulting from a lack of scalar similarity is assumed to be small compared to our measurement uncertainties. Ruppert et al. (2006) found that, CO_2, H_2O, and T, with a deadband of 0.6, the relative flux error due to a lack of scalar similarity is < 10 %, without systematic under- or overestimation of the flux. Similar results were obtained by Oncley et al. (1993), Foken et al. (1995), and Ammann and Meixner (2002). For hyperbolic REA, relative flux errors are expected to be larger (up to 40 %) during time periods of low scalar correlation, with a tendency for underestimation of the flux (Ruppert et al., 2006). Nevertheless, this is still substantially smaller than the mean measurement uncertainty of 70 % observed in the few (N=7) HREA measurements taken in Munich.

In summary, the good correlation between CO_2 and $^{14}CO_2$ flask measurements, and the similar results obtained from more extensive studies on $^{13}CO_2$, indicate good scalar similarity between CO_2 and $^{14}CO_2$. In the future, scalar similarity between CO_2 and $^{14}CO_2$ should be investigated further using slow response gas analyzers for $^{14}CO_2$ (once these are available with sufficient temporal resolution and accuracy, see e.g., Lehmuskoski et al., 2021; McCartt and Jiang, 2022). Even for periods with low scalar similarity, flux errors are typically assumed to be < 10 %, which is small compared to our mean measurement uncertainties.

9, Line 150-151. Since the storage flux correction is not feasible for REA measurement. What did the storage term look like compared with your EC results? Additional discussion may be helpful to demonstrate the minimal impact of storage flux correction.

For the well-mixed REA measurements, the mean (median) ratios between the storage flux estimates and the total CO_2 fluxes were -3.7% (0.4 %) in Zurich, 4.7% (-0.1%) in Paris, and -1.6% (-0.2%) in Munich. The impact of storage flux correction is therefore considered negligible compared to other errors (particularly the measurement error of $^{14}CO_2$ in the laboratory).

We added a note in the discussion of the results in Sect. 4.1 (Line 359 of the resubmitted manuscript).

10, Line 214-215. I can see that the inlet of the MGA instrument was located close to the flask sample inlet, so most likely, the intake line of the MGA was the same setup as the flask sampling. If so, please clarify in the manuscript. My question is, do both intake lines always share the same flow rate? Given the 100 m intake lines in Munich (assuming the same intake line setup), pressure drop could be an issue for the MGA instrument. Is there evidence that shows that the MGA instrument provides solid readings with low inlet pressure? Given the fact that MGA results were in better agreement with flask results than IRGASON, the quality check of the MGA instrument is important.

As suspected by the referee, the intake line of the MGA[7] was set up in the same way as the flask sampling lines. We clarified this in Line 215 and Table 2.

Accordingly, all intake lines in Munich were 100 m long. The flow rates, on the contrary, were different. Due to the long inlet, there were significant spectral losses. However, these spectral losses were smaller than anticipated and comparable to those in Paris (an installation of the MGA^7 closer to the intakes in Munich had been intended, but this was not possible due to a lack of space further up in the tower). The suitability of the MGA^7 for turbulent flux measurements in all three cities is demonstrated in Fig. R2 and Fig. R3. Figure R2 shows the average normalized spectra and co-spectra against normalized frequency and theoretical slopes according to Kaimal et al. (1972). The (co-) spectra of the open-path IRGASON are shown for reference. In each city, the low-frequency, i.e., large energy-carrying eddies, are well captured by the MGA^7. The slopes of the (co-) spectra agree well with the theoretical expectations of -2/3 and -4/3, respectively. For high normalized frequencies, the impact of spectral attenuation in the intake line of the MGA^7 is visible. However, these losses were corrected by applying high frequency spectral corrections according to Fratini et al. (2012). The large contribution of high frequencies observed in the IRGASON data in Paris is speculated to be related to the proximity of a strong electromagnetic source (antenna), which may be the reason for the increase in the observed IRGASON noise (personal correspondence with the manufacturer). Fortunately, the IRGASON noise was not correlated to the sonic velocity, and was hence mostly filtered out by the covariance computation itself, which showed comparable w'T' and $w'CO_2'$ spectra (Bignotti et al. 2026, in prep.). This is confirmed by Fig. R3 showing that the CO_2 fluxes of IRGASON and MGA^7 agreed well in all three cities.

