Reply to Referee #2: Kenneth Davis

We would like to thank Kenneth Davis for his detailed and constructive feedback on this manuscript. We restructured the paragraphs and sections as best as possible and hope this has increased the readability and the impact of the work. 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 manuscript presents a substantial and successful effort to employ 14CO2 measurements to disaggregate urban CO2 flux measurements into fossil and non fossil components. This is an important contribution to urban eddy covariance flux literature that is definitely worthy of publication. Eddy covariance flux work is prone to detailed debates over methodology. I have no significant concerns, however, with the methods chosen by these investigators. They have documented their choices well and all appear defensible and reasonable. I doubt that alternative choices would have a significant impact on their results.

My primary concerns are with the presentation of their results. The results and conclusions, and the associated elements of the abstract, deserve, in my opinion, significant revision to ensure that this manuscript is as impactful as possible. I will call this "major revisions" but they are mainly editorial revisions, not major revisions to the scientific content of the manuscript. If the authors do not find these suggestions to be helpful I would respect their decision. And as always, apologies for any misunderstandings. I look forward to your responses.

Major suggestions.

The results are not clearly written. The text of the results wanders from methods to descriptions of the content of figures (which should be contained in the figure captions) to a description of the results. The main results are often buried near the end of the relevant sections and are not always stated strongly and clearly. I suggest that the authors pick out the main results - their primary findings - and reorganize the text of the results around these findings. Second, the portions of methods in the results that are unique (means of analysis of the flux measurements) should be moved to the methods section of the document. Redundant methodological text should be removed. The sharpened focus on the main results should be translated into the abstract. Finally, I suggest removing the summary of results from the conclusions and focusing the end of the manuscript on a strong set of take away messages, passing on the recommendations this work yields for how our community should move forward. I believe that a revision with these guidelines (see more notes in the detailed comments) will significantly improve the clarity and impact of the manuscript.

We followed the suggestions as follows:
- At the beginning of each subsection of the results section, we summarized our main findings from that paragraph.
- We moved the description of methods from the results section to an additional subsection in the methods sections (Sect. 2.3 in the resubmitted manuscript). We

removed redundant text and added a flow chart that provides an overview of the individual steps involved in the REA measurements and analysis.
- We revised the abstract to include the main results, focusing more on the progress of the methodology.
- We split the conclusions section of the original manuscript into separate "summary" and "conclusions" sections. Having first summarised our results with respect to the research questions raised in the introduction in a separate section, we extended the conclusions section to contain only the key messages and recommendations for future investigations.

I have a second overall suggestion. The authors discuss the use of CO for disaggregating ff and nf CO2 fluxes and the benefits of using 14CO2. The results, however, do not revisit this issue at all. Are CO measurements available? Is it possible to add a discussion of the CO/CO2ff ratios implied by the 14CO2 measurements to this document? It would improve the impact of the manuscript even further. If this is beyond the scope of the authors for this manuscript, could the manuscript at minimum state the availability of CO measurements for this purpose?

CO measurements are available from the MGA[7] (continuous high-frequency measurements for flux calculation) and from the REA flasks (concentrations of updraft and downdraft samples). As the manuscript under review is already very comprehensive, an analysis of the CO/ffCO$_2$ flux ratios observed during the REA sampling periods and subsequent calculation of a continuous CO-based ffCO$_2$ flux record are currently being prepared for a separate publication, as part of the dissertation of the first author. We added a note on the availability and potential use of CO measurements at the beginning of the Methods section (Line 66). In addition, we refer to these multi-species MGA[7] and flask measurements in the respective sentence in the conclusions (Line 611 of the original / Line 610 of the resubmitted manuscript).

A third major (but relatively minor note): The flagging approach is confusing. I suggest trying to clarify this presentation.

We revised the description of the flagging of REA measurements in Sect. 2.2.4 and added a flow scheme to visualize our approach (Fig. 2 in the resubmitted manuscript). Please see our response to Comment 16 for further details.

1. Abstract.
I believe that demonstration of this methodology and its strengths and weaknesses is the primary contribution of the manuscript. The abstract primarily presents the somewhat limited information about the mixture of ffCO2 and nfCO2 fluxes that can be derived from the available observations without saying much at all about what the authors learned regarding how to make this methodology work. I would put more focus in the abstract on the methodological progress and use these urban system flux results as examples of what can be learned from this methodology.

We agree that the abstract should summarize what we learned from and about the methodology. Despite the limited representativity and the large uncertainties of our results, we would also like to focus on the actual flux observations, since the technical details of the REA system and the partitioning approach have already been discussed in Kunz et al. (2025).

We rephrased the abstract to focus more on the results from the Munich measurements as an example and proof of the improvements in the methodology.

2. Line 4, "with mean ffCO2 contributions of about 80%."
Is this a magnitude? How is the potential for the changing sign of biological fluxes (even in winter) taken into account?

Please see the following answer below Comment 3.

3. Line 7. "the observed nfCO2 fluxes were predominantly positive (~50% of net summer fluxes),"
Could you be more precise about what this figure represents? CO2 fluxes vary a lot with time.

