Atmospheric Measurement Techniques

Response to Reviewers for

Theoretical assessment of the ability of the MicroCarb satellite city-scan observing mode to estimate urban CO_2 emissions

We thank both reviewers for taking the time to review our manuscript and to provide constructive suggestions and comments that have helped us improve the manuscript. Please find below the reviewers' comments in regular italic and point-by-point responses in bold font. We have revised the manuscript based on reviewer comments and quote the revised text.

Reviewer #1

This publication investigates the potential of the MicroCarb satellite mission planned to be launched in the 2023/2024 timeframe to estimate urban CO_2 emissions. The potential is analyzed for Paris and London using synthetic model experiments for a total of 15 days distributed over three representative months of the year with varying illumination conditions, biospheric activity, and cloud cover.

The MicroCarb mission will have three different observation configurations including a "city-mode", which allows observing a target with a two times higher spatial resolution (2.25 km x 2.25 km) and a two or three times wider swath than in the nominal nadir mode by performing two or three pitch maneuvers when flying over the target. The study is important as it provides a scientific basis for the decision on whether a two-sweep or three-sweep mode should be implemented.

The manuscript is very well written and concise, the model simulations are state-of-the-art, and the figures are informative. I have only a few minor points and one main concern that actually not only applies to this study but to other similar studies as well. I consider the manuscript acceptable after addressing these (primarily minor) points.

We sincerely appreciate your support for this study and this manuscript. We have revised the manuscript based on these suggestions and comments.

Main concern:

Like other studies employing an OSSE approach with synthetically generated observations, the conclusions on the ability of the satellite to constrain (urban) emissions are quite optimistic, and I would argue too optimistic. Some of the limitations are addressed in the manuscript, e.g. in the "Conclusions and discussion" section where the results are presented as "best-case" scenario. I am worried, however, that papers like this generate unrealistic high expectations of city governments or other stakeholders, which the scientific community, once the satellites are in orbit, will not be able to meet. The study is very carefully designed and executed, but in my view the discussion on the limitations of the study induced by the different assumptions is falling short. Probably the easiest solution would be to add a paragraph or two to the "conclusions and discussion" section, which so far actually contains very little discussion.

Agreed. We have revised the manuscript by extending the "Conclusion and Discussion" to discuss more possible limitations and uncertainties in using MicroCarb satellite data to estimate urban CO_2 emissions.

Here is a (incomplete) list of my concerns:

- *My general concern is that systematic errors in any of the components (e.g. atmospheric transport, observations, biospheric fluxes) are much more critical than random errors. The latter can always be reduced by increasing the number of observations, and since most OSSEs only consider random errors, the results tend to be overly optimistic. In the present study, the three-sweep mode simply leads to better results because of the larger number of observations, but this is not the only point that should be considered when comparing the two modes. There is only a short discussion on the critical issue of determining the background against which the urban enhancements are measured. Reliably determining the background concentration levels will likely be essential, especially in the presence of systematic observation biases, and will need a sufficient number of cloud-free pixels outside of the urban plume. This may be another important advantage of the three-sweep mode that is not addressed in this study.*

  Fair point. Systematic errors are generally more critical than random errors. We have revised the manuscript to discuss more about this issue and emphasize the advantage of the three-sweep mode on estimating the atmospheric background concentration.

  The following text has been added in the revised manuscript.

  "We demonstrate the ability of using synthetic unbiased MicroCarb data to infer urban CO_2 emissions, but there are additional limitations and uncertainties introduced by using real data that will have systematic and random errors (Broquet et al., 2018). Reducing random errors can be achieved by increasing the number of independent measurements. Addressing regional systematic errors is more difficult. For example, we find that column-averaged urban CO_2 enhancements are typically less than 1% of the atmospheric background column concentration (about 415 ppm), so that a regionally varying systematic measurement error of the order of 1 ppm would significantly degrade emission estimates inferred from the data. We expect that larger scale systematic errors would be identified and corrected by ground-based remote sensing instruments (e.g., TCCON). Establishing city-scale ground-based atmospheric remote sensing networks would help correct regional systematic errors.

