Reviewer 1 - Martijn Pallandt

The authors wish to thank Martijn Pallandt (Reviewer 1) for his constructive review. The reviewer's comments are below in black and our responses are in blue. Due to significant changes to the text, not all changes are detailed here, and we direct the reviewer to the full tracked-changes version of the resubmitted manuscript.

General

The authors set out to create a unique multispecies dataset of urban fluxes (comprising of CO2, CO, NOx CH4 and N20) to test a new method to partition urban GHG fluxes, specifically into stationary combustion, road transport, and biosphere components. Both this dataset and the methods are novel and of importance since there is a great need for tools to properly partition fluxes in heterogeneous environments. Overall, the writing is clear, the reasoning is sound and the manuscript follows a logical order. Two problems arise during this research. One being unrealistic low values in NOx measurements which are then scaled to local measurements. The other is that the novel partition method struggles in this high flux, high heterogeneity environment, resulting in a mixture of reasonable and unrealistic partition results. The authors are clear on these limitations and suggest linking this method to more components, flux footprints, and temporal explicit emission inventories as potential improvements. Even though the results aren't entirely as envisioned at the start, they are still valuable and serve as an important stepping stone for further research on this or similar partition methods. As an in-situ case study this research sheds a light on seasonal and daily patterns both in observed and ratios of GHG fluxes in Zurich, and is able to characterize this city as a net source of CO2, CO, NOx, CH4 and N20. It was an interesting read and I'm curious how this research further develops.

Specific

20: Cities leading these efforts if quite a statement, the references here need some work or the claim needs to be modified: the European commission report doi 404ed on my side, the stad Zurich website probably has no place here ->“ Informal or so-called "grey" literature may only be referred to if there is no alternative from the formal literature.”(and this is just a website not an archived report). A quick scan of the Lwasa executive summary seems to point out that GHG emissions will very likely increase through cities modernization construction, and further urbanization, though they note there is a need for decarbonization for cities.

The European Commission report had an incorrect DOI, this has been updated.

The city of Zurich website reference has been updated to an archived report of the city's most recent net zero 2040 progress report. This is grey literature but appropriate here as it

demonstrates that the City of Zurich has a more ambitious climate plan than exists on the national scale.

The sentence makes two claims: that cities are critical areas for emission reduction efforts, and that many cities have ambitious plans independent from / in addition to those at the national level. We have modified the sentence structure to make it clearer, and softened the latter claim. The new sentence reads:

Cities are therefore crucial for emission mitigation initiatives (European Commission, 2021; Lwasa et al. 2022), and some cities are already pursuing ambitious plans for reduction of carbon dioxide (CO_2) emissions above and beyond national-level plans (e.g., Stadt Zürich, 2024)

108: Is there a particular reason why these months were chosen? A clearer distinction between winter and summer would be observable if June, July and August were chosen instead. Related to this at line 156 /table 2: the winter period is considerably longer at 5 months vs 3 and probably with worse weather, it can be informative to indicate the distribution of (successful) samples between summer and winter.

These months were chosen as a compromise between competing deliverables and milestones of the ICOS PAUL project, namely the desire to have measurements across three pilot cities in sequence (Zurich, Paris, and Munich), to capture cold/warm months, and to fit within the limited project timeline. Therefore the instrument was first installed in Zurich in late July 2022, and needed to move to the following city (Paris) after March 2023 due to project timelines. A longer time series in one city is desirable but was not possible within the 3-year project timeline, measuring 3 cities (including installation of three towers and three campaigns). Table 2 has been updated along with reviewer 2's comments for clarity and to make the seasonal difference in data retention and total preserved 30-minute averaging periods clear.

165: It is unfortunate when equipment malfunctions, time dependent in-situ measurements as these cant be redone and you have to work with what you have. However, this section goes fairly quickly from noting that observations didn't match expectations to scaling them up. Some essential steps appear to be missing.

Was the source of the error investigated, (e.g. was it a mechanical or a calibration issue), why did it only affect NOx fluxes, and have similar problems been reported with this instrument? Without identifying the source of the issue, it is hard to justify a correction.

The NOx flux was corrected based on local measurements, but why not perform an actual recalibration/correction with calibration gasses? That would give more certainty in its accuracy than a local measurement which adds several layers of uncertainty such as transport and their measurement errors. It would also allow for additional tests on any drift, to verify that the adjusted slope is stable over time.