These are the error-weighted average $ffCO_2/CO_2$ flux ratios (or $1 - ffCO_2/CO_2 = nfCO_2/CO_2$) of the well-mixed REA measurements, excluding measurements potentially influenced by emissions from the district heating plant in Zurich or the brewery in Munich, as presented in Sect. 4.3.3. Summer measurements were taken between 06:00 and 21:00, and winter measurements between 08:00 and 19:00 local time.
We specified this and provided the exact numbers from the summer and winter measurements in Munich. We removed the information from the Zurich and Paris measurements, since the results were not significantly different to the null hypothesis of 100 % fossil emissions, due to the small number of measurements and the large measurement uncertainties. Please see also our response to Comment 1.

4. Line 38. "air is conditionally collected for one hour in an updraft or downdraft reservoir"
Air is separated into updraft and downdraft reservoirs, correct? Each hour doesn't result in either an updraft or a downdraft reservoir being filled. Please clarify / rewrite.

Agreed. We clarified in Line 38: On a tall tower over the city, air is conditionally collected during one hour in two separate reservoirs (an updraft and a downdraft reservoir) using fast-switching sampling valves. The valves respond to a 20 Hz vertical wind signal from a 3D ultrasonic anemometer.

5. Line 70-72. "The 20 Hz CO2 measurements of the IRGASON were despiked by discarding measurements where the median absolute deviation of less than three

To remove erroneous spikes from the raw 20 Hz IRGASON CO_2 time series caused by instrument malfunction or obstructions in the path of the IRGASON gas analyzer (e.g., animals, dirt, rain, snow), we detected and removed short-time outranged data prior to flux calculation with EddyPro. A comparison of the measured CO_2 differences of the REA flask sample pairs with the IRGASON- or MGA[7]-based estimates revealed that the standard despiking methods by Vickers and Mahrt (1997) and Mauder et al. (2013) were too strict for the urban environment, as they removed real peaks in concentrations, leading to a systematic underestimation of REA flask concentration differences. Therefore, we modified the Median Absolute Deviation (MAD) method described by Mauder et al. (2013). Measurements were considered as spikes if no more than three consecutive 20 Hz measurements exceeded the range defined by Mauder et al., (2013). With this despiking method, the flask, IRGASON, and MGA[7] data agreed much better (see Appendix E in the manuscript).

We expanded the description in Line 72.

6. Line 72. "were upsampled to 20 Hz"
What does it mean to be upsampled?

To merge the 10 Hz MGA[7] data with the 20 Hz IRGASON data, we upsampled the MGA[7] data to the timestamps of the IRGASON measurements using a nearest-neighbor approach with a 50 ms search window. This essentially duplicated the MGA[7] CO_2 measurements.

We expanded the description in Line 72. Please see also our response to Referee #1 Comment 2.

7. Line 75. "the time lag was determined via linear interpolation to correct clock drift"
Linear interpolation of what? What clock drift?

The upsampled MGA[7] and the 20 Hz IRGASON CO_2 measurements will be shifted in time due to clock drift between the two instruments (the MGA[7] was synced to an NTP server once per week, the IRGASON logger only periodically) and due to the travel time of the sample gas to the MGA7 instrument. To synchronize the two data sets, we identified the time lag at which the high-frequency CO_2 concentrations showed maximum correlation and shifted the MGA[7] measurements accordingly. However, for individual 30 min periods during which the correlation between the IRGASON and the MGA[7] CO_2 signals is poor, e.g., due to low IRGASON signal strength during a rain event, this approach yields unreasonable time lags. For these periods, we removed the erroneous time lags and determined them by linear interpolation of the preceding and subsequent time lags.

We specified and expanded the description in Line 75 and added further information in an additional section of the Appendix (Appendix D in the resubmitted manuscript). Please see also our response to Referee #1 Comment 2.

Yes, we compared the spectra of the MGA[7] and the IRGASON. The spectral analysis revealed high-frequency losses, but spectral corrections for both instruments led to very good agreement between the flux estimates. Please refer to our response to Referee #1, Comment 10 for details. We added the spectra and co-spectra of both instruments in an additional section in the Appendix (Appendix D in the resubmitted manuscript).

Yes, the mean vertical wind velocities show a directional dependence (Fig. R1), which can be attributed to instrument tilt and flow distortion caused by the tower structure. However, deviations from horizontal wind are considered small, especially in Zurich and Munich (see also Hilland et al. (2025), Lan et al. (2024), and our response to Referee #1 Comment 11). Above all, these deviations are taken into account in the EC measurements through coordinate rotation during data processing, and in the REA measurements through adjustment of the deadband. While the method used to compute the deadband affects the magnitude of the concentration differences between the updraft and downdraft reservoirs, the beta coefficient calculated from CO_2 for each measurement period individually accounts for the deadband used when calculating fluxes (Pattey et al., 1993). Therefore, the results are independent of the method used to compute the deadband.

