Review of Ponomarev et al. (2025)

This study presents an investigation of urban CO_2 fluxes in Zurich and Paris through inverse modeling. It is interesting that the same modeling framework is applied to two different cities with different sizes, emission distributions and set-ups of their measurement network. However, I have some concerns regarding the set-up of the inversions, which are mostly mentioned in the paper, but not addressed in any type of sensitivity tests. I appreciate that the study covers a lot of ground, making it hard to dig deep everywhere. However, in the absence of sensitivity tests, the final emission estimates need to be presented more carefully. They should be accompanied by a more comprehensive discussion of the limitations and uncertainties, with suggestions on how these can be addressed in future work.

My primary concerns:

1. I think the authors undersell how complex their inversion set-up is compared to previous urban work. The system co-optimizes anthropogenic emissions *and* GPP *and* RE *and* boundary conditions. Previous work has shown that co-optimizing anthropogenic emissions and NEE with only CO_2 observations is already difficult, if not situationally impossible (e.g., Sargent et al., 2018). This is one of the key difficulties in urban inversions. Yet, in most of the manuscript, the authors consider that all these components can be separately estimated, which is something that, to my knowledge, has never been shown before. It would therefore be important that the authors address how good their estimates are.
   The authors do address that the biosphere fluxes are likely too low in the prior, which makes co-optimization easier. Then, however, I think it would be an easy sensitivity test to run the inversion with biosphere prior x2 or x5. Additionally, given the inversion method, the authors can investigate the posterior covariances to see if estimation of GPP, RE and anthropogenic emissions are correlated. The current manuscript does not include analyses of this novel aspect of the study. I would really like to see some analysis of posterior covariances between flux categories, which is possible without additional inversions.
   As a specific example, I find it hard to believe that separate estimation of RE and GPP is possible, and that the interpretation of these two separately is meaningful (L490-493), especially with only afternoon observations. My hypothesis would be that the posterior covariances will indicate that the solutions for RE and GPP are correlated, and if this is not the case I would be very interested in the authors' interpretation of this result.

2. With the lack of sensitivity tests, the authors rely on the posterior error statistics. While these have value, they are very optimistic as absolute uncertainties. I think the authors should be much more explicit that these statistics provide a limited

view of the real uncertainties, which generally require sensitivity tests to be quantified. Reading the abstract would give the reader the (in my view false) impression that there are tiny posterior uncertainties and the system works incredibly well. In reality, many major uncertainties are not included in the current uncertainty estimate. In fact, my interpretation is that for Zurich the authors find that in winter the estimate is biased due to transport errors, and in summer the estimate is biased due to interference from the biosphere. Lian et al. (2023) also report a 3x higher posterior uncertainty than this work (Table 3), which is not really discussed in the manuscript.

In my view, the conclusions and abstract should reflect these concerns with respect to these final emission estimates, which the authors themselves raise. I appreciate this study as a first step that covers a lot of ground, but, based on the presented analysis, I do not think that the Zurich emission estimatecan be trusted within the currently stated posterior uncertainties.

3. Given that the authors present an impressive amount of work, I would not suggest additional inversions to be included. But I would like to see a more start-to-end discussion of what the major limitations of the current framework are, and, most importantly, how (your) future work can address these. A non-exhaustive list:
    a. The influence of the biosphere, which is discussed as being too low. As mentioned, there are some simple sensitivity tests you can do by scaling up the prior biosphere fluxes. Additionally, there is an urban version of VPRM that includes, e.g., surface porosity and the urban heat island effect, likely to result in higher NEE as well (e.g., Hardiman et al., 2017).
    b. Are there any potential steps forward you can take in separating biosphere and anthropogenic fluxes?
    c. The problematic winter observations that are not well-simulated could be filtered out. Additionally, one could consider including filtered nighttime observations (Monteiro et al., 2024), although I understand that's quite a new development. E.g., in Lian et al., 2023, filtered morning observations are included.
    d. Given the limitations of EnKF posterior uncertainties, sensitivity tests can help quantify the uncertainties better.
    e. Any key lessons on comparing emission estimates for different cities? To my knowledge this is the first urban inversion that includes more than one city.

Other points:

- The introduction doesn't explain previous literature in enough detail. For example, what specifically does previous work say about separating anthropogenic emissions from biosphere fluxes, beyond that it is difficult? Given that Paris has

quite an extensive history of high-quality urban inversion work, I would also expect some more specific discussion beyond that your posterior totals agree. Generally, I would expand the introduction with specific, relevant findings from previous urban studies. It also meanders into too much detail in L44-48, which is more fitting for the methods (see next point).

- The manuscript is very long. At times this is necessary, but not everywhere. I think that some adjustments can strengthen the presentation of the most important results. Some suggestions:
    - Are all 15 figures required to be in the main text? Fig. 10, 13 could be supplementary in my view. Possibly others at discretion of the authors.
    - Much of the discussion of the results outside Zurich and Paris (i.e., the European domain) could be supplementary, with a few sentences addressing this in the main text.
    - L254-268: Given that off-diagonal errors for R are not included, they do not need to be explained in this much detail.
    - L281-286: This can be 1 or 2 sentences.
    - L311-315: This is just aggregation of uncertainties, I don't think it needs to be explained.
    - Conclusions are very verbose.

