Review comments of Pilz et al. (2026)

In the present study, the authors describe their Bayesian approach to constrain urban scale CO2 emissions in the city of Berlin using 2018 aircraft campaign observations. They explicitly solve one of the major inverse problems – background concentration estimation. The results suggest that the proposed method performs well for the studied campaign. It gives a flux estimate that agrees better with emission inventory data and has lower uncertainty in comparison to a standard but more simplistic mass balance approach. The optimized background concentrations are verified using independent concentration measurements above the boundary layer.

The manuscript is well structured and written in a clear, plain language. Overall, the work is in the journal scope and provides a relevant contribution to urban scale flux estimation methods. The major concern is a relatively small amount of data used for inverse experiments and to draw the conclusions about method performance. The authors should discuss this limitation and provide additional justification for why their conclusions would hold when more data is used. The study is worth publishing after minor revisions and clarifications about the described method.

We thank Referee #1 for their positive assessment of our manuscript and for recognizing the relevance of our methodological contribution.

We acknowledge the reviewer's concern regarding the limited observational dataset. Indeed, aircraft campaigns over urban areas remain relatively rare, and not all meteorological conditions are suitable for inverse modelling, as a sufficiently accurate representation of atmospheric transport is a prerequisite for reliable flux estimates. This limits the number of observations that can meaningfully be included in the inversion. While we cannot solve the lack of data, in the revised manuscript, we more explicitly acknowledge the limitations associated with the sparse observational dataset and clarify that the primary contribution of our study is methodological.

In particular, we acknowledge that the limited dataset does not allow us to establish the general performance of the approach across different meteorological conditions and urban environments, but the proposed framework can be applied and further evaluated in future studies using more extensive observational datasets. We therefore consider our approach a valuable complement to established mass-balance methods.

Specific comments

Line 8: 'The emissions rate is about 48% higher than the prior total annual emissions …'

I would suggest explicitly stating the emission rate over the given day, since CO2 fluxes have significant interannual variations.

> Since our information about daily anthropogenic emissions rates only comes from TNO scaling factors (month-in-year, day-in-week, hour-in-day), we chose to not use daily fluxes in order not to give the reader a false sense of accuracy. We are hereby following the convention of Klausner (2020).

Line 15: 'Urban areas account for the majority of anthropogenic CO2 emissions …'

Please add relevant references.

> We have added IPCC, 2022 and Crippa et al., 2021 as references for this statement to the manuscript. These studies highlight the significant contribution of urban areas to global anthropogenic CO_2 emissions.

Lines 38-41: 'However, the spatial representativeness of the networks is constrained by the limited number of stations and their proximity to heterogeneous near surface sources within the urban canopy, which can strongly bias inferred city-wide fluxes and if not carefully treated can pose issues, as they are not able to directly and homogeneously sample the urban plume downwind of the city (Li et al., 2026).'

This sentence is quite long and contains several statements. I would suggest splitting it into several sentences. I agree with the statement that airborne observations can provide better spatial coverage in comparison to ground-based in-situ networks. However, the argument about proximity to heterogeneous near-surface sources biasing inferred fluxes is not clearly described. It would only hold if we are optimizing total city emissions and do not allow the inversion to correct intra-urban distribution. Otherwise, the stronger CO2 signal in in-situ observations should help us to better constrain those sources and their spatial gradients. I think the problem is rather related to atmospheric transport errors inside the canopy layer and the inability of the models to capture the air flow (plume direction, vertical mixing, wind speed etc.). However, this does not diminish the value of in-situ measurements. Rather, it incentivizes researchers to improve atmospheric models.

> We agree that splitting the different arguments within this sentence helps to make the points more clearly. We now address the typically cost-driven limits of in-situ networks explicitly. Additionally, we point towards the modeling of transport as the main source of uncertainty

introducing biases for near surface sensors within the urban canopy. We adjusted the manuscript as follows: “However, constraints in network cost often require a trade-off between quality and number of sensors either introducing random errors and biases or limiting the spatial representativeness of the network (Turner et al., 2016). Additionally, modeling the transport of heterogeneous near-surface sources within the urban canopy to the sensors in their proximity is prone to errors, which can strongly bias inferred city-wide fluxes and if not carefully treated (Li et al., 2026). “