We added a note regarding the independence of the results on the method used to compute the deadband to Line 110 (Line 122 in the resubmitted manuscript).
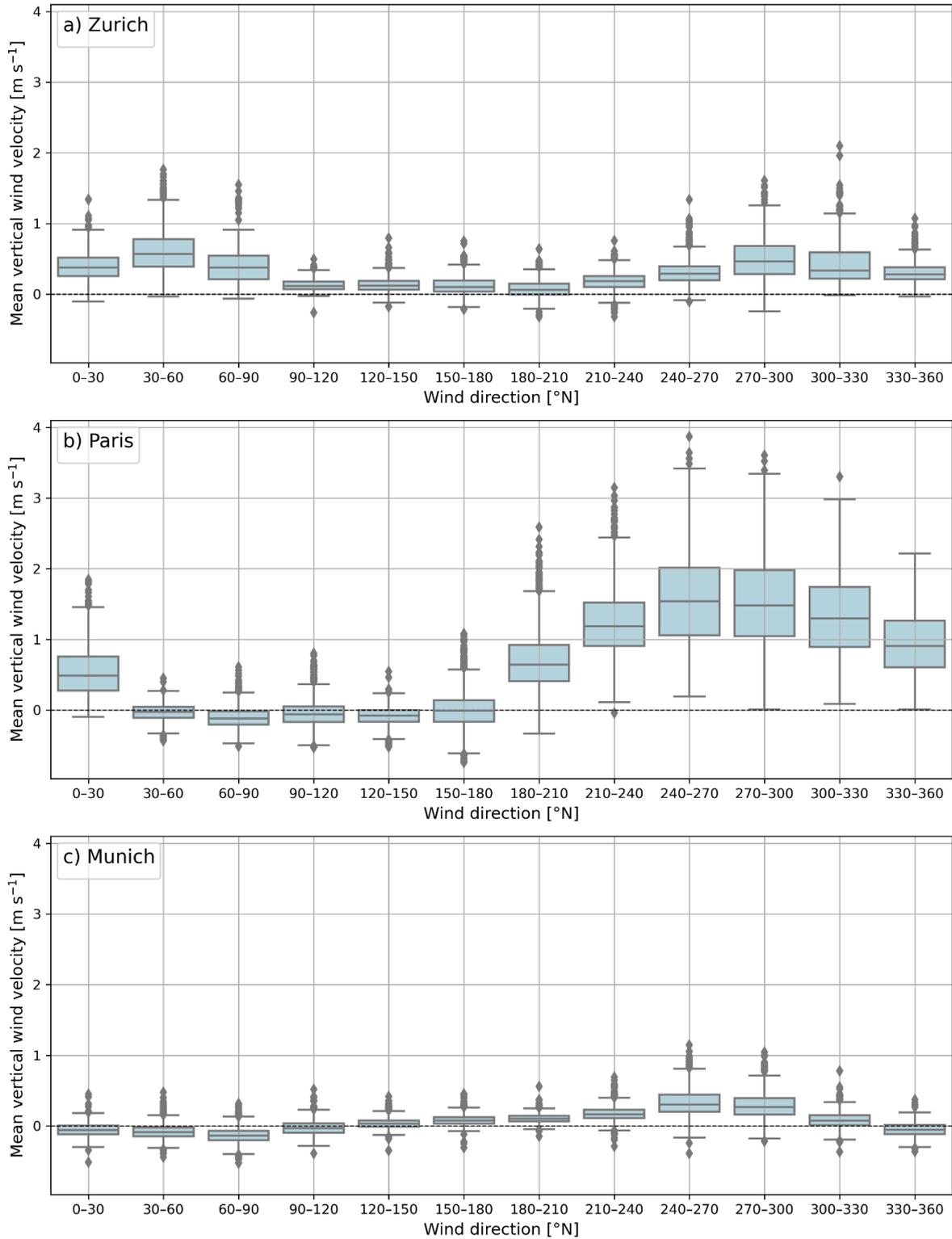
Figure R1. Mean vertical wind velocity in dependence on the horizontal wind direction, based on all measurements where wind speed was > 1 m s^{-1}.

Agreed. We changed it to "net CO_2 flux" in Line 126.

11. Line 137. There is also turbulent sampling error in the vertical mole fraction differences. I cannot, however, recommend readily accessible approaches from the literature for estimating this uncertainty. But it would be worth noting that this exists. The upper limit could be estimated as the sum of random errors in each individual mole fraction measurement. Surface layer variances from similarity theory and turbulent sampling error theory (e.g. Lenschow and Stankov, 1986) could be used to estimate this uncertainty.

We added a note in Line 138 (Line 147 in the resubmitted manuscript) about the existence of additional uncertainties, including the turbulent sampling error, which are considered less relevant and not taken into account in our analysis.

12. Line 147-149. "The ffCO2 contribution to the storage fluxes equals the ratio of the flux averages over the period during which CO2 accumulated below the measurement height, which does not necessarily equal the surface flux ratio during the measurement period."
I do not object to the authors' choice to exclude the storage flux from their flux estimates and to present data from low turbulence conditions separately but I do not understand this statement. Storage flux represents accumulation in the column which occurred during the time period of the turbulent flux measurement. I don't understand how the time period for storage can be significantly different from the time period for the turbulent flux measurement.

We agree that, for a per convention positive storage flux representing accumulation below the measurement height during the time period of the turbulent flux measurement, the time periods for storage and for the turbulent flux measurement are identical. In this case, the $ffCO_2/CO_2$ ratio of the storage flux should equal the ratio of the turbulent flux. However, for a negative storage flux, i.e., the venting of CO_2 from below the measurement height, we assume that the $ffCO_2/CO_2$ ratio of vented air will equal the ratio of the integrated fluxes over the period during which the CO_2 accumulated. If the CO_2 accumulated at night, when traffic and heating emissions are reduced, the $ffCO_2/CO_2$ ratio should be much lower than that of the turbulent (= surface) flux measured during the morning rush hour, for example. This assumption is supported by the generally lower $ffCO_2/CO_2$ ratios observed in the low-turbulence and storage measurements in the morning (Sect. 4.4).