We added Fig. R2 and the related discussion to an additional section in the appendix (Appendix D in the resubmitted manuscript). For further details on the Zurich measurements, see Hilland et al. (2025).

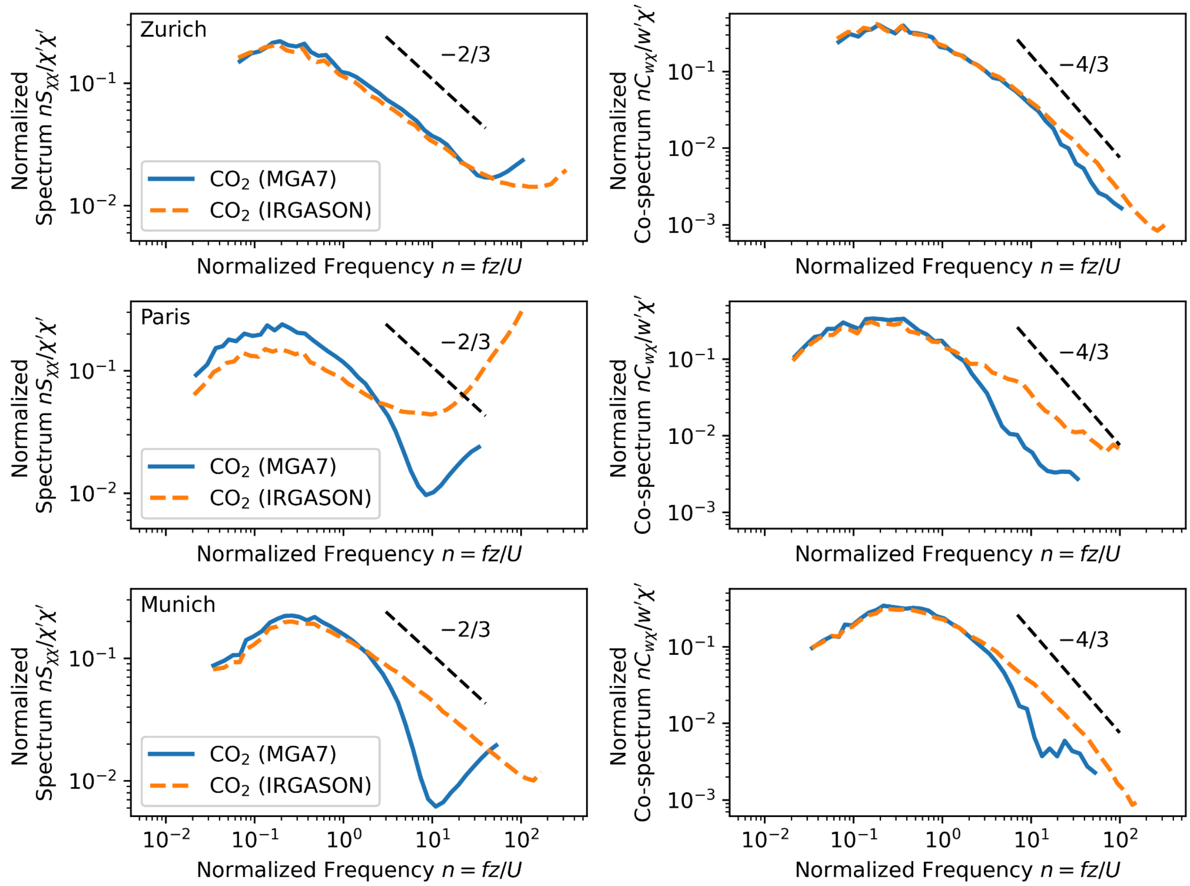

Figure R2. Average normalized spectra and co-spectra against normalized frequency and theoretical slopes according to Kaimal et al. (1972).