Lines 57-62: This is an important paragraph and could be expanded. I would suggest the authors mention that there is a difference in background definition in Eulerian and Lagrangian models. Eulerian transport models produce background concentration directly, since it can be transported explicitly as a separate tracer forced only by initial and boundary conditions. In contrast, Lagrangian particle dispersion models output only a sensitivity (footprint) field relating flux to concentration enhancement. In the latter case, the background concentration is not a native model output and must instead be assigned externally, e.g. by sampling a separate concentration field at the trajectories’ point of origin. Since this study uses a Lagrangian model (FLEXPART-WRF) for footprint calculation, it is important to make this distinction explicit, and to discuss the uncertainty arising from background definition in Lagrangian models and how it compares to other sources of errors as well as the emission signal. I would also suggest citing other studies and providing their estimates, e.g. Karion et al. (2021) find summer monthly-mean scatter of 1 to 3 ppm between different background definition methods, with hourly RMSE reaching up to 8 ppm. These errors, at least partially, represent an additional source of background uncertainty on top of errors coming from boundary condition products and model errors.

Yes, the treatment of background concentrations in Eulerian and Lagrangian transport models is an important distinction. In particular, as our approach combines the Langragian approach derived from FLEXPART-WRF with the background concentration fields provided by the Eulerian WRF-Chem simulations and explicitly accounts for background uncertainty within the Bayesian inversion. We have therefore expanded the introduction to explain how background concentrations are represented in Eulerian models and how they are commonly determined in Lagrangian inversion systems.

Further, we have added a discussion of the uncertainties associated with different background definitions, drawing on the findings of Karion et al. (2021). We also emphasis the important of the background treatment in the revised manuscript.

The additional paragraph reads:
“The treatment of background concentrations in regional and urban inversions depends on whether an Eulerian or a Lagrangian model is used. Eulerian models simulate the background directly as a tracer that is forced only by initial and lateral boundary conditions and transported

with the concentration enhancements generated by surface fluxes within the modelling domain. In contrast, Lagrangian particle dispersion models like STILT or FLEXPART, which underlie many inversion systems (e.g. InTEM, ELRIS or Jena CarboScope-Regional; Rödenbeck et al., 2015; Henne et al., 2016; Manning et al., 2021), provide footprints which describe the source–receptor sensitivity, relating concentrations at the measurement location to the underlying surface fluxes. Background concentrations must therefore be determined separately. These backgrounds are usually derived from measurements at upwind or remote background sites or determined by sampling a three-dimensional concentration field from a larger model at the spatio-temporal coordinates where the particles leave the domain (Karion et al., 2021).

All these options are accompanied by their own uncertainties and errors. Karion et al. (2021) compared observation- and model-based backgrounds for the Washington DC and Baltimore tower network. For CO_2 they found summer monthly-mean biases of up to 2 ppm with a scatter of 1 to 3 ppm, and hourly root-mean-square errors (RMSE) ranging from 1 to 8 ppm. Since these values can be comparable to modelled in-domain concentration enhancements, the choice of background estimation method can substantially affect the inferred fluxes and their uncertainties. This motivates the explicit treatment of background uncertainty within the inversion framework."

Lines 85 – 87: 'Flights on the other days are not considered, as the urban plume was either not captured or the boundary layer was not clearly separated (18th, 23rd, 24th, 25th July).'

Most of the days were discarded, do the authors expect the limited number of usable flight days to be a common limitation for similar campaigns? Could the authors discuss how flight planning could be optimized to maximize the number of observations suitable for inversion, and for the future work, how could one estimate the number days required to constrain fluxes monthly/seasonally?

We agree that determining the number of flight days required to constrain monthly or seasonal emissions is an important question that will depend strongly on the variability of the emissions source and the uncertainty of the emission estimate. Answering this question would require many assumptions and a dedicated analysis e.g. as a Observing System Simulation Experiment (OSSE), which is outside the scope of this study.

However, we have expanded the discussion on flight planning and refer to Klausner et al. (2020) for further details on the original campaign planning and its limitations.

Line 96: Why were the MYJ and the YSU boundary-layer schemes selected?

These two boundary layer schemes have been used in the MACRO-2018 dataset. They were selected due to their good performance over urban areas. For more information about MACRO-2018 and its genesis, please refer to Pilz (2026).

Lines 99-100: The VPRM setup needs a clarification. Could you include the following information: Which VPRM code version was used? What input products for satellite indices (MODIS or Sentinel-2), temperature, radiation, land cover data did you use? Also, what is the spatio-temporal resolution of the resulting fluxes?