We specified this in Line 149 (Line 160 in the resubmitted manuscript). It is also discussed in more detail in Appendix B4.

13. Line 193-194. "Following Hensen et al. (2009) and Osterwalder et al. (2016), measurements for which Eq. (3) does not provide reasonable ffCO2 fluxes were flagged according to β (Eq. 2)."

This criterion for flagging isn't clear to me. How are "reasonable" ffCO2 fluxes defined? And what does it mean to flag a flux "according to beta"?

Please see the answer below Comment 16.

14. Line 194-196. "Measurements with large uncertainties due to the limited resolution of the 14CO2 differences between updraft and downdraft samples were flagged based on the signal-to-noise ratio SNR, defined as the minimum of the relative FffCO2 and FnfCO2 uncertainties."
Similarly, this criteria isn't entirely clear. Is there a threshold SNR below which a flag is applied? If so, what is that threshold? Or is there a SNR estimate given to each flux measurement?

Please see the answer below Comment 16.

15. OK, I see that my questions are answered mostly by Table 1. Please point out Table 1 before describing these flagging cases so that the reader can readily find the threshold values while reading the conditions.

Please see the answer below Comment 16.

16. Table 1. I am a little slow, but the flagging scheme isn't clear to me. Conditions followed by "/" mean the data are not considered further. What does the & represent? "The other flags were only assigned if all criteria were met" I don't really understand what that means.

To Comments 13, 14, 15, and 16:

In the Sect. 2.2.4, we had first defined the relevant variables (QC, β, SNR, u_*, and $F_{CO2,strg}$) and later specified the thresholds for categorization of the measurements, as given in Table 1 of the original manuscript. To clarify, we added a Table for the description of the variables (Table 1 in the resubmitted manuscript) and replaced the original Table 1 by a diagram (Fig. 2 in the resubmitted manuscript), to which we refer at the beginning of the section. Thereby, we now describe low-turbulence and storage measurements as one category to emphasize that all measurements are either categorized as "well-mixed measurements", "low-turbulence and storage measurements", or "not considered further". See also our response to Comments 24 and 25.

17. Lines 214-217 are redundant to earlier text, no? Please combine / condense the texts.

Agreed. We removed the sentence in Line 215 - 217, as the handling of time lags has already been explained in Sect. 2.1.

Please specify the averaging interval used to create these source area estimates.

The flux footprints were derived for each 30 min averaging interval, for which EC turbulence data was available as input data. In addition, we calculated an aggregated footprint using every 30 min time step at which well-mixed REA measurements were conducted (Fig. 1).

We specified this in the text, but moved the description to the methods section (Sect. 2.3.1 in the resubmitted manuscript).

19. The methods section lacks the methods for analysis of the resulting flux measurements. Please describe, briefly, your analysis methods.

Agreed. We added a subsection (Sect. 2.3 in the resubmitted manuscript) to the methods section, where we briefly explain the calculation of flux footprints and mean land cover fractions (taken from Sect. 3), the calculation of error-weighted average $ffCO_2/CO_2$ ratios and average $nfCO_2$ fluxes (taken from Sect. 4.3.3), and the calculation of regional $ffCO_2$ enhancements (taken from Sect. 4.5). For clarity, we also provided an overview of the individual steps in the measurement and analysis procedure (Fig. 1 in the resubmitted manuscript). Please see also our responses to Comments 22 and 45.

20. Line 274. "Regular" not "Reqular"

Corrected, or more precisely removed, as described in Comments 21 and 22.

21. Line 274-277. "However, as the quality of the REA measurements varies depending on the micrometeorological conditions during sampling and the signal-to-noise ratios, the analyzed REA measurements were flagged as well-mixed measurements, low-turbulence and storage measurements, or were not considered further (Sect. 4.1)." Wasn't this already described in the text associated with Table 1? Please reduce redundancy in the text.

We removed the entire paragraph, as suggested in Comment 22.

22. Line 274. This paragraph is methods. Please merge this into the methods section and remove this from the results. These are not results.

We removed the paragraph in Line 274 ff, as suggested. The description of our flux analysis was included in the Methods (Sect. 2.3 in the resubmitted manuscript), and a reference to the good results of the quality control tests and spectral analyses, shown in the Appendix, was added in Line 273 in the resubmitted manuscript.

23. Line 285-286. "the criteria finally considered, describing suitable well-mixed conditions"
This is verbose and ambiguous. Do you mean just well-mixed conditions, or do these cases satisfy all of the criteria outlined in Table 1? Please clarify.

We meant that 30 out of 87 measurements were flagged as well-mixed measurements as defined in Sect. 2.2.4. We clarified it in Line 285 (Line 337 in the resubmitted manuscript).

24. Table 3 caption. "flagged as well-mixed, low-turbulence, storage or not considered further (Sect. 2.2.4)."
Apologies, but I don't understand what this means. Please make your data filtering choice easier for your readers to follow. Perhaps a flow chart would be helpful, instead of a table, at some point in the methods. I believe that Table 3 is showing a progression of filtering choices.

Please see the following answer below Comment 25.

25. Table 3. I don't see a row that presents how many flux measurements you DO consider. I also can't tell if the ""not considered" cases overlap, for example, with the well-mixed cases. The number of cases don't add up to the first row. I think that the first three rows add up to 100% of cases. Maybe you could add the statement that all ffCO2 flux cases are "categorized either as well-mixed, low-turbulence/storage, or not considered due to quality control flags".