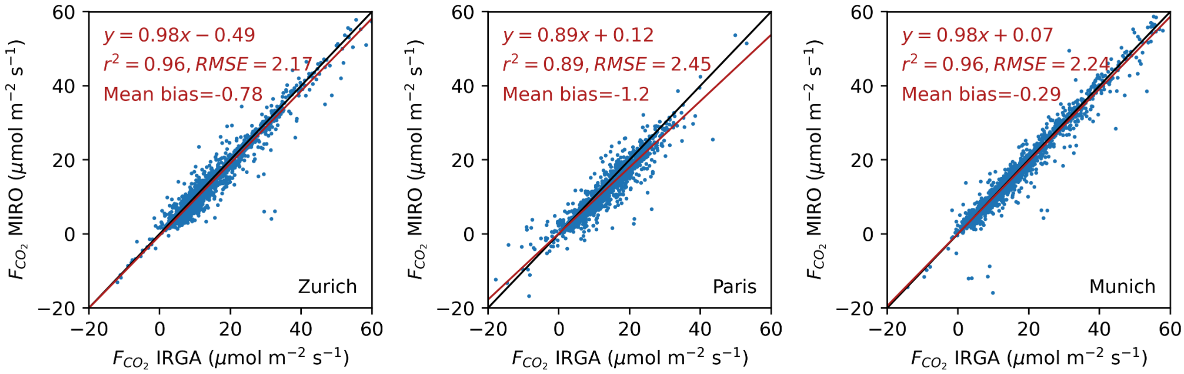

Figure R3. Comparison of the CO_2 fluxes derived from IRGASON and MGA[7] (MIRO instrument) measurements. Shown are the high- and medium-quality data (quality control flag 0 or 1 according to Mauder and Foken, 2004).

11, Line 243. The flux sites were located on the edge of a telecommunications tower. Please discuss the impact of the tower structure on the micrometeorological environment.

Due to the massive structure of the telecommunications tower at the Paris site, wind distortion effects were observed between ~70 °N and 120 °N. Figure R4 shows the pitch angle in dependence of the horizontal wind direction in Paris. The sinusoidal pattern indicates instrument tilt, while the “flattening” between ~45° and 160° is likely related to the wide tower platform and the strong distortion in the marked area to distortions from the main tower, the pylon, the instrument body etc.. To minimize the effect of flow distortion, REA measurements from this wind sector were rejected (Table 2 and Line 251 in the manuscript). As shown in Fig. F1 in the manuscript, most REA measurements in Paris were taken during south-southwesterly wind. Consequently, flow distortion effects are assumed to be negligible for the REA measurements.

We added a comment on the impact of the tower structure on the micrometeorological environment in Line 245. Please also see our response to Comment 9 of Referee #2, and the discussion in Lan et al. (2024), particularly Fig. R4, which presents results from the same tall towers and discusses possible building influence.