We have improved the description of the VPRM product and have added the following to the manuscript: "The VPRM emissions were calculated using pyVPRM by Glauch et al. (2025) and have an hourly resolution of 500 m. They use the EVI and LSWI 1-day MODIS products from the Terra and Aqua satellites and the 2 m temperature and downwelling shortwave radiation product from ERA5-Land (C3S, 2019). As land-use data, they use the 100 m resolved Copernicus Dynamic Land Cover Collection 3 (Buchhorn et al., 2020). Further information on the VPRM product can be found in Pilz (2026)."

Line 103: Why did you decide to use that specific CAMS experiment? Have you tried using different CAMS products?

The CAMS experiment was not specifically selected for the study but was provided as part of the MACRO-2018 dataset. Using this background, therefore, ensures consistency with the existing WRF-Chem simulations, in which the CAMS fields were used as initial and lateral boundary conditions. Also note that the CAMS background represents the large-scale atmospheric CO_2 distribution, including contributions from sources and sinks outside the European modelling domain. The regional concentration enhancements are then simulated separately using WRF-Chem.

We therefore did not test alternative CAMS products. However, an important aspect of our Bayesian approach is that the background concentrations are explicitly optimized rather than prescribed by the CAMS product alone. Furthermore, we independently evaluate the optimized background concentrations against aircraft observations above the boundary layer, providing an additional assessment of their reliability.

Lines 112-115: 'Because we also perform a background estimation in parallel to the emissions estimation, we need prior background concentration information for all measurements. These prior background concentrations are provided by MACRO-2018, which includes a first estimate thereof using the previously mentioned background concentration field from outside the Europe domain.'

Please provide a description of how the background concentration is sampled for each observation here in the methods section or in the appendix.

The Eulerian background concentration provided by MACRO-2018 (CO2_BCK) corresponds to the forward transported background fields provided by CAMS using WRF-Chem as a transport model and is available for every grid cell and time step. We then simply linearly interpolated onto the space-time coordinates of the measurements. We have added "We sample MACRO-2018's Eulerian CO2_BCK concentration field at the spatio-temporal coordinates of the aircraft measurements." to the manuscript.

Line 125: Did you consider re-scaling annual totals to correspond to the official self-reported Berlin emissions?

We did not rescale the emissions to match the officially reported Berlin emissions. The TNO inventory provides sector-specific temporal profiles, including hourly, weekly, and monthly scaling factors for individual GNFR sectors, whereas the officially reported emissions are only available as annual totals. Rescaling the inventory to match these totals would therefore require additional assumptions regarding the sectoral distribution of emissions and their temporal variability, which we are reluctant to introduce.

Instead, we acknowledge that the limited observational coverage of our campaign does not allow us to independently assess the representativeness of the inferred emissions for the entire year. In our Results section, you can find the sentence: "Given the short-term nature of the aircraft observations, the resulting estimate is not representative of the full annual mean, which complicates comparisons with annual inventories such as EDGAR, CAMS-REG and the Berlin inventory.".

Line 137: How was the uncertainty estimate of 6.8 $MtCO_2a^{-1}$ derived? Is it just 75% of the prior that you describe in the Section 2.5.1?

Yes, it is 75% of the prior. However, we have removed the uncertainty from this section in the manuscript where uncertainty was not discussed yet and have moved it to the end of Section 2.5.1. There, we have added the clarifying sentence "This corresponds to a prior emissions estimate of 9.1 ± 6.8 MtCO2a−1."

Line 157: Why did you decide to also include separate regions 3 and 4 for Germany upwind and downwind? Does it have an advantage over having only three regions 0 – 2, i.e. by increasing the size of Berlin upwind and downwind to cover 3 and 4 regions, correspondingly?

We chose our inversion set-up such that it can distinguish between the Berlin upwind and downwind regions (regions 1 and 2) and the corresponding larger-scale German regions (regions 3 and 4) to allow the inversion to adjust local and regional emission contributions

independently. This distinction is particularly relevant in the context of our background optimization, which addresses only the large-scale and smoothed background originating outside the European modelling domain.

Combining regions 1 and 3, and regions 2 and 4, respectively, would impose the same emission scaling factor on sources close to Berlin and sources farther away, whereas separate scaling factors allows the inversion to exploit this spatial information and account for potential discrepancies in regional emissions without necessarily attributing them to either emissions in the vicinity of Berlin or the large-scale background.