  Estimating the regional background CO_2 column concentration is important for quantifying the magnitude of urban CO_2 enhancements due to net urban emissions, including anthropogenic and biospheric CO_2 fluxes. A range of methods have been employed in different studies to quantify regional background values (Kort et al., 2012, Ye et al., 2020, Kiel et al., 2021). Here, we have sidestepped this issue by taking advantage of the closed-loop experiment configuration. To estimate city-scale CO_2 emissions from real MicroCarb data, we will need to consider the additional uncertainty associated with the calculation of the elevated CO_2 column. Here, we find that the three-sweep observing mode outperforms the two-sweep mode because of the

larger number of observations that help to reduce random errors. The three-sweep mode also has a wider scan outside the urban plume that will help to reduce bias in estimating regional background CO_2 values."

- *The closed-loop experiment assumes unbiased transport errors. Although this limitation is acknowledged by the authors, there is essentially no discussion on how such errors could affect the emission estimation. Is there no information available from the meteorological community about biases and errors in their analyzed wind products? Relative errors in wind speeds are almost certainly largest under low wind conditions, but these are the conditions that produce the strongest signals in an OSSE experiment as presented here. Thus, the situations working best in this OSSE may actually be the most problematic in reality. Probably there is an optimal range between low wind speeds, where relative errors are large, and large wind speeds, where the urban enhancements are too small to be detected. Unfortunately, the present study presents no insights into this question.*

  We agree. We have revised the discussion section to expand this point:

  "Our study also assumes unbiased and uncorrelated atmospheric transport errors. However, these errors are likely to be correlated at the sub-city scale. Also, relatively low (high) wind speed will produce the strong (weak) urban enhancement signals but are associated with relatively large (small) transport errors. Better understanding atmospheric transport and {\em a~priori} flux errors is essential to improve the accuracy and precision of high-resolution urban {\em a~posteriori} CO_2 flux estimates. Our assumptions about atmospheric transport errors likely result in the best-case scenario for error reduction that can be achieved by the MicroCarb city-scan observing modes."

  "This study assumes unbiased and uncorrelated atmospheric transport errors. However, these errors are likely to be correlated at the sub-city scale. We acknowledge that X-STILT used in the study is not tuned for simulating elevated emissions given the use of vertical-integrated 2-D footprint. Future work will include the use of the vertical resolved footprint associated with each air parcel, and the vertical profiles of sector-specific emissions to improve the hyper-nearfield simulations of point sources following prior approaches (Maier et al., 2022; Mallia et al., 2018). Also, relatively low (high) wind speed will produce the strong (weak) urban enhancement signals but are associated with relatively large (small) transport errors. Better understanding atmospheric transport and *a priori* flux errors is essential to improve the accuracy and precision of high-resolution urban *a posteriori* CO_2 flux estimates. Our assumptions about atmospheric transport errors likely result in the best-case scenario for error reduction that can be achieved by the MicroCarb city-scan observing modes."

  We partially tested the impacts of wind speed by simulating footprints in different dates (Figs. A3 and A4), and the results in Figs. 8 and 9 partially include the effect of different wind speeds. However, we didn't try to identify all the individual sources of uncertainties in inversion configurations. Identifying the impacts of wind speeds on our ability to quantify urban CO_2 elevation is beyond the scope of this present study.

- *Cloud cover is also treated in a random fashion. Depending on cloud cover, a probability is estimated for a pixel to be cloud-free or not, which results in a random pattern of cloud-free observations. However, clouds tend to be organized and to obscure one part of the image more than another part. As a result, the satellite will likely be blind to emissions in some parts of the city, an effect that is not captured by a purely random pattern. Furthermore, with increasing cloudiness, also cloud-free pixels tend to show biases in XCO2 due to 3D cloud radiation effects (e.g. Massie et al., 2017; https://doi.org/10.1002/2016JD026111). As a result, the ability to constrain urban emissions will likely drop faster with increasing cloud cover than suggested by Fig. 6. Figure 6 would tell the operators of Microcarb that a 50% cloud cover would reduce the error reduction for Paris only from 13% to 10% as compared to a cloud-free case (for a 1 ppm observation uncertainty). I doubt that this will be true in reality.*