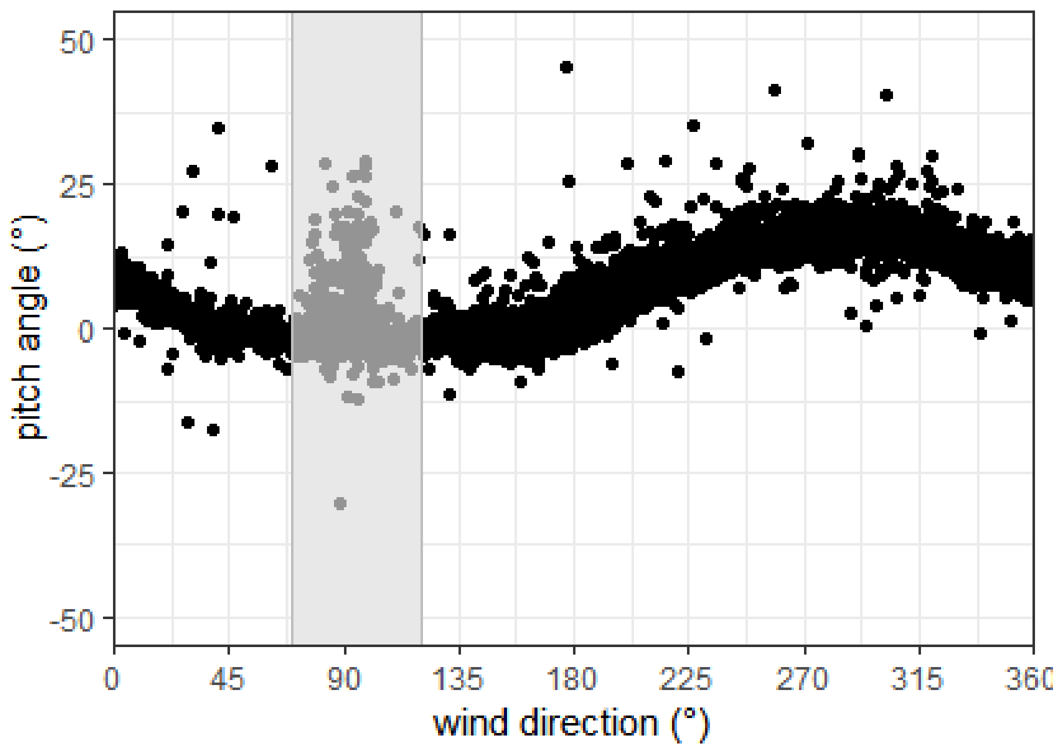

Figure R4. Pitch angle as a function of the horizontal wind direction at the Paris site between July 2023 and April 2024. Between 70 °N and 120 °N, flow distortion effects from the tower structure are visible.

12, Line 270 and Table 2. Please add the flow rate of the REA intake lines and the corresponding Reynolds number to show that turbulent flow through the lines was well-maintained. The flow rates of 7 L/min in Paris and 11 L/min in Munich may result in laminar flow in the intake lines. Please add discussions about the potential impact. The information about the setup for MGA sampling should also be included, as well as the lagtime determined.

For the REA measurements, the Reynolds number and its impact on the flow in the intake lines are not relevant, as the fast response sampling valves that determine the conditional collection of air depending on the vertical wind velocity, are mounted on top of the tower, close to the inlets. Consequently, it is not important whether the air is mixed in the lines or not, as it all ends up in the same flask. Any residual air in the intake lines at the start of a REA measurement, is flushed out. See Kunz et al. (2025) for details. To avoid any misinterpretation, we have not included the Reynolds numbers in Table 2.

For the MGA[7], spectral losses due to the intake line were corrected according to Fratini et al. (2012), leading to good agreement between the IRGASON and MGA[7] fluxes, as discussed in Comment 10.

13, Line 345. I totally agree with the authors that great uncertainty exists in determining the storage term, especially for urban flux measurement. In the manuscript, storage was calculated using the single-point profile. It would be great to conduct additional storage calculations using the CO2 concentration measured at two heights in Munich and compare with the single-point method. The authors set a threshold of 20 μmol m−2 s −1 to flag extremely large storage terms. Could you clarify the criterion used to determine this threshold?

We agree that the deployment of two midcost sensors at different heights in Munich provides a great opportunity to study the uncertainty-related estimation of storage fluxes in an urban environment. However, Crawford and Christen (2014), for example, showed that using vertical profile measurements does not necessarily improve the storage flux estimate due to the high horizontal variability in the CO_2 concentrations closer to the urban canopy layer. Moreover, we did not use the storage flux estimates for quantitative flux estimation, but only for a qualitative classification of the REA measurements. Therefore, we believe that comparing different storage flux estimates is beyond the scope of this study. A detailed analysis of storage flux estimation, using not only the two mid-cost sensors, but also a spatially distributed network of low-cost rooftop sensors in Zurich and Munich, is currently being prepared for publication. Preliminary results from this work show that the single-point method is already describing the magnitudes and temporal profiles of storage fluxes in Zurich and Munich accurately, and that the information from the vertical CO_2 profiles fine-tune the fluxes only slightly (Sigmund et al., 2025).