Therefore, the choice of state vector elements is a compromise between sufficient flexibility to represent different emission contributions and limiting the number of parameters to be optimized due to limited number of measurements.

We have added the following explanation of this to the manuscript: "We separate the upwind and downwind regions into the Berlin and Germany upwind and downwind parts in order to give the inversion system the possibility to adjust the emissions between the background (coming from outside the Europe domain in MACRO-2018) and the edge of the FLEXPART footprints independently from local emissions."

Line 170: Does 'CO2_BCKfield' include the forcing from Berlin emissions?

The CO2_BCK field of MACRO-2018 does not directly include the forcing from Berlin emissions. It represents the background tracer initialized and driven by CAMS concentrations at the boundaries of the European MACRO-2018 domain (Pilz, 2026), while concentration enhancements from surface fluxes within the domain are simulated separately. However, since CAMS is a global atmospheric CO_2 simulation that includes anthropogenic emissions and assimilates observations, the CO2 fields used here as background may in principle contain a contribution from Berlin emissions. Given the distance between Berlin and the boundaries of the European modelling domain, this contribution is expected to be small and spatially diffuse due to atmospheric transport and mixing.

Lines 187-194: Please provide the formula you used for computing the gain and average kernel matrices, and comment on how it follows from this formula that values outside [-1, 1] are considered non-physical.

The formulas we use are from Rodgers (2000) and are added to the manuscript now:

$$\mathbf{G}_y = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \mathbf{K}^T \mathbf{S}_\epsilon^{-1}$$

$$\mathbf{A} = \mathbf{G}_y \mathbf{K} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K}$$

We have removed the statement regarding non-physical behavior, as the reviewer is correct that this interpretation is not generally valid. This does not affect the results or conclusions of the manuscript. Consequently, we have also removed the corresponding argument regarding non-physical behavior in Lines 208–209.

We now simply state that we use a prior uncertainty of 75%. As we show the sensitivity of the inversion results to different prior uncertainties in Fig. A1, the reader can evaluate the implications of this choice.

Lines 236-237: 'The model reproduces the general structure well up to about 2km but underestimates and smooths the mixing-layer height, likely due to limited vertical resolution (cf. Table 1).'

The wind performance seems to be the same between both altitude ranges, I suggest mentioning this in the text. Also, based on the wind profiles (Figure 2), it seems like the vertical structure is quite different between the schemes. Could you please provide possible reasons for this difference.

We have added a reference to the similar wind performance between the two models and altitude ranges to the text. This paragraph now includes the sentences: "Both model configurations have similar fidelity in reproducing the winds, and wind velocity performance is similar between the two altitude ranges. However, their difference in vertical structure underscores the importance of using multiple model setups to get a sense of modeling variability (cf. Fig. 2)".

Regarding the reason for the vertical structure, this is based in the difference in PBL parametrization between the two schemes. While more detailed information is available in Pilz, 2026 the largest difference is that the MYJ scheme is a local TKE closure scheme while the YSU scheme is a non-local, first-order (K-profile) closure scheme. These are known to have different vertical structures and biases – e.g. Hu et al. (2010) found that the MYJ scheme produced lower vertical mixing, which can cause the larger wind velocities near the PBL top which are visible in Fig. 2.

Lines 272-273: 'This is achieved without increasing its correlation to the observations (prior and posterior correlation both 0.46).'

You mention this statement several times in the text, could you please provide a few comments in the text on why it is important to keep the correlation between measurements and background unchanged. The background concentration can also include plumes of elevated concentration in it because CO2 is a near-conservative tracer on synoptic scales, so the enhancements can be advected from outside of the model domain as well. How do you make sure those data are not used for the inversions?

The unchanged correlation between the background concentrations and the observations provides an indication that the optimization does not artificially attribute the observed urban plume enhancements to the background.

We agree that background concentrations can contain enhancements originating from emissions outside the modelling domain, particularly because CO_2 is a relatively long-lived tracer. However, our background optimization specifically addresses the contribution from the edge of the FLEXPART footprints outwards. This edge is situated 200-300 km away from the region of interest. At these large spatial scales, we expect individual emission plumes to be substantially mixed and smoothed by atmospheric transport, resulting in relatively smooth background concentration fields. We therefore do not expect this contribution to exhibit the same spatial patterns as the local urban plume.