As noted at Comment 15, we added a flow chart in the methods to emphasize that all REA measurements with ffCO$_2$ fluxes were categorized as either well-mixed measurements, low-turbulence/storage measurements, or measurements not considered further. The well-mixed measurements are analyzed in most detail, the low-turbulence and storage measurements are only briefly analyzed, and all other measurements are, as the name says, not considered further.

We adapted the structure of Table 3 and specified the caption.

26. Line 287-298. "In Paris, low-turbulence and storage measurements were mostly discarded."
Did you use different filtering criteria for the Paris data? Or is it just that the low-turbulence data in Paris were mostly filtered out by the quality control flags?

In Line 287, we referred to the rejection of low-turbulence and storage measurements during the Paris measurement campaign, meaning that flask samples collected when u_* < 0.2 m s^{-1} or $|F_{CO2, strg\ max}|$ > 20 µmol m^{-2} s^{-1} were usually not selected for laboratory analysis. In Zurich and Munich, on the contrary, several low-turbulence and storage measurements were intentionally selected to analyze the ffCO_2/CO_2 ratio of nocturnally accumulated CO_2. The differences therefore stem from changes in the focus of our research.

We clarified this in Line 298. The different selection strategies were also addressed in Lines 286 - 289 and 306 - 307.

27. Line 314-315. "To illustrate the principle and to show an example of partitioning net fluxes of CO2 collected via EC into fossil and non-fossil CO2 flux components using REA, Fig. 2 presents data collected on 09 October 2024 in Munich. On that day, micrometeorological conditions were suitable for REA measurements and six flask pairs, sampled between 08:00 and 19:00 local time (UTC+2), were analyzed for 14CO2. The hour-long sampling periods are highlighted in Fig. 2 in light blue."

Apologies to be an old professor, but allow me to make a suggestion. I would tell my students that these sentences are a combination of a statement of methods and content suited for the caption of the figure. This is not appropriate for the topic sentence of a paragraph in the results section. Move these sentences to the methods orto the figure caption. Please begin paragraphs in the results sections with topic sentences that summarize the results that you are presenting - tell us what you have learned. Use the paragraph to explain that topic sentence in more detail. I believe that if you follow this advice your text will be more concise and focus more clearly on your primary results.

We re-organized the paragraphs in the results as suggested. For this purpose, we moved the conclusive sentences in all subsections to the beginning of the respective paragraphs. Please see also the answers to the comments below.

28. Line 327-329. "The continuous EC measurements (Fig. 2 d) show that the turbulent CO2 fluxes at 85 m height are approximately 10 µmol m−2 s−1 in the early morning, increase after sunrise, and reach a maximum of more than 60 µmol m−2 s−1 at noon, before they decrease again."

Please excuse another old professor comment. This text describes the contents of the figure. The figure does that already. Use the text to tell your readers the story of what you have learned from the figure.

We removed the description in Lines 327 - 330 and moved the discussion in Line 337 ff further up.

We specified our main results, including the successful separation of $ffCO_2$ and $nfCO_2$ fluxes, and moved this part to the beginning of the paragraph. In addition, we emphasized the unexpectedly high $nfCO_2$ flux observed in the early afternoon, with reference to the discussion of the brewery emissions in Sect. 4.3.2, and the absence of expected traffic peaks in the $ffCO_2$.

Agreed. We used the city names as synonyms for the measurement sites; the $ffCO_2/CO_2$ ratios, of course, are specific to the measurement sites and measurement campaigns. We replaced the introductory sentences in Line 354 - 359 with the main findings from Sect. 4.3.1 (see also Comment 31, 33, and 34).

We replaced the introductory sentences in Line 354 - 359 with the main findings from Sect. 4.3.1.

As all relevant information is already provided in the figure caption, we removed the description in Lines 360 - 364.

33. Lines 369-370. "For comparison, the median fluxes of the continuous EC CO2 measurements are shown in the left panels of Fig. 3 for both summer and winter."

Figure caption please. What did you learn? I can't yet tell. Start by stating your results - what you have learned. Then explain how you have learned it. Let the figure captions do the job of explaining the contents and structure of the figures.

As mentioned also in response to Comments 30, 31, 32, and 34, we removed the description of the figure from the text and started the paragraph with our main findings.

34. Section 4.3.1 and the other results sections should be rewritten, in my opinion, with these principles in mind. As written the text is a mixture of methods, figure captions and some of what you learned, but the "learning" (the results!) are difficult to extract. Lead with what you have learned.

Agreed. In the resubmitted manuscript, we begin each paragraph with our key findings. In addition, we moved the description of methods from Sect. 4.3.2, 4.3.3, and 4.5 to the Methods section (Sect. 2.3 of the resubmitted manuscript) or the Appendix. Figure descriptions were removed, as all relevant information should be provided in the figure captions. Please also see our responses to the other Comments.

35. Figure 3. I would strongly recommend showing the fluxes in local time, similar to Figure 2. Human and ecosystem fluxes are best described as a function of local time, not UT.

Done.

36. Figure 3. The error bars in the y-direction on the REA fluxes do not include the uncertainties estimated in the 14CO2 measurements? The caption suggests that only random sampling error is included. Why not, as noted in the methods, include the merged impact of both significant sources of random error?