I assume that the throughout the paper the adjusted NOx values are used as if they were the actual measurements, and the increased uncertainty was not propagated throughout the rest of the analysis. In either case, please state clearly how this was handled.

The difficulties with the NOx measurements are frustrating and are a significant source of uncertainty that propagates through the study. We agree that these concerns were not sufficiently addressed in the initial submission. We have added additional text to the main body (S. 2.3) which details the quality control and calibration approaches that were used, namely a zeroing tank of N_2 was used to zero reactive species NO and NO_2 and a co-located flask sampler, air samples from which were analysed by the ICOS Flask and Calibration Laboratory in Jena, Germany, was used to verify good performance for species CO_2, CO, CH_4, and N_2O. Unfortunately the flask sampler is not suited to measuring NO_x and therefore the use of a background reference station was employed as part of quality control.

We suspect two contributions to the low NO_x measurements:

1) losses due to the inlet construction on the order of ~20%. The inlets used in the study are a standard part from LiCOR consisting of rain/dirt cap and approximately 1 m of steel tubing. These inlets are designed for measurements of CO2 and H2O fluxes and it is now clear that they are not desirable for NO_x.

We can estimate the losses due to the inlet system based on the subsequent measurement campaign that was conducted in Paris from January - June 2025 following re-calibration and upgrades by the manufacturer. Then testing of the instrument alongside standard NO_x analyzers. The inlet system was identical between both Zurich and Paris campaigns, as was the inlet length (33 m +/- 2 m). The instrument was run co-located with instruments operated by colleagues at the Climate and Environment Sciences Lab (LSCe) in France. Here we found a bias of ~20%, which we attribute to the steel inlet.

2) assumed errors in the spectroscopic setup and retrieval functions accounting for the remaining bias, inclusive of errors associated with the reference measurement station and atmospheric transport to the tower. As noted we are unfortunately unable to independently verify this as the problem was only identified after the end of the field campaign and after the instrument had been returned twice to the manufacturer for various upgrades and repairs.

The large correction applied to the NO_x measurements is not ideal, but we note that it is applied to 8 months of high-quality and standardised measurements collected using a Teledyne T200 chemiluminescent monitor by the Swiss Federal Laboratories for Materials Science and Technology (EMPA) and that we have not simply scaled the fluxes to reach a pre-desired value, but rather corrected the raw concentrations based on a high-quality and regularly-calibrated reference instrument, albeit not in the most perfect circumstances, and re-calculated the fluxes from these data. More information on the reference station measurement program is available in the document below (only in German):

Technical Report on National Observation Network for Air Pollutants (NABEL). https://www.empa.ch/documents/56101/246436/Technischer_Bericht_2024/0d7b63e5-70a1-4fba-ad3a-599347447a32

We have updated the text of this section to incorporate this information, and along with other requests from the reviewer, made additional explicit references to the NO_x uncertainty throughout the manuscript. We hope therefore that while the uncertainties are large, the reader is sufficiently well-informed of these.

Please see the updated section 2.3 for the updated text as well as our response to the second reviewer for individual plots relating to the quality control.

179: Why this 4x4 square and not the entire city or a larger part of the tower footprint?

The 4 x 4 km^2 was a simplification chosen to cover the majority contribution of the individual 30-minute footprints used in the analysis. Nevertheless, this 4 x 4 km^2 box covers 67% of the long-term footprint, representing the majority of the emissions measured at the tower, and also covers the most densely populated portion of the city of Zurich.

Most importantly though, the analysis presented in the paper does not make use of individual footprints: rather the box is used only to extract characteristic sector-specific emission ratios, and further subdivided to determine the variability in those ratios with direction. As the portions of the 70% to 80% ensemble footprint extend largely towards vegetated areas (SW) or similar land uses as within the box (NW, NE), expanding this box does not significantly change the ratios. We have updated Table 3 to include the ratios obtained from considering the whole city, and these fall within the directional variability within the 4 x 4 km^2 area.

Figure 3: Not essential but a 6th panel in this style with temperatures would be interesting since these would be the main drivers in differences between winter and summer.

Agreed: a 6th panel with air temperature has been added to the figure (now Fig. 4 in the new version).

301: If you can conclude based on figure 5 that the ratio can be a mixture of road transport and stationary combustion, why not a combination of respiration and road transport?

