This paper describes an inverse modelling approach developed for use with aircraft data, with explicit optimisation of the background within the inversion framework. The approach is demonstrated by applying it to a single flight around Berlin. It is found that this inverse modelling based approach yields more a more reasonable emission estimate, with a lower uncertainty, than a mass balance approach using the same data. I agree that the mass balance method is too crude to apply to most urban areas and I also agree that the mole fractions of the inflowing air need to be optimised within the inversion. However, I have some questions regarding the approach taken here that I think need some extra thought.

My main issue is that I don't really understand the motivation behind this particular construction of the state vector. As far as I understand it, in this case it contains 5 emission scaling factors and 15 boundary condition scaling factors. The emission scaling factors are assumed to be uncorrelated (if I've understood that correctly then it should be stated explicitly in section 2.5.1), while the background offset correlation is adjusted so that the background concentrations don't become more correlated in the posterior. The rationale for this last part seems unclear to me. It is true that it would be undesirable to push the plume into the background offsets, and that this would lead to a higher correlation in the background. But an upwind plume that should be in the background (but is not well captured by the prior) would surely have the same effect. So increasing the background correlation in the posterior does not seem inherently bad.

Previous studies (e.g. Lopez-Coto et al., 2020) have handled the background by optimising the flux field at the grid scales (with appropriate spatial correlations) in a wider domain. This makes good use of the upwind transect – the inversion can use this data to optimise the upwind flux map, and this optimised background is also propagated onto the downwind transects. Of course there is still some risk that the inversion responds to a large downwind city plume by increasing sources far upwind, but the upwind transect acts against this (especially if the model-data mismatch errors have an element that is proportional to the magnitude of pollution and/or the variability within the averaging period). It seems like this spatial information regarding the background is essentially lost in the method presented here, and that is a shame because it is one of the big advantages of flying the aircraft. I'm not saying that the authors should necessarily adopt the approach of Lopez-Coto et al., 2020, but I think this element of the inversion needs more thought, especially as the treatment of the background within the inversion is the main focus of this paper.

We thank Joseph Pitt for the constructive comments. We agree that the motivation for our state-vector construction and background optimization requires further clarification. We now explicitly state that the prior uncertainties of the five emission scaling factors are assumed to be uncorrelated.

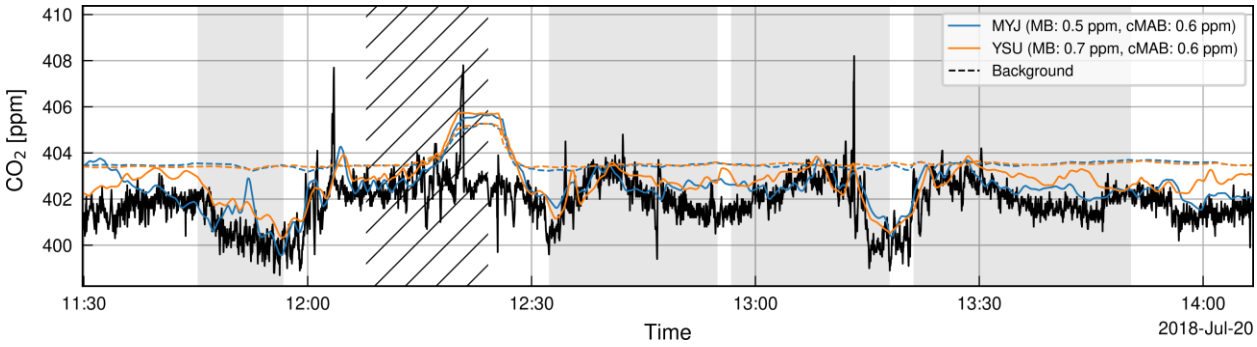
We also agree that an increased correlation between optimized background concentrations and observations is not inherently undesirable as this could also arise from a plume transported into the domain.  However, our approach is not designed to reconstruct fine background plumes, but rather to correct for larger scale biases. In our setting, the background contribution considered here originates from outside the European modelling domain, so far away from the domain of interest. Therefore, we expect plumes from outside of the European domain to have smoothed out considerably when reaching Berlin. In contrast, emission contributions within Europe are explicitly represented by the emission scaling factors. Further, the limited number of aircraft observations does not provide sufficient information to independently constrain both strongly varying background concentrations and local emissions contributions. Allowing the background to vary too freely would therefore introduce additional ambiguity between the background and emission components and could lead to the attribution of urban plume enhancements to background adjustments. For this reason, relatively smooth background conditions are an important prerequisite for applying our approach. We verified that these conditions were met during the selected flight (cf. Figure below) and ensure that the urban plume is not pushed into the background by checking that the correlation of the background does not increase.