The flagging of extremely large storage terms with a threshold of ± 20 μmol m^{-2} s^{-1} is depicted in Fig. R5. The x-axis shows the absolute CO_2 flux during the REA period and the y-axis shows the maximum of the two 30 min storage flux estimates covering the REA measurement period. We considered the individual 30 min periods (instead of the mean storage flux during the 60 min REA measurement) to also flag measurements where a large positive storage flux was followed by a large negative storage flux, i.e., where the absolute mean storage flux was small. This was occasionally observed in the morning, when the CO_2 concentration first increased and then decreased, possibly due to venting of nocturnally accumulated CO_2 below the measurement height. Measurements with $|F_{CO2,\ strg\ max}|$ > 20 μmol m^{-2} s^{-1} (gray shaded area) were flagged as storage measurements. These measurements were analyzed together with the low-turbulence measurements, which had a friction velocity u_* < 0.2 m s^{-1} (shown as stars). An absolute storage flux threshold was used instead of a relative threshold of, for example, 50 or 100 % of the total CO_2 flux (dashed lines), because the latter would have flagged a large number of measurements with small

F_{CO2}, which seemed inappropriate given the large uncertainty of the storage flux estimates. Figure R5 shows that most measurements with $|F_{CO2,\ strg\ max}|$ > 20 µmol m^{-2} s^{-1} were already flagged due to other flagging criteria and were not further analyzed due to QC < 2, β < 0.1, β > 1, or SNR < 100 % (semi-transparent markers), or were regarded as low-turbulence measurements due to a friction velocity u_* < 0.2 m s^{-1} (stars).

The uncertainty of our classification of REA measurements is acknowledged and discussed in the text, e.g., in Lines 345, 482 ff, 610, 707. We added Fig. R5 to the discussion of the storage flag in Appendix B4.