In the left panels on the y-axis, the total CO_2 fluxes obtained directly from EC measurements are shown. Therefore, only the random errors are considered. Uncertainties from $^{14}CO_2$ measurements are only relevant for the $ffCO_2$ and $nfCO_2$ flux estimates and shown in the right panels in Fig. 3. As the derivation of flux uncertainties is already explained in the methods section, we removed explanations of the uncertainties from the figure caption.

The IRGASON fluxes shown in Fig. 3 of the manuscript cover the entire time periods of the measurement campaigns (Table 2 in the manuscript). After quality control, between 40 and 100 measurements, or 40 to 70 % of the total number of time periods, were available for each 30 min period (Fig. R2). It is important to note that the continuous IRGASON flux measurements were only shown as reference. For quantitative analysis of the REA measurements, the fluxes measured during the individual REA measurement periods by the MGA[7] or, if not available, by the IRGASON, were used.

Nevertheless, we added the information in Appendix D of the resubmitted manuscript with a reference in the caption of Fig. 3.

a) Zurich

b) Paris

c) Munich
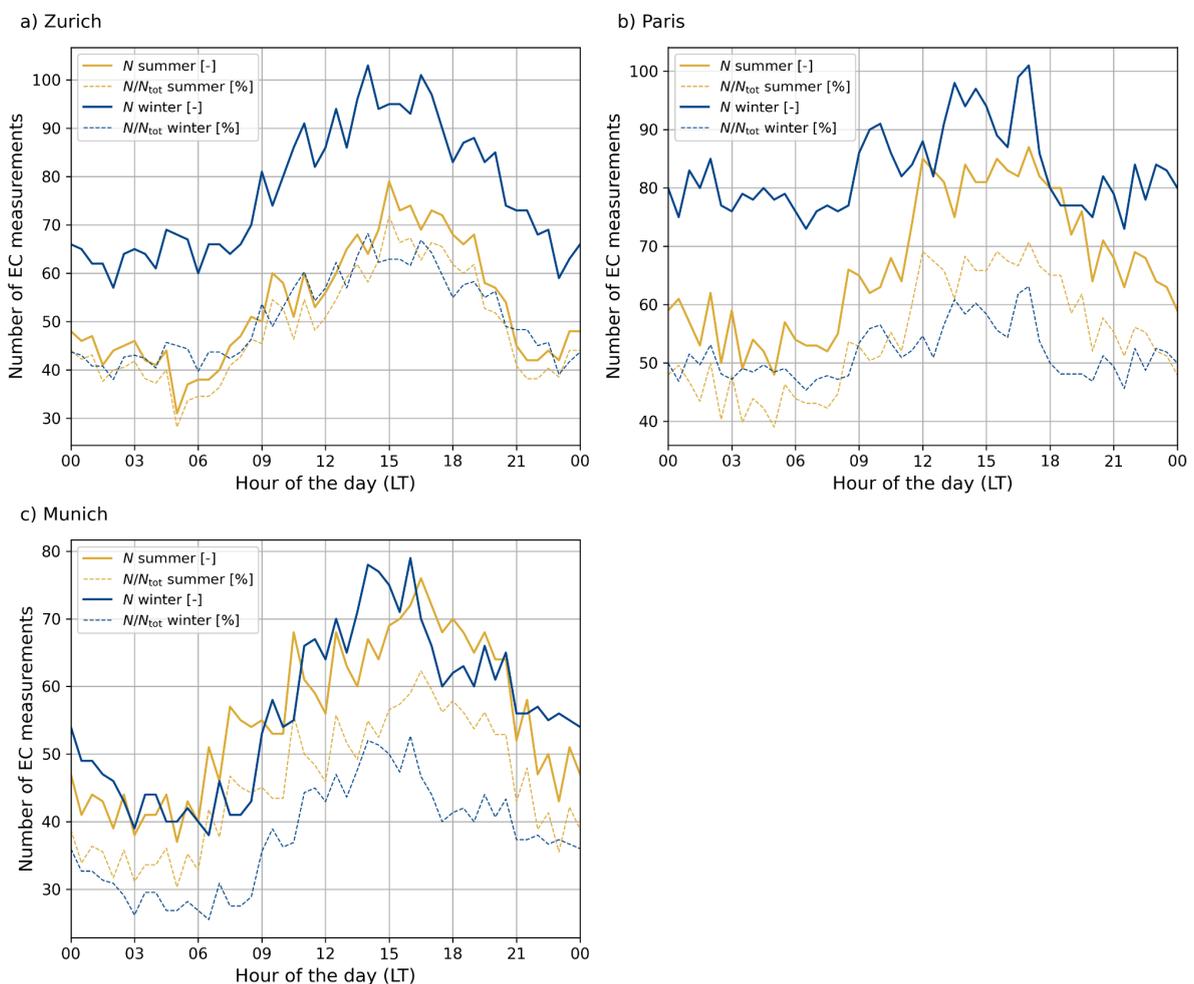
Figure R2. Number and percentage of retained measurements after quality control and filtering of the continuous CO_2 flux measurements from the IRGASON.