We agree that any measurements made can (and likely do) also include a respiration flux of CO_2. We are just noting here that in the winter months, the measured fluxes are well-bounded by the inventory references. We have added a sentence to this section to make it clear that a non-fossil CO_2 flux from human respiration is present in both seasons.

350 / figure 6: This explanation mainly fits the winter period. Summer 2-8 is largely indistinguishable from 6-7 when I look at the figure, and the afternoon rush seems to start at 12 already. Summer seems different from winter here.

Yes this is a fair point: this difference is mostly driven by source area and a greater prevalence of NE winds during the summer during these hours compared to the winter. We have modified this sentence to address winter and summer separately, noting this difference and reiterating the source-area contribution. Nevertheless, this species pair does not show a statistically significant difference between the seasons in aggregate.

375 /figure 7A: The nocturnal fluxes would be the yellow area around y0.7 x1.3 just outside the dashed lines? Without a temporal component to this figure, it is not directly clear which the nocturnal NOx fluxes are. Can you clarify.

Yes these fluxes, which form a mode just outside the dashed lines, are what we're referring to. The text has been modified to make this explicit.

400 and 418: What can be done about this unrealistic allocation that goes well beyond measured quantities? While the section starting around 427 discusses improvements in general, can you give specific advice on preventing this problem in future cases? Maybe adding more components or limiting allocation to know maxima?

The unrealistic allocation can occur where the measured ratios fall outside the limits defined by the reference ratios from the inventory. One could avoid this entirely by only considering measurements where all ratios fall within the defined reference limits, but this dramatically reduces the number of available measurements (and as we note, treating ratios as constants is probably not sufficient in itself). As we conclude later, the solution is likely improved (measurement-driven) and time-resolved reference ratios.

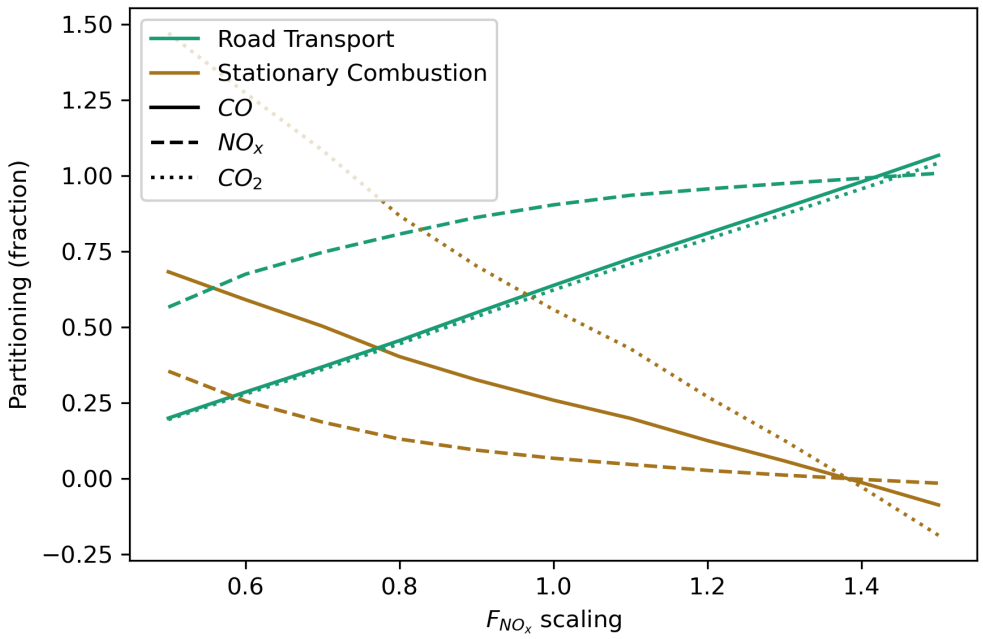
453-455: If I followed the equations in section 2.5 correctly al equations for the final partition components include the NOx term directly or indirectly. Therefore, wouldn't this affect all ratios?

Correct, all partitioning is ultimately affected by the NO_x ratios (or more precisely: the relationship between measured NO_x ratios and the reference ratio definitions). We have modified this sentence to make it explicit that elevated NO_x may lead to over-attribution to road transport of all species.

Did you at some point test the sensitivity of this model to FNOx? Even if you are not certain of your measurements it would give an indication of the impact of such errors.