  These are non-exhaustive, I think it would help the manuscript's readability if it were more concise.

- It is common practice to include only afternoon observations in CO_2 inversions. However, since the overpass time of air over a city is short, afternoon observations only inform on afternoon fluxes, and perhaps late morning. I would like to see a brief discussion of the potential impact on flux estimation, given that it can depend on a city's size and diurnal cycle in fluxes. E.g., do you completely miss the morning rush hour? How high are nighttime fluxes (anthropogenic and biosphere), and what is their contribution to the CO_2 budget? What are the consequences of scaling the non-afternoon fluxes along with the afternoon fluxes, when only the latter are constrained by observations?

Minor points:

- L20-22: While it is true that there are not that many urban inversions, there are a lot more than these 4, so I suggest to add: "for example".
- L22-23: "*cities are major contributors"* Please add reference and possibly an estimated number (e.g., 44% from Seto et al., 2014).
- L44-48: This is too detailed for the introduction, it is more fitting for methods.

- Why is the prior error for anthropogenic emissions higher (100%) than for biosphere fluxes (50%)? It is repeatedly mentioned how uncertain biosphere fluxes are.
- Fig. 2: Maybe mark resolution with e.g., 1 km x 1 km, so it is clear that it is a resolution. Also, I am not sure what the dashed lines in the inner figure mark without x/y labels, so I suggest removing them. Given that no coordinates are given, a scale in the inner zooms would help: it's easier to interpret than the area number, and useful for comparing to the spatial correlation scale.
- What is the temporal resolution of scaling factors in the state? Is it one per week?
- Why are the errors in R smoothed over 5 weeks? How does that affect the impact of anomalously high model-data differences in winter that are currently affecting the inversion result quite a bit?
- L266: How much is the impact? I would expect that correlations between sites reduce the information in the system by a lot, so I am surprised by this result.
- L339: Mention that it's Pearson r at first mention of the correlation, no need to mention the python function.
- L380-381: How so? Is the next paragraph describing this unrealistic behavior? Could be more explicit.
- L385-386: I do not follow this logic. If anthropogenic emissions and biosphere were interchangeable, then the inversion would stick to the prior, since that gives a lower cost function. Therefore, the fact that it is adjusting the two in opposite directions means that they are to some extent separable. The question is: is this real or does it have to do with e.g., overfitting?
- L424: Suggest Paris instead of French capital.
- Fig. 12c: These local reductions are huge, scaling down certain grid cells by >50%, and that is on the annual mean where the text mentions it mostly comes from winter. Is there any reason to believe these reductions are realistic or does it only have to do with transport errors? Does it fully disappear in summer?
- Fig. 14: Expand description. E.g., RMSE/bias between simulated and actual observations.
- I would like to see a supplementary figure with spatial distributions of anthropogenic emissions, GPP and RE and their diurnal cycle in the supplements, given that this is the information that the system allows you to separate the three (L39-40). Partly this information is already in existing figures, but not everything (e.g., diurnal cycles, or GPP and RE split) and not in a way that makes it easy to compare.

References

1. Seto, K.C., Dhakal, S., Bigio, A., Blanco, H., Carlo Delgado, G., Dewar, D., Huang, L., Inaba, A., Kansal, A., Lwasa, S. and McMahon, J., 2014. Human settlements, infrastructure, and spatial planning.

2. Hardiman, B.S., Wang, J.A., Hutyra, L.R., Gately, C.K., Getson, J.M. and Friedl, M.A., 2017. Accounting for urban biogenic fluxes in regional carbon budgets. Science of the Total Environment, 592, pp.366-372.
3. Sargent, M., Barrera, Y., Nehrkorn, T., Hutyra, L.R., Gately, C.K., Jones, T., McKain, K., Sweeney, C., Hegarty, J., Hardiman, B. and Wang, J.A., 2018. Anthropogenic and biogenic CO2 fluxes in the Boston urban region. Proceedings of the National Academy of Sciences, 115(29), pp.7491-7496.
4. Lian, J., Lauvaux, T., Utard, H., Bréon, F.M., Broquet, G., Ramonet, M., Laurent, O., Albarus, I., Chariot, M., Kotthaus, S. and Haeffelin, M., 2023. Can we use atmospheric CO_2 measurements to verify emission trends reported by cities? Lessons from a 6-year atmospheric inversion over Paris. Atmospheric Chemistry and Physics, 23(15), pp.8823-8835.
5. Monteiro, V.C., Turnbull, J.C., Miles, N.L., Davis, K.J., Barkley, Z. and Deng, A., 2024. Assimilating morning, evening, and nighttime greenhouse gas observations in atmospheric inversions. Journal of Geophysical Research: Atmospheres, 129(17), p.e2024JD040998.