  <<CLOUD DISTRIBUTIONS>>
  To clarify, cloud cover in this study is not totally treated in a random fashion. We have added clarifying text in the revised manuscript:

  "For each individual measurement pixel, we calculate its cloud-free probability (*f* in Equation 1) based on ERA5 total cloud cover reanalysis data at 0.25º × 0.25º resolution, then we generate a random number for each measurement to compare with its cloud-free probability (*f*) and determine whether a measurement is cloud-free. The first step is related to the organization of cloud distribution based on the ERA5 data, and the second step is related to the randomness of the impact of cloud on satellite measurements."

  We have used this approach to describe cloud cover because we could not find cloud cover data at the spatial resolution of the city-scan mode (2.25 km × 2.25 km). As such, Fig 2 looks like a random pattern that does not show the organized structure of cloud.

  We have also extended the following discussion in the revised manuscript:

  "As with other polar-orbiting satellites, MicroCarb provides a limited amount of data over urban areas due to seasonal cloud cover (especially for mid-to-high latitude cities) and elevated aerosol loading in urban areas, and because of their orbital configuration. As a result, the instrument will not see CO_2 immediately overhead of some parts of the city and some cloud-free pixels, particularly between cloud breaks, will likely have a bias due to 3-D cloud radiation effects (Massie et al., 2017). Table A1 lists the number and percentage of cloud-free data over our test cities. In Paris, four of the nine days collect cloud-free data (the other five days are covered by cloud at 12 UTC). For these four days, only 40-50% of the data are cloud-free. Lei et al. (2021) found only 17% of OCO-2 soundings over the 70 most populated cities are of high-quality, consistent with our calculations (18-22%). The number of available measurements is reduced further considering the effective footprints over urban areas. Using satellites to collect data throughout the day over a particular location only be achieved by with a constellation of satellites or with a satellite in a geostationary orbit, neither solution is currently planned."

To clarify the results in Fig. 6, which is the spatial mean value of flux error reduction. A 3% difference of the spatial mean error reduction (where cloud cover increases from 0 to 50%) is not trivial after integrating it for the whole domain.

- *Point sources are excluded in the OSSE, but point sources are an important part of reality, not only outside but also inside cities. In many cities, emissions from waste incinerators, combined heat and power plants or industrial stacks make up a large fraction of the total CO_2 emissions and the emissions are usually not constant in time. X-STILT, to my knowledge, only computes the sensitivity to emissions at the surface but not to elevated emissions. The mixture of emissions released near the surface and from stacks will pose a major challenge to quantify urban emissions.*

<<LARGE POINT SOURCES>>
Thanks for identifying this point. We would like to clarify that we only exclude extremely strong point sources (e.g., power plants and some industrial emissions) whose emissions strengths are more than 100 times larger than the other conventional emission pixels. We keep most of the conventional emission sources (including areal and point sources). Fig. 1 in this response shows the filtering process of those extremely strong point sources (the Y axis is plotted at the logarithmic scale). To clarify this point, we added the following text in the discussion section:

"This study excludes a few very extremely large emission sources (e.g., power plants and some industrial emissions) that are more than 100 times larger than other (areal and point sources) emission pixels but have lower uncertainties. In total, nine pixels are excluded in Paris (about 0.1% of 10,080 pixels) and 22 pixels are excluded in London (about 0.2% of 10,368 pixels). We exclude these sources because: 1) the atmospheric transport model is not sufficiently accurate and precise to infer the location and emissions strength of these sources; and 2) these anomalously large emissions complicate the matrix-based inverse method calculation. We acknowledge that real measurements will likely detect these large emission sources, subject to changes in cloud cover and wind direction and speed. For meteorological circumstances where the resulting plume takes the form of thin, anomalously large CO_2 enhancement, only very high-resolution models will be able to describe the plume and estimate its parent emissions with fidelity."