We agree that the very small $ffCO_2$ fluxes in Zurich and Paris, and the small difference between the summer and winter measurements in Munich, are surprising. We think that these results are not representative of the true seasonal averages, particularly in summer in Zurich and Paris, where the number of well-mixed summer measurements was < 10. In Munich, part of the small difference could be explained by the relatively large proportion of buildings in the footprint of the tower which are heated by district heating. In Zurich, on the contrary, the very large $ffCO_2$ fluxes observed in winter were potentially influenced by emissions from a district heating plant near the tower (Sect. 4.3.2). Average $ffCO_2/CO_2$ ratios excluding the measurements potentially influenced by the district heating plant in Zurich and a brewery in Munich are discussed in Sect. 4.3.3. A rough analysis of the emission inventories for the year 2022 in the areas around the three measurement sites could not explain the observed differences. In an approximately 2 x 2 km^2 area centered around the three tall towers, road transport accounted for about 20 % of the total annual CO_2 emissions (23 % in Zurich, 14 % in Paris, 24 % in Munich), while stationary combustion accounted for about 50 - 60 % in all three cities (56 % in Zurich, 62 % in Paris, 52 % in Munich) (Dröge et al., 2024). An ongoing study considering the actual flux footprints will further clarify expected differences.

We added a discussion of seasonal differences in Line 414 in the resubmitted manuscript. Please also see our response to Comment 15 from Referee #1.

In our view, the exceptionally high fluxes observed during the selected REA sampling periods emphasize that the dataset is biased towards high fluxes and is therefore likely to be unrepresentative of true seasonal or yearly averages. As mentioned in Line 384, this bias is caused by the systematic selection of flask pairs with significant differences in CO_2 concentration.

We followed the advice and started the paragraph with this description. We also added a note on the potential bias in Line 425 in the resubmitted manuscript.

We decided to keep Table 4 because, in our opinion, it shows the exact number of summer and winter measurements, the reduction in measurement uncertainty, and the low correlation between the CO_2 and $ffCO_2$ fluxes more clearly than Fig. 3. To further increase the value of Table 4, we added a note on the mean air temperatures in Line 421 in the resubmitted manuscript to emphasize that the observed differences in seasonal $ffCO_2$ flux variations between the three sites are not due to significantly different weather conditions.

41. Lines 411-414. "As emissions from point sources are generally not representative of the average fluxes in a city, and the comparison of measured and modeled point source emissions on an hourly basis is limited by uncertainties in the emissions inventory and transport models, we attempted to identify the REA measurements which were potentially influenced by emissions from the district heating plant."

I would argue that the city is a collection of point sources and no single measurements are "representative of the average fluxes in a city." I do not believe that you should try to represent average city fluxes with these measurements. I suggest that you drop the apologetic text. I fully support the effort to interpret your measurements with a flux footprint model and knowledge of likely large point sources. It is of course true that, "comparison of measured and modeled point source emissions on an hourly basis is limited by uncertainties in the emissions inventory and transport models." But isn't that part of what we are trying to test and learn - that is, the limits of our understanding of these fluxes and quantification systems? Embrace your effort. Don't apologize because the tools and methods are not perfect. All models are wrong.

We agree that the definition of a point source in a city is ambiguous. We also agree, and emphasized in the manuscript, that our limited number of measurements is not capable of representing the average EC fluxes (e.g., Lines 456, 596). However, in addition to the exemplary analysis of a single day (Sect. 4.2) and of individual high-flux measurements (Sect. 4.3.2), we believe that the REA measurements should also be used to estimate the approximate significance of non-fossil CO_2 fluxes in an urban environment without large point-source emissions. Moreover, large point-source emissions may violate the principles of stationarity and horizontal homogeneity and prevent the attribution of EC measurements to surface fluxes. Therefore, we decided to treat the measurements which were potentially influenced by emissions from the district heating plant in Zurich and the brewery in Munich separately. Nevertheless, we acknowledge that this discussion mostly concerns Sect. 4.3.3, in which we determined the average $ffCO_2/CO_2$ ratios and average $nfCO_2$ fluxes, excluding the large-source affected measurements for the aforementioned reasons.

To clarify the motivation and purpose of our analyses, we removed the text in Lines 411 - 414. Discussion of average CO_2 flux compositions obtained from our REA measurements has been moved to Sect. 4.3.3., so that Sect. 4.3.2. focuses on the proof of concept of our method, as suggested (see also Comment 44). In addition, we revised the manuscript to ensure that the term "point-source" is used appropriately, and that

brewery and district heating plant emissions are labelled as "identified/large/…
point-source emissions" if at all.

42. Section 4.3.2. There are lots of methods in this text that is supposed to be the
results. Please move the analysis methods to the methods section of the manuscript.
Focus here on the results, please.

Agreed. As we describe the identification of measurements that were potentially
influenced by emissions from the district heating plant or the brewery in Appendix F, we
removed the description in Sect 4.3.2.. We also moved the description of the ^{13}C
analysis to Appendix F (see Comment 43).

43. The discussion of the attempt to use 13C is interesting but, in the end, not used.
This is worth reporting for others who might try similar methods but perhaps this could
be an aside?

Agreed. We moved the discussion to Appendix F, where we describe the identification of
REA measurements potentially influenced by emissions from the district heating plant
and the brewery.

44. Lines 443-444. "The results thus emphasize that tall-tower measurements in urban
environments can often be affected by individual point sources."

I would argue that these results show that your measurements are doing what you hope
they will do. Of course large point sources have an impact - and you detect this impact!
- even when this is not fossil emissions and you have to use your decomposition to tell
the difference. I would change the focus of this analysis to one of verification of the
efficacy of your methodology, rather than an apology for having measurements
"contaminated" by point sources. Cities are complex. They are not homogeneous. If
we are going to make measurements in cities we must be able to deal with sources that
are mixed in space, time and type. You are demonstrating progress toward these
objectives and scientific needs. Embrace your progress and mission. You do state this
(later in this paragraph, someone buried). I suggest that you move these statements to
the fore - including the main message summarized in the abstract.