Relatively smooth background conditions are nevertheless an important prerequisite for our approach, as strong background gradients or distinct transported plumes could introduce ambiguities between background variability and local emissions. We verified that the selected flights exhibited sufficiently smooth background conditions for the inversion. This can be checked in Fig. 3c) where we show the prior background concentrations.

Importantly, emission contributions originating within the European modelling domain, including those from other regions in Germany, are treated separately in our inversion through the optimization of domains 3 and 4. This allows us to account for regional emission contributions without attributing them to the large-scale background. We have clarified these distinctions in the revised manuscript and expanded the discussion of the assumptions underlying the background optimization.

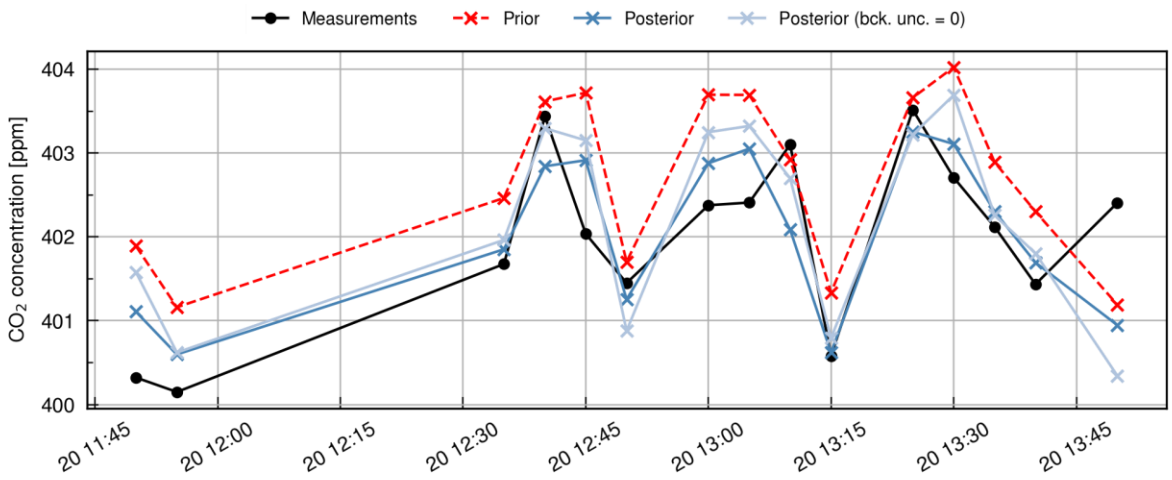
Lines 292-293: Why are the gain matrix values for cell 1 significantly smaller than for cell 0? Whereas for the diagonal elements of the average kernel matrix they are much closer to each other. Also, what is the meaning of the sign of the gain matrix values? Why does the sign alternate throughout the day?

The gain and the averaging kernel are two different quantities. On the one hand, the gain gives the flux change per unit of model–data mismatch, whereas the averaging kernel on the other hand is the gain multiplied by the sensitivity (A = GF) and is dimensionless. Cell 1 is a much larger region, so the measurements are considerably more sensitive to it. Therefore, the conversion from gain to averaging kernel involves a larger factor. However, because cell 1 also has a smaller prior uncertainty both cells end up with similar averaging-kernel values while the gain for cell 1 remains smaller. In short, the observations constrain both cells to a similar degree relative to their prior uncertainty, but cell 1 needs a smaller flux change to explain the same share of a given mismatch.

Negative gain values are not unusual. They result from the measurements seeing the city and its surroundings in varying proportions. A measurement has a negative gain for the city when its share of sensitivity to the city, relative to the surroundings, is below the information-weighted average across all measurements. Such a measurement mainly informs about the surroundings: a positive mismatch at that measurement raises the estimate for cell 1, which in turn explains part of the enhancement in the city-sensitive measurements, so the estimate for the city is lowered. As the share of city sensitivity changes over the day with wind direction, boundary-layer height and flight pattern, the sign of the gain changes accordingly. As this behavior is not unusual, but requires an excursion to inversion theory to properly explain, we have decided to not expand this discussion in the revised manuscript.

Figure 4. It would be good to include a similar figure, but that also includes posterior estimates when the background was not optimized, to directly show what the impacts are on the concentration timeseries.

We thank the reviewer for their suggestion and have added an additional trace for the posterior with the background uncertainty of zero to Fig. 5. The comparison shows that jointly optimizing the background and emission contributions generally improves the agreement between simulated and observed CO_2 concentrations (without bg optimization: RMSE 0.846 ppm, correlation 0.704; with bg optimization: RMSE 0.641 ppm, correlation 0.787). This illustrates the benefits of accounting for background uncertainty.