The sensitivity to FNO_x was tested by scaling the NO_x fluxes and running the model for a series of linear scalings from 0.5 to 1.5 where 1.0 represents the fluxes used through the

manuscript, i.e., those calculated from concentrations scaled against the reference station. The results are presented in the figure below.



As expected the model is sensitive to the NOx inputs as the partitioning of the net flux to source categories is essentially determined by the measured FNO_x:FCO along with the reference ratios, and these fluxes are of a similar magnitude.

456 - 465: I am somewhat surprised to find new methods and results after the discussion of the model's limitations. And it is not clear to me what you have done here. These two paragraphs and table 6 can benefit greatly from a rewrite to clarify the method and intent. I would advise to add a subsection to the methods section describing this sensitivity analysis in a bit more detail, and discuss these results probably before ~427

Specifically:

1. Table 4 nor its description make mention of a and b. what is exactly being combined here?
2. In the section on the linear mixing model no mention is made of the mmol mol resolution, in which way is it different here?
3. You are testing the sensitivity of the linear mixing model to changes in what exactly? The discussion in the second paragraph also doesn't make clear what sensitivity has been tested here. A typical sensitivity analysis tests a range of values of certain parameters.
4. Continuing with table 6: The first '% of total' refers to table 3 where you have 5 categories and the second 'relative' to only the two listed in this table 6 (sc and rt)? Please clarify.

5. And the model outputs should be compared to the relative column since the model only uses the sc and rt categories? If the model outputs are per wind direction, it doesn't make much sense to compare them to the general reference inventories, wouldn't it be more interesting then to have 5 relative inventory references: all and the 4 directions?
6. In line 496 it is mentioned the sensitivity to reference inputs was tested here?

We agree that this section is more suitable before the assumptions of the model limitations, and we have reorganised/rewritten to try and better clarify what we have done. In response to specific points:

1. The parameters a and b are defined according to Eq. 4-9 and the reference values used from the inventory are those found in Table 4. We have updated Table 4 to make this definition explicit, where before it was implicit.
2. This was poor wording: rather the model was run for every combination of *a* and *b*, within the range found in Table 4, at increments of 0.05 mmol mol-1 (0.1 was from a previous run but provided insufficient granularity for some ratios with smaller ranges, this has been updated). I.e., a set of input ratios was made for a_rt: [3.00, 3.05, 3.10, 3.15 … 4.40], and for a_sc: [0.95, 1.00, 1.05 … 1.40], and the same for b_rt and b_sc. All possible combinations of the inputs were tested and the distribution of model outputs statistically analysed in this section. The wording has been updated to try and make this clearer.
3. Our response to (2) should hopefully make this clear.
4. Correct; we have updated the caption and text to make this more explicit.
5. The relative column is probably more relevant, yes, as it incorporates the assumptions of only 2 combustive source categories. The second part of the question we hope has been addressed by our response to (2).
6. See response to section (2).

Please see the tracked changes version to see all specific changes made to the text.

468 That would be a valuable continuation of this research, though if you continue with this dataset really aim to get a proper recalibration of the NOx data.

Unfortunately this is not possible, as noted above.

477-489: Since not everyone might look at the methods (in detail) a disclaimer on NOx uncertainty in the second or third paragraph is appropriate.

A disclaimer / note on the NO_x correction and associated uncertainty has been added to this section.

466: "complex and heterogeneous urban environment may ultimately pose too great a challenge for the application of such a model with fixed emission factors over long periods

of time and large flux footprints.” And **501**: “ however in the complex and heterogeneous urban environment this information is difficult to exploit on its own.”

This is unfortunate, since these are the environments where flux partitioning is especially important. In a homogeneous environment partitioning of fluxes is of lesser importance. Hopefully next steps will improve on this method further.

Indeed, and the following work on this dataset will try to address this problem. The partitioning model developed here essentially asks the question: Can we use measured ratios, reference ratios, and some simplifying assumptions to partition net fluxes even in a complex environment? The results provide a springboard to help us in the next stages to target areas where we can reduce or eliminate the need for certain assumptions as the partitioning becomes more complex to match the real urban environment.

Technical

~427: not essential but you could put a section break here with everything after ~427 a discussion of the assumptions and where they are met/failed.

After rearrangement of the text in this section we think the progression is more logical and that the discussion flows better without a section break.

483: tower..

Fixed.