Following the advice of the Referee, we revised Sect. 4.3.2. to focus on the efficacy of
the ^{14}C-based separation of $ffCO_2$ and $nfCO_2$ fluxes. The implications for urban tall-tower
measurements in general are provided in the conclusions. To this end, we removed the
text in Lines 443 - 444 and included a statement on the benefits of these measurements
at the beginning of the paragraph and in the Abstract. Please see also our response to
Comment 41.

Agreed, we moved the paragraph to Sect. 2.3.3 in the resubmitted manuscript.

In general, we agree that a separate discussion could improve the structure of the paper. However, we found it difficult to separate the results and discussion sections without repeating ourselves and decided to keep them together. We will consider the suggestion for our next manuscript, which is currently being prepared for publication.

We agree that the manuscript would benefit from a more concise conclusion section. However, as a summary of results is too comprehensive to be included entirely in the abstract, and as not everyone will read the whole paper, we believe that providing a brief summary of the main results, answering the research questions raised in the introduction, is helpful.

Therefore, we deleted Lines 563-565 and split the original conclusions section into Summary ("1. - 3.", i.e., Lines 566 - 604 of the original manuscript) and Conclusions ("Outlook", i.e., Lines 606 - 613 and concluding sentences in Lines 588 - 598). To emphasize that the subsections refer to the three research questions raised in the introduction (Lines 55 - 61), we labeled the questions and the corresponding headers in the conclusions with Q1, Q2, Q3. In addition, we expanded the conclusion section, building on the previous "outlook" section, as suggested (see Comments 49 - 53).

It is correct that the detailed results of the z-tests mentioned in the conclusions had only been shown in Appendix G. We added the information in Line 507 in the resubmitted manuscript, but did not change Tables 5, G1, and G2 for clarity.

We agree that reducing the measurement height is a noteworthy option for increasing the concentration differences between updraft and downdraft samples, and thereby the SNR. However, it should be noted that the concentration difference will only be increased if the variance of the vertical wind velocity decreases with height, i.e., below the inertial sublayer, and if the sources are within the respective flux footprint, which will be smaller and closer to the tower. For our purpose of urban measurements at the neighborhood to city scale, lowering the measurement height was not considered as a good solution.

We added a note regarding lowering the measurement height in Lin 604 in the resubmitted manuscript. For technical details we refer to Kunz et al. (2025), where this topic and further recommendations for future measurements had already been discussed. See also Referee Comments to Kunz et al. (2025), page 3, https://egusphere.copernicus.org/preprints/2025/egusphere-2024-3175/egusphere-2024-3175-AC1-supplement.pdf).
Strengths and limitations of the ^{14}C-based partitioning approach are summarized as the answer to research question 1 (Lines 566 - 577 of the original / Sect. 5.1 of the resubmitted manuscript). Further recommendations for future investigations were included in the Conclusions.

As mentioned in response to Comment 47, Lines 579 - 587 have been retained as part of a "Summary" section. The following paragraph (Lines 588 - 598) has been extended

to address the general issue of considering point sources and mixed fluxes in cities in the "Conclusions".

51. Lines 597-598. "This again highlights the fundamental challenges of extrapolating local observations to derive emissions at the scale of an entire city."

I would not suggest direct extrapolation as an objective of urban EC flux measurements. EC measurements are well suited to testing process-based models. These models are the correct tools for extrapolation.

Agreed. We implicitly meant that process-based models should be used to extrapolate the EC measurements, but we did not state this properly. With the revised structure of the Summary / Conclusion (see Comment 47), the sentence in Lines 597-598 was no longer necessary and was deleted.

52. Lines 599-604.
This isn't compelling text for the conclusions. This restates the results section and should be fairly well understood. The take home message here, "be careful selecting your background," is reasonable but vague.

Please see our response to Comment 47. Lines 599 - 604 have been retained as part of a "Summary" section.

53. I suggest building on the "outlook" section for a sharper, targeted conclusions section of the document.

Agreed. Please see our response to Comment 47.

References:

Dröge, R., H. Denier van der Gon, Perrussel, O., David, L., Aigner, P., Kuehbacher, D., Chen, J., Hinderer, J., Brunner, D., and Constantin, L.: Final version of high-resolution city emission inventory for GHGs and co-emitted species for 2018, 2020 and 2022, ICOS ERIC - Carbon Portal, https://doi.org/10.18160/0GEF-R5XX, 2024.

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.

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

Mauder, M., Cuntz, M., Drüe, C., Graf, A., Rebmann, C., Schmid, H.P., Schmidt, M., Steinbrecher, R. 2013. A strategy for quality and uncertainty assessment of long-term eddy covariance measurements, Agricultural and Forest Meteorology, Volume 169, Pages 122-135, ISSN 0168-1923.

Pattey, E., Desjardins, R. L., and Rochette, P.: Accuracy of the relaxed eddy-accumulation technique, evaluated using CO2 flux measurements, Bound.-Lay. Meteorol., 66, 341–355, https://doi.org/10.1007/BF00712728, 1993.

Vickers, D. and L. Mahrt. 1997. Quality control and flux sampling problems for tower and aircraft data. Journal of Atmospheric and Oceanic Technology, 14: 512-526.