Lines 323-324: Please add comments on how you think the inversion is able to effectively separate the background and emission signals from observations. Is it because some parts of the observation were unaffected by local sources and were correlating only with background adjustments?

The separation of background and emission contributions becomes possible because of the distinct spatial sensitivities of the observations to the individual state-vector elements. In particular, the observations are collected at different locations, both inside and outside the urban plume. Their footprints, i.e. their sensitivities to surface emissions, differ for each location. Consequently, the individual observations respond differently to changes in the various emission sectors and regions represented in the state vector. The Bayesian inversion exploits these differences in sensitivity to jointly constrain the background concentrations and the emission contributions. We have added "Since the measurement locations where the footprints are spawned are both upwind and downwind of the city as well as inside and outside of the urban plume, the difference between the footprints allows the inversion system to separate urban from peri-urban fluxes." to the revised manuscript.

Lines 339-341 and 362: 'This substantial uncertainty reduction highlights the ability of the Bayesian Inversion framework with explicit background estimation to more efficiently extract the available information from the aircraft observations.'

'We reduce the flux uncertainty by around a factor of five compared to the one in Klausner et al.

(2020)…'

How realistic do you think the posterior uncertainty estimate is? Can it be decreased by the inversion too much? Are there any other sources of uncertainty that could not be included in your study? Please add some comments to the Discussions section.

> We agree, of course, that this is not a final objective truth, but rather an estimate subject to assumptions and methodological uncertainties (Ganesan et al., 2014, Pitt et al., 2022). To the paragraph presenting the posterior uncertainty estimate, we have added "However, as with all Bayesian Inversion methods, our posterior uncertainty estimate is not only a function of the data and measurements but also of the prior uncertainty assumptions (Ganesan et al., 2014, Pitt et al., 2022). While we have aimed to base the uncertainty assignment on transparent and data-driven assumptions (cf. Secs. 2.5.1 and 3.2), the posterior uncertainty remains an estimate subject to these underlying assumptions. In particular, the posterior uncertainty may be underestimated if systematic errors or non-Gaussian errors occur. Finally, uncertainties could be further reduced by averaging across multiple flights as shown by Pitt et al. (2022). These findings highlight that the posterior uncertainty derived from a single inversion may not fully capture uncertainties associated with alternative modelling configurations."

Technical corrections

Table 1. Please provide variable name descriptions in the caption.

> We have changed this as requested. The caption now reads: "Mean absolute bias comparison of MACRO-2018 variables to flight measurements for the July 20th 2018 vertical profile. With θ for potential temperature and RH for relative humidity. Comparison is separated into values below and above 2km altitude. "

Figure 2. The difference in wind speed profiles between the two model schemes is quite significant. Could you please change zorder on the profile plots and bring the smooth model lines to the front, otherwise it is difficult to see them behind fluctuating observations.

We have updated Figure 2 as suggested and agree with the reviewer that this is a much clearer presentation.

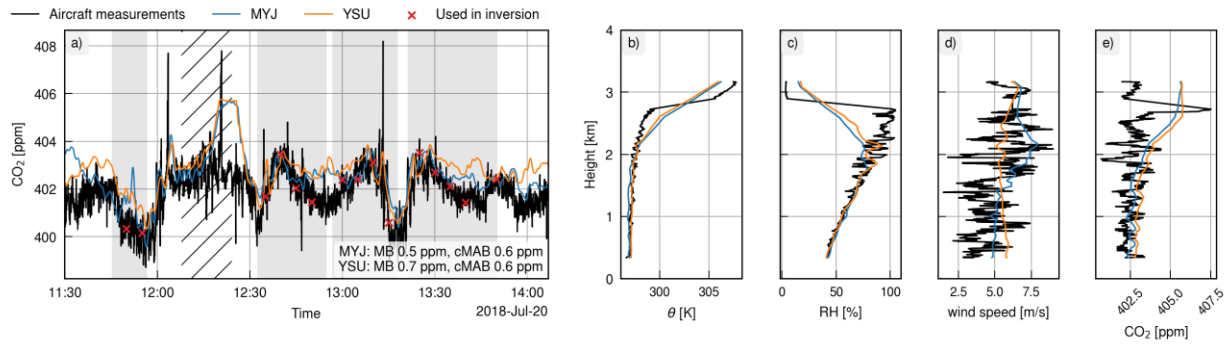


Figure 3 e). What do the 'Posterior' units represent? Please add an explanation in the caption.