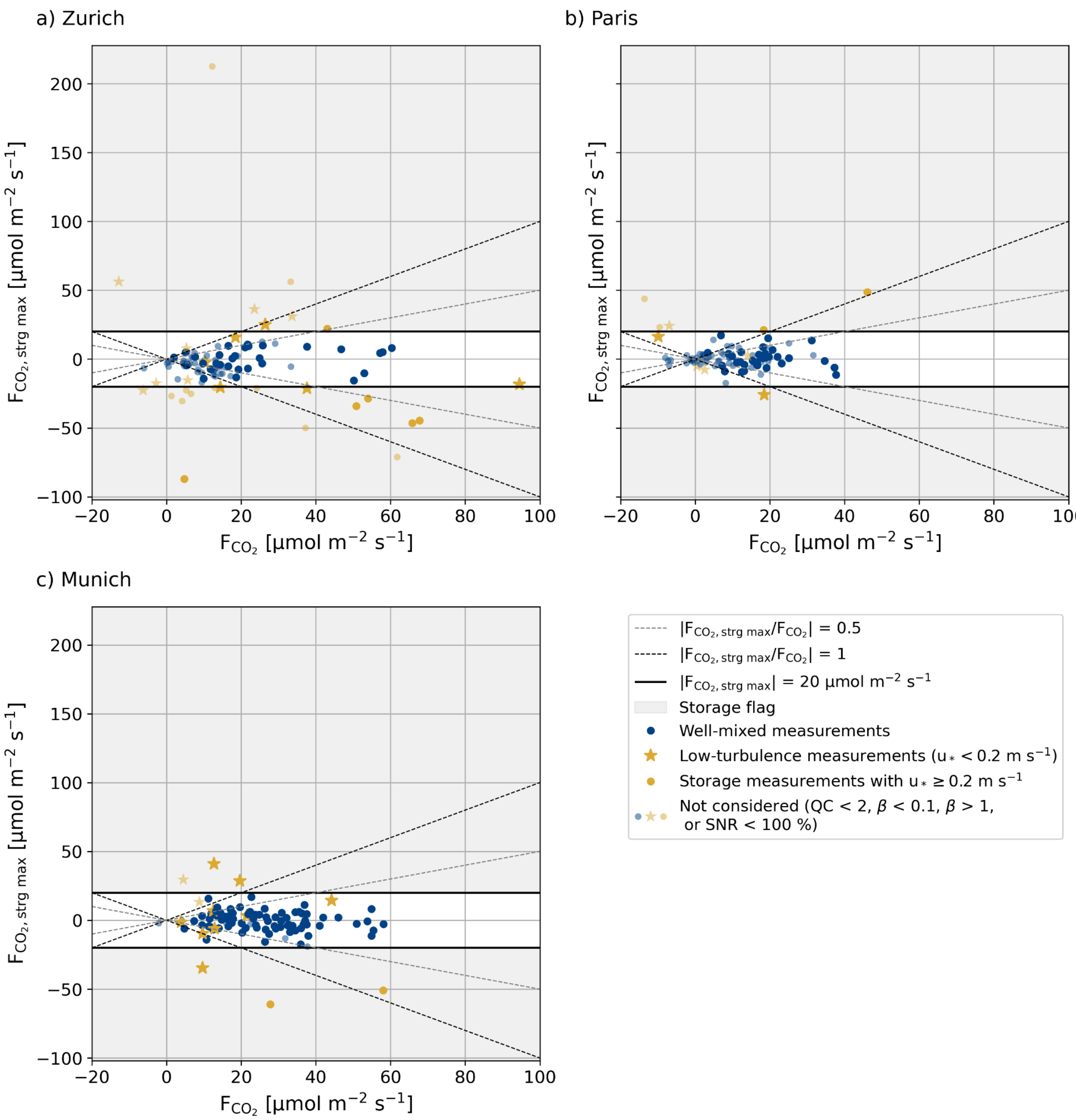

Figure R5. Maximum of the 30 min storage flux estimates with respect to the mean absolute CO_2 fluxes during the REA measurement periods. Measurements with $|F_{CO2,\ strg\ max}|$ > 20 µmol m^{-2} s^{-1} were flagged as storage measurements and analyzed together with the low-turbulence measurements ($u*$ < 0.2 m s^{-1}).

14, Line 365-368. Obviously, there were many fewer negative ffCO2 flux data points in Munich than other two cities. It may be worth noting this to highlight the advantages of the REA measurements in Munich.

We have gladly included a corresponding note in Line 368.

15, Line 385-391. Based on the data measured in Zurich and Paris, the ffCO2 fluxes in winter were systematically higher than in summer, while it looks like there was no significant difference between summer ffCO2 fluxes and winter ffCO2 fluxes in Munich. Could you add some discussion? Probably the contribution of natural gas consumption?

Table 2 shows the means and standard deviations of the $ffCO_2$ fluxes of the well-mixed summer and winter measurements in each city. As described by the Referee, the difference between summer and winter is much larger in Zurich and Paris than in Munich. We were mainly puzzled by the very small $ffCO_2$ fluxes in summer in Zurich and Paris, which we believe are not representative given the large uncertainties and small number of measurements. In Munich, $ffCO_2$ fluxes are still larger on average in winter than in summer, but we agree that the difference is not significant. One possible reason for this could be a relatively large contribution from district heating in the area surrounding the tower. However, a rough analysis of emission inventories suggests that the relative contributions of stationary combustion and road transport in the vicinity of the tall towers are similar at all three sites, so this is unlikely to fully explain our observations. Please also see our response to Comment 38 from Referee #2.

We added a note in Line 415 of the resubmitted manuscript.

Table 2. Mean and standard deviation of the $ffCO_2$ fluxes of the well-mixed REA measurements in µmol m^{-2} s^{-1}.

	Summer measurements	Winter measurements
Zurich	6 +- 8 (*N* = 6)	22 +- 20 (*N* = 24)
Paris	-1 +- 9 (*N* = 8)	16 +- 23 (*N* = 25)
Munich	12 +- 11 (*N* = 40)	19 +- 8 (*N* = 38)

16, Line 491-492. It would be great to provide the exact proportions of the vegetated area relative to the footprint areas and have this discussed in the nfCO2 fluxes comparison between cities.