"For our calculations, we exclude a small number of extremely large emissions sources (e.g., power plants and some industrial emissions) that are more than 100 times larger than other (areal and point sources) emission pixels but have lower uncertainties. In total, we exclude nine pixels over Paris (about 0.1% of 10080 pixels) and 22 pixels are excluded in London (about 0.2% of 10368 pixels). We exclude these sources because: 1) the atmospheric transport model is not sufficiently accurate and precise to infer the location and emissions strength of these sources and errors in the locations of these point sources diminish the ability to infer sub-city scale emissions (Oda et al., 2017; Roten et al., 2022); and 2) these anomalously large emissions complicate the matrix-based inverse method calculation. We acknowledge that real

measurements will likely detect these large emission sources, subject to changes in cloud cover and wind direction and speed. For meteorological circumstances where the resulting plume takes the form of thin, anomalously large CO$_2$ enhancement, only very high-resolution models will be able to describe the plume and estimate its parent emissions with fidelity."

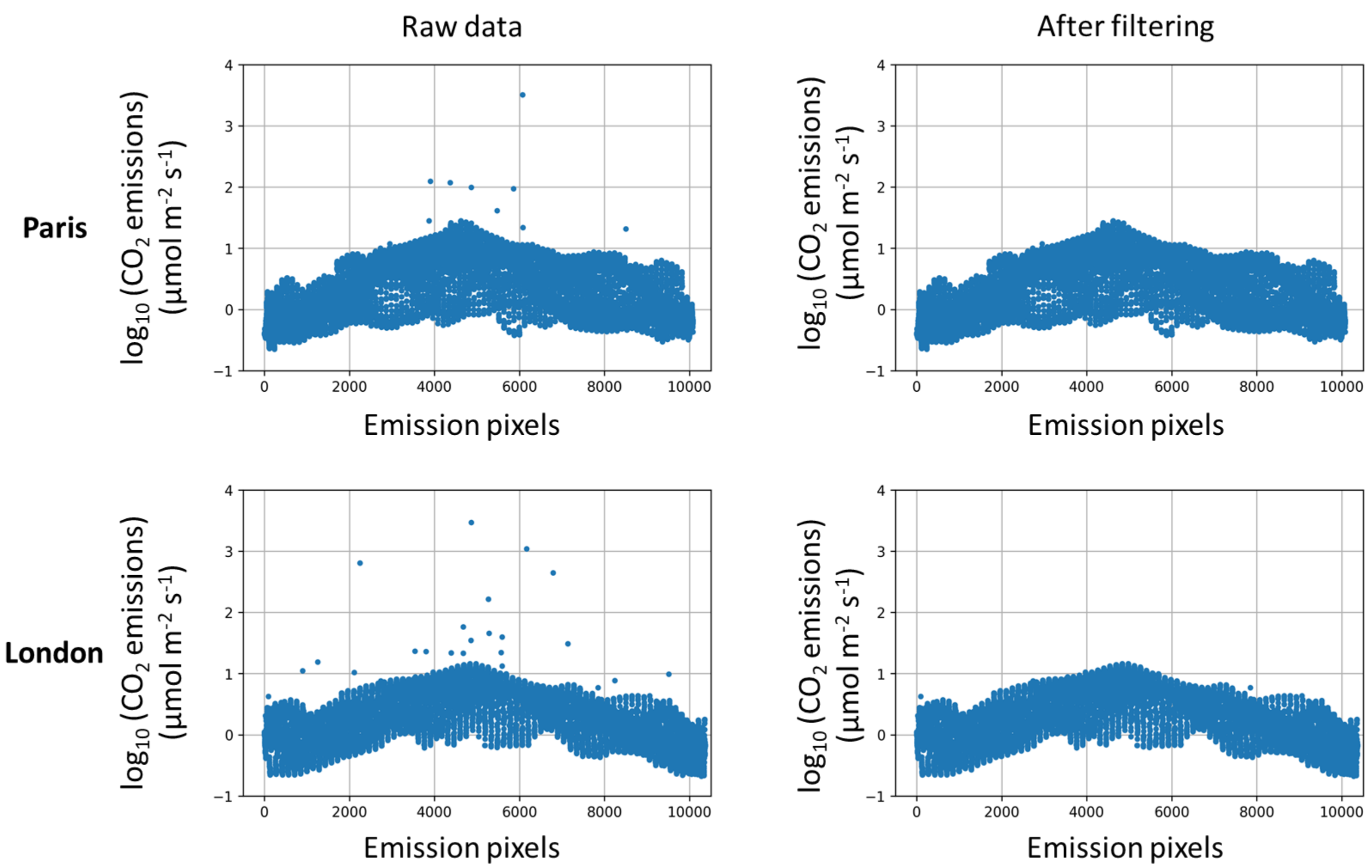

Fig 1. ODIAC emissions in April 2018 in Paris (top panels) and London (bottom panels) plotted in a logarithmic scale for identifying extremely strong point sources like power plants and some industrial emissions whose emissions strengths are more than 100 times larger than the other conventional emissions pixels.

For simulating the mixture of emissions released near the surface and from stacks in X-STILT, we agree that properly accounting for the stack height and even the plume height from point source emissions or elevated emissions is critical when using an atmospheric transport model. Although X-STILT used in this work was not particularly tuned for point sources, these challenges can be tackled via future model implementations (based on existing STILT-related studies, e.g., Mallia et al., 2018; Maier et al., 2022). Specifically, we clarify that the default final product of the (X-STILT model is the vertical-integrated spatial (2D) map of footprint, which assumes that emissions are instantaneously well mixed within the mixed layer (i.e., equal contributions to enhancements from every air parcel within the mixed layer). The intermediate product of X-STILT is the vertically resolved "footprint" related to each air parcel with proper averaging kernel and pressure weighting function.

To better acknowledge this model limitation, we have revised the manuscript:

"We acknowledge that X-STILT used in the study is not tuned for simulating elevated emissions given the use of vertical-integrated 2-D footprint. Future work will include the use of the vertical resolved footprint associated with each air parcel, and the vertical profiles of sector-specific emissions to improve the hyper-nearfield simulations of point sources following prior approaches (Maier et al., 2022, Mallia et al., 2018)."

- *Microcarb as other satellites will only measure at a single time of the day and will observe a given city only very rarely in a given year (actually, it would be good to know how many samples can be obtained over a city under optimal conditions). Emissions have a diurnal cycle and change from day-to-day. However, cities would like to better know their annual mean emissions and how they evolve over the years. Because of the uncertainties in temporal variability of emissions, observations on a few days per year (if possible at all) will provide only small constraints on annual mean emissions. Some of these challenges were addressed in Broquet et al. (2018; https://doi.org/10.5194/amt-11-681-2018), which also performed an OSSE for Paris.*

  Yes, we agree that MicroCarb will only provide a limited number of data in urban areas and we appreciate the discussion in Broquet et al. (2018). We have revised the manuscript:

  "As with other polar-orbiting satellites, MicroCarb provides a limited amount of data over urban areas due to seasonal cloud cover (especially for mid-to-high latitude cities) and elevated aerosol loading in urban areas, and because of their orbital configuration. As a result, the instrument will not see CO_2 immediately overhead of some parts of the city and some cloud-free pixels, particularly between cloud breaks, will likely have a bias due to 3-D cloud radiation effects (Massie et al., 2017). Table A1 lists the number and percentage of cloud-free data over our test cities. In Paris, four of the nine days collect cloud-free data (the other five days are covered by cloud at 12 UTC). For these four days, only 40-50% of the data are cloud-free. Lei et al. (2021) found only 17% of OCO-2 soundings over the 70 most populated cities are of high-quality, consistent with our calculations (18-22%). The number of available measurements is reduced further considering the effective footprints over urban areas. Using satellites to collect data throughout the day over a particular location only be achieved by with a constellation of satellites or with a satellite in a geostationary orbit, neither solution is currently planned."