We also agree that optimizing spatially resolved upwind fluxes is an alternative and valid inversion set-up. However, this approach has limitations when applied to a small observational dataset. Increasing the number of independently optimized flux elements introduces additional degrees of freedom, which may not be sufficiently constrained by the observations. Nesser et al. (2025) recently compared the correction of background concentration errors through the optimization of boundary concentrations with an alternative approach that optimizes surface fluxes. They found that directly optimizing boundary concentrations generally provides a more accurate and computationally efficient correction of background errors. While these findings depend on the setting of the problem, they generally support our choice of set-up to optimize the background concentrations directly. Note that, in practice, our approach is a mixture of background flux optimization (domain 3 and 4), and background concentration optimization.

We acknowledge that our approach cannot fully account for unresolved background patterns such as transported plumes. However, given the limited observational dataset, we consider the direct optimization of background concentrations a suitable compromise between accounting for background uncertainty and limiting the degrees of freedom of the inverse problem. The objective of our study is not to reconstruct the underlying upwind

flux distribution, but to improve the treatment of background uncertainty when estimating urban emissions.

We have expanded the discussion to explain the rationale behind our state-vector construction, in particular of the correlation in the posterior background, and clarify the limitations of our method.




One potential issue with employing a more spatial approach to the background optimisation is the limited amount of data. This raises two other questions for me. Firstly, why were footprints calculated every 5 minutes? At the speed the aircraft flies it seems to me like you would get considerably more information by releasing footprints every 1 minute. I know that adds to the computational expense, but that is nothing compared to the expense of flying the aircraft, and if you only take a point every 5 minutes I feel again that these aircraft data are not being used to their full potential. The second question I have concerns the decision to "exclude measurements that exhibit negative prior correlations in background concentrations". This needs much clearer explanation and justification – at the moment I don't understand how this filter works. It would also be good to state explicitly how many measurements are there before filtering. When I look at Figure 1 it seems like there are hardly any remaining red crosses on the upwind transect, which poses an issue.

The 5-minute interval used in our study corresponds to approximately 10 km along the flight track, and therefore reducing this interval to 1 minute would, indeed, provide a considerably denser spatial sampling.

Our choice of a 5-minute interval is motivated by three main arguments: (1) Computational costs, (2) Independence of information of neighbouring data points and (3) spatial resolution of inversion.

The FLEXPART-WRF simulations are driven by meteorological fields from MACRO-2018, which are available at hourly temporal resolution. Although footprints calculated at shorter intervals would differ due to the changing aircraft position, neighbouring observations are

expected to exhibit substantial correlations in their sensitivities to surface emissions, particularly at the spatial scales represented by our state vector.

Furthermore, our inversion aims to constrain emission scaling factors for five relatively large regions, rather than reconstructing the spatial distribution of emissions at the original 1 km model resolution. The aircraft sampled the Berlin plume repeatedly along three downwind transects, providing observations with different spatial sensitivities to the urban emissions. We therefore considered the 5-minute interval sufficient, while keeping the computational requirements manageable.

We nevertheless acknowledge that a higher sampling frequency could provide additional information, particularly near plume boundaries and in regions with strong spatial gradients. We explain the sampling interval in the revised manuscript and discuss limitations.

The purpose of the filtering is to identify observations for which the assumptions underlying our background optimization are appropriate. As pointed out, these are measurements corresponding to background concentrations exhibiting relatively smooth spatial variability. Observations associated with strongly different background conditions may violate this assumption and introduce ambiguities between background variability and local emission contributions. We therefore applied a filtering criterion based on the correlations in the prior background concentrations to restrict the inversion to observations that are consistent with the assumed background structure.

We acknowledge that this filtering reduces the number of observations available for the inversion, including observations along the upwind transect, which are particularly valuable for separating background and urban emission contributions. However, the filtering removes only one upwind and one downwind measurement, so most of the transect coverage used to separate these contributions is retained. We have therefore expanded the description of the filtering procedure. We agree that after filtering, we are left with a relatively small dataset and made this limitation more explicit in the revised manuscript.