We added the percentage of vegetated area within the mean footprints to Table 2 (33 % in Zurich, 17 % in Paris, and 23 % in Munich), and expanded the discussion in Lines 491 ff. To obtain comparable data for all three cities, we used the WorldCover product with 10 m resolution provided by the European Space Agency (https://esa-worldcover.org, last access: 12 December 2025).

17, Line 492-495. Do the 2.5 μmol m−2 s −1 values represent the mean annual human respiration fluxes for one city, or are they the same for all three cities? Please clarify.

The ~ 2.5 μmol m^{-2} s^{-1} refers to the mean annual human respiration flux in a 2 x 2 km^2 area around the tower for all three cities. In Zurich, the exact value for the year 2022 was 2.5 μmol m^{-2} s^{-1}, in Paris 2.3 μmol m^{-2} s^{-1}, and in Munich 2.8 μmol m^{-2} s^{-1}. We clarified this in Line 495.

18, Line 727-729. Although the authors acknowledge the high-frequency attenuation caused by the MGA intake lines, no spectral analysis was provided to evaluate this effect, nor were potential factors affecting the coverage of energy-carrying eddies discussed. Please add more related information.

We included the spectral analysis together with the other MGA[7]-related information regarding the setup, data processing and comparison with IRGASON measurements (see Comments 3, 4, 10, 12) in an additional section in the appendix (Appendix D in the revised manuscript).

References:

Ammann, C. and Meixner, F.: Stability dependence of the relaxed eddy accumulation coefficient for various scalar quantities, J.-Geophys.-Res. 107 (D8), https://doi.org/10.1029/2001JD000649, 2002.

Bowling, D. R., Delany, A. C., Turnipseed, A. A., Baldocchi, D. D., and Monson, R. K.: Modification of the relaxed eddy accumulation technique to maximize measured scalar mixing ratio differences in updrafts and downdrafts, J. Geophys. Res., 104, 9121–9133, https://doi.org/10.1029/1999JD900013, 1999.

Bowling, D. R., Tans, P. P., and Monson, R. K.: Partitioning net ecosystem carbon exchange with isotopic fluxes of CO2, Global Change Biol. 7 (2), 127–145, https://doi.org/10.1046/j.1365-2486.2001.00400.x, 2001.

Crawford, B. and Christen, A.: Spatial variability of carbon dioxide in the urban canopy layer and implications for flux measurements, Atmos. Environ., 98, 308–322, https://doi.org/10.1016/j.atmosenv.2014.08.052, 2014.

Foken, T., Dlugi, R., and Kramm, G.: On the determination of dry deposition and emission of gaseous compounds at the biosphere-atmosphere interface, Meteorol. Zeitschrift 4, 91–118, https://doi.org/10.1127/metz/4/1995/91, 1995.

Fratini, G., Ibrom, A., Arriga, N., Burba, G., and Papale, D.: Relative humidity effects on water vapour fluxes measured with closed-path eddy-covariance systems with short sampling lines, Agr. Forest Meteorol., 165, 53–63, https://doi.org/10.1016/j.agrformet.2012.05.018, 2012.

Hilland, R., Hashemi, J., Stagakis, S., Brunner, D., Constantin, L., Kljun, N., Molinier, B., Hammer, S., Emmenegger, L., and Christen, A.: Sectoral attribution of greenhouse gas and pollutant emissions using multi-species eddy covariance on a tall tower in Zurich, Switzerland, Atmos. Chem. Phys., 25, 14279-14299, https://doi.org/10.5194/acp-25-14279-2025, 2025.

ICOS - CAL, ICOS - Central Analytical Laboratories, https://www.icos-cal.eu/, last access: 01 February 2026.

Kaimal, J., Wyngaard, J., Izumi, Y., and Cote, O.: Spectral characteristics of surface-layer turbulence, Quarterly Journal of the Royal Meteorological Society, 98, 563–589, https://doi.org/10.1002/qj.49709841707, 1972.