We have added the sentence "The gain describes for each measurement the change in fluxes per model data mismatch in units of flux/ppm." to the caption.

Please make sure that the posterior uncertainty is consistent throughout the manuscript. In conclusions (line 360), it is specified as ± 4.3, while in the abstract (line 8) you write ± 4.6.

This was a typo in the manuscript, which we have amended. The uncertainty of the final emissions estimate is a square sum of the posterior emissions estimate uncertainty (4.3 Mt a-1) and the uncertainty introduced from the sensitivity study (1.5 Mt a-1). This yields a final uncertainty of 4.6 Mt a-1.

References

Li, J., Li, P., Han, P., Cheng, Z., Li, J., Zhang, T., Chen, D., Zheng, Y., Zeng, N., and Zhang, G. (2026). Advances in the design of urban CO_2 emission monitoring networks: a review. Carbon Research, 5, 3. https://doi.org/10.1007/s44246-025-00239-z

Karion, A., Lopez-Coto, I., Gourdji, S. M., Mueller, K., Ghosh, S., Callahan, W., Stock, M., DiGangi, E., Prinzivalli, S., and Whetstone, J. (2021). Background conditions for an urban greenhouse gas network in the Washington, DC, and Baltimore metropolitan region. Atmospheric Chemistry and Physics, 21, 6257–6273. https://doi.org/10.5194/acp-21-6257-2021

Klausner, T., Mertens, M., Huntrieser, H., Galkowski, M., Kuhlmann, G., Baumann, R., Fiehn, A., Jöckel, P., Pühl, M., and Roiger, A. (2020). Urban greenhouse gas emissions from the Berlin area: A case study using airborne CO_2 and CH_4 in situ observations in summer 2018. Elementa: Science of the Anthropocene, 8, 15. https://doi.org/10.1525/elementa.411

References:

Crippa, M., Guizzardi, D., Pisoni, E., Solazzo, E., Guion, A., Muntean, M., et al. (2021). Global anthropogenic emissions in urban areas: patterns, trends, and challenges. *Environmental Research Letters*, *16*(7), 74033. doi: 10.1088/1748-9326/ac00e2

Ganesan, A. L., Rigby, M., Zammit-Mangion, A., Manning, A. J., Prinn, R. G., Fraser, P. J., et al. (2014). Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods. *Atmospheric Chemistry and Physics*, *14*(8), 3855–3864. doi: 10.5194/acp-14-3855-2014

Hu, X., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of Three Planetary Boundary Layer Schemes in the WRF Model. J. Appl. Meteor. Climatol., 49, 1831–1844, https://doi.org/10.1175/2010JAMC2432.1.

IPCC. (2022). *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (P. R. Shukla, J. Skea, R. Slade, A. A. Khourdajie, R. van Diemen, D. McCollum, et al., Eds.) [Book]. Cambridge, UK and New York, NY, USA: Cambridge University Press. doi: 10.1017/9781009157926

Karion, A., Lopez-Coto, I., Gourdji, S. M., Mueller, K., Ghosh, S., Callahan, W., et al. (2021). Background conditions for an urban greenhouse gas network in the Washington, DC, and Baltimore metropolitan region. *Atmospheric Chemistry and Physics*, *21*(8), 6257–6273. doi: 10.5194/acp-21-6257-2021

Klausner, T., Mertens, M., Huntrieser, H., Galkowski, M., Kuhlmann, G., Baumann, R., et al. (2020). Urban greenhouse gas emissions from the Berlin area: A case study using airborne CO2 and CH4 in situ observations in summer 2018. *Elem Sci Anth*, *8*. doi: 10.1525/elementa.411

Pilz, L. N. (2026). *Monitoring urban greenhouse gas emissions using high-resolution models* (Heidelberg University Library). Heidelberg University Library. doi: 10.11588/HEIDOK.00037983

Pitt, J. R., Lopez-Coto, I., Hajny, K. D., Tomlin, J., Kaeser, R., Jayarathne, T., ... & Shepson, P. B. (2022). New York City greenhouse gas emissions estimated with inverse modeling of aircraft measurements. Elem Sci Anth, 10(1), 00082.

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding*. WORLD SCIENTIFIC. doi: 10.1142/3171