On a separate note, it would be useful to have more information on how the hour-in-day scaling factor was applied to the TNO inventory in the prior, and how the time of day was dealt with in the VPRM run. The flight itself spans the middle of the day, but what was the timeframe for the sampled air passing over Berlin? This seems especially relevant when making comparisons between the posterior and the inventory in the discussion section – the net biospheric flux changes rapidly during the morning from net emission to net uptake. As a related point, more information is needed on the VPRM configuration, especially regarding

changes made for urban areas. This should include reference to Li et al. (2026), even if a different framework has been used here.

The scaling factors of the TNO inventory were applied in the following way. Pick a datetime (Y-M-D, H:00), calculate the weekday WD, look up the scaling factors for M, WD, H in the tables provided by TNO and multiply the with the average annual emission to get hourly emissions.

Time-of-day was explicitly simulated in the VPRM run. The VPRM fluxes are outputs of a WRF simulation getting its inputs from pyVPRM and were provided by Theo Glauch (Glauch et al., 2025). 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). This VPRM product was further developed into Li et al, 2026."

The timeframe for the sampled air passing over Berlin is resolved implicitly by the 4-D footprints being multiplied with the underlying VPRM fluxes. They ensure the sensitivity to the appropriate fluxes, which are indeed strongly varying over the diurnal cycle.

We added a reference to Li et al., 2026 in the following context: "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).". We also added a reference to Li et al., 2026 as in context of the VPRM setup: "This VPRM product was further developed into Li et al, 2026."

L6 – "using the Weather Research and Forecasting (WRF) model"

We have fixed this typo.

L103 – "the concentration data for the background field"

We have fixed this typo.

L111 – check citation format

We fixed the citation format.

L144 – I would say that background values are just as often interpolated between the edges of the downwind transects. In any case both approaches are in common use so it probably makes sense to mention them both here.

We have mentioned the different methods of implementing the integration.

L165 – More clarity with the terminology would be good here. If x contains emission scaling factors then presumably the various components of f are not footprints (at least according to the standard definition) but a multiplication of the footprint by the emissions in each grid cell?

In fact, these are not scaling factors in the traditional sense. More precisely, they are the average emissions per emission cell, which is used to scale the structure that is unresolved in the state vector. We adjusted the name accordingly to "average emissions per cell in the emission state vector element".

L187 – "gain and averaging kernel". Also I'm not sure that these are usually capitalised?

We have amended the capitalization.

L190-192 – should x_0 be x_a here for consistency with above?

Indeed, x_a is correct. We have changed this.

L265 – the reference to "the sensitivity study" here is ambiguous – be explicit that this is an attempt to quantify uncertainty in the posterior related to uncertainty in the specification of the prior covariance matrix.

We have changed this accordingly.

L364 – repeated sentence

We have removed the sentence.

References:

Israel Lopez-Coto, Xinrong Ren, Olivia E. Salmon, Anna Karion, Paul B. Shepson, Russell R. Dickerson, Ariel Stein, Kuldeep Prasad, James R. Whetstone; Wintertime CO2, CH4, and CO Emissions Estimation for the Washington, DC–Baltimore Metropolitan Area Using an Inverse Modeling Technique. Environ. Sci. Technol. 3 March 2020; 54 (5): 2606–2614. https://doi.org/10.1021/acs.est.9b06619

Li, J., Chen, J., Glauch, T., Brunner, D., Marshall, J., Ponomarev, N., et al. (2026). Fine-scale estimation of urban biogenic CO2 fluxes: A novel framework integrating multiple versions of

vegetation photosynthesis and respiration models and in situ measurements. Earth's Future, 14, e2025EF007458. https://doi.org/10.1029/2025EF007458

References:

Glauch, T., Marshall, J., Gerbig, C., Botía, S., Gałkowski, M., Vardag, S. N., & Butz, A. (2025). pyVPRM: A next-generation Vegetation Photosynthesis and Respiration Model for the post-MODIS era. *EGUsphere*. doi: 10.5194/egusphere-2024-3692

Nesser, H., Bowman, K. W., Thill, M. D., Varon, D. J., Randles, C. A., Tewari, A., et al. (2025). Predicting and correcting the influence of boundary conditions in regional inverse analyses. Geoscientific Model Development, 18(23), 9279–9291. doi: 10.5194/gmd-18-9279-2025