Kunz, A.-K., Borchardt, L., Christen, A., Della Coletta, J., Eritt, M., Gutiérrez, X., Hashemi, J., Hilland, R., Jordan, A., Kneißl, R., Legendre, V., Levin, I., Preunkert, S., Rubli, P., Stagakis, S., and Hammer, S.: A relaxed eddy accumulation flask sampling system for 14C-based partitioning of fossil and non-fossil CO2 fluxes, Atmos. Meas. Tech., 18, 5349–5373, https://doi.org/10.5194/amt-18-5349-2025, 2025.

Lan, C., Mauder, M., Stagakis, S., Loubet, B., D'Onofrio, C., Metzger, S., Durden, D., and Herig-Coimbra, P.-H.: Intercomparison of eddy-covariance software for urban tall-tower sites, Atmos. Meas. Tech., 17, 2649–2669, https://doi.org/10.5194/amt-17-2649-2024, 2024

Levin, I., Karstens, U., Eritt, M., Maier, F., Arnold, S., Rzesanke, D., Hammer, S., Ramonet, M., Vítková, G., Conil, S., Heliasz, M., Kubistin, D., and Lindauer, M.: A dedicated flask sampling strategy developed for Integrated Carbon Observation System (ICOS) stations based on CO2 and CO measurements and Stochastic Time-Inverted Lagrangian Transport (STILT) footprint modelling, Atmos. Chem. Phys., 20, 11161–11180, https://doi.org/10.5194/acp-20-11161-2020, 2020.

Mauder, M. and Foken, T.: Documentation and instruction manual of the eddy-covariance software package TK2, Arbeitsergebnisse, Universität Bayreuth, Abt. Mikrometeorologie, 26, 2004.

McCartt, A. D. and Jiang, J.: Room-Temperature Optical Detection of $^{14}CO_2$ below the Natural Abundance with Two-Color Cavity Ring-Down Spectroscopy, 7, 3258–3264, https://doi.org/10.1021/acssensors.2c01253, 2022.

Lehmuskoski, J., Vasama, H., Hämäläinen, J., Hokkinen, J., Kärkelä, T., Heiskanen, K., Reinikainen, M., Rautio, S., Hirvelä, M., and Genoud, G.: On-Line Monitoring of Radiocarbon Emissions in a Nuclear Facility with Cavity Ring-Down Spectroscopy, Anal. Chem., 93, 16096–16104, https://doi.org/10.1021/acs.analchem.1c03814, 2021.

Oncley, S. P., Delany, A. C., Horst, T. W., and Tans, P. P.: Verification of flux measurement using relaxed eddy accumulation, Atmos. Environ. 27 (15), 2417–2426, https://doi.org/10.1016/0960-1686(93)90409-R, 1993.

Pearson, R. J., Oncley, S. P., and Delany, A. C.: A scalar similarity study based on surface layer ozone measurements over cotton during the California Ozone Deposition Experiment, J. Geophys. Res. 103, 18919–18926, https://doi.org/10.1029/98JD01479, 1998.

Ruppert, J.: CO2 and Isotope Flux Measurements above a Spruce Forest, PhD thesis, University of Bayreuth, Germany, 2008.

Ruppert, J., Thomas, C., and Foken, T.: Scalar similarity for relaxed eddy accumulation methods, Bound.-Lay. Meteorol. 120 (1), 39–63, https://doi.org/10.1007/s10546-005-9043-3, 2006.

Sigmund, A., Stagakis, S., Hilland, R., Brunner, D., Christen, A., Feigenwinter, C., Vogt, R., Emmenegger, L., and Kalberer, M.: Direct measurements of net CO2 emissions using a tall-tower eddy-covariance system and a rooftop CO2 sensor network in the city of Zurich, 12th International Conference on Urban Climate, Rotterdam, The Netherlands, 7–11 Jul 2025, ICUC12-713, https://doi.org/10.5194/icuc12-713, 2025.

WMO (World Meteorological Organization): 20th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Measurement Techniques (GGMT-2019), GAW-Report No. 255, Geneva, Switzerland, 2020.