<u>*Minor points and corrections*</u>*:*

- *I had to look up other documentation of MicroCarb to fully understand the city-mode. If the satellite would only perform pitch-maneuvers (as I thought after reading section 2.1), it would miss many cities or observe them only partially. However, as stated on https://microcarb.cnes.fr/en/MICROCARB/GP_mission.htm, the pointing-mirror allows also movements along the roll-axis by pointing the line of sight on either side of the satellite* *ground track up to ±200 km. This should be explained more clearly, because otherwise it is difficult to understand why the cities are exactly in the center of the scans. Furthermore, I assume that the sampling pattern shown in Figure 2 only holds when the city is exactly in nadir. How does the pattern look like when the satellite points some 100-200 km to the side (would be good to show e.g. in the Annex)? This will likely be necessary in many cases to have good coverage of a given target. Is the distortion of the pattern and the degradation of the resolution important or not? Would that affect the conclusions of the study?*

The across-track scanning mirror helps to enlarge the swath during the backwards and the forwards sweeps. This scanning mirror is also used to shift all the pattern across-track when pointing at an off-nadir city.

We revised the texts to explain this better:

"The two-sweep and three-sweep city-scan sampling configurations (Figure 2}) are being considered by the MicroCarb science team. Both city-scan configurations cover a city the size of Paris or London. For any acquisition mode, the spatial across-track (ACT) swath is acquired by a detector that has 90 pixels in the spatial dimension and 1024 pixels that collects spectral information for the four infrared bands described above. In the nominal nadir mode of MicroCarb, each individual ACT sounding results from the binning of 30 ACT pixels. The spatial along-track (ALT) dimension is determined by the satellite velocity during the 1.3 s measurement integration time, which leads to the nominal pattern of three soundings of 4.5 km ACT x 8.9 km ALT, leading to a 13.5 km ACT swath and a continuous ALT sampling. The city-mode differs from the nominal nadir pattern in three ways: 1) the satellite makes a permanent pitch rotation to slow the projection of the line of sight on the ground, leading to an ALT footprint size of 2.25 km (for the same 1.3 s measurement integration time); 2) each individual ACT footprint results from the binning of 15 ACT pixels, leading to six ACT footprints each with an ACT size of 2.25 km; and 3) the ACT swath of the complete city-mode measurement configuration is constructed by the juxtaposition of two or three ALT sweeps, enabled by satellite pitch manoeuvres before and after flying over the target that provides forward and backward viewing of the target. Each successive sweep is shifted by 13.5 km ACT using a roll-axis pointing mirror to get contiguous spatial observations. The roll-axis pointing mirror also provides the opportunity to centre the city-mode sampling pattern over cities that lie aside of the satellite track. Further instrument details that enable the city-scale mode can be found in Jouglet et al. (2021).

We study two configurations in the MicroCarb satellite city-scan observing mode (Figure 2): 1) the two-sweep configuration has only forward and backward observations, leading to a ∼40 km ACT swath (2 x 13.5 km including the off-nadir pointing footprint deformation); and 2) the three-sweep configuration has forward, nadir, and backward observations, leading to a ~60 km ACT swath. These are typical values including off-nadir centred city modes. For a city mode centred at nadir, the ACT swath is 34 km for the two-sweep mode, 52 km for the three-sweep mode. The allowed duration of each sweep mode leads to an ALT swath of ~40 km."

- *It is probably correct that urban areas are responsible for ~70% of global emissions, but at the same time it is clear that only a fraction of this is actually emitted within the city boundaries, since a large part of the power consumed by cities is generated by power plants outside. I always find this number of 70% misleading when it is presented without further explanation.*

  Thanks for identifying this point. We revised the texts as follows:

  "Urban areas account for about 70% of global energy-related anthropogenic CO_2 emissions (including Scope 1 and 2 emissions), while comprise only 3% of Earth's surface area (Turnbull et al., 2018)"

- *The introduction section misses a number of references that I consider important for this paper:*

  Other CO2 satellite OSSE studies for urban areas: Broquet et al. 2018 (see above) and Kuhlmann et al. (2020; https://doi.org/10.5194/amt-13-6733-2020).

  Study on CO2 emissions of the city of London based on aircraft observations: Pitt et al. (2019; https://doi.org/10.5194/acp-19-8931-2019).

  Study on large CO2 emitters (including cities) as observed from OCO-2: Chevallier et al. (2022; https://doi.org/10.1029/2021GL097540).

  Thanks for suggesting these references, which have been cited in the manuscript.

- *Page 3, Line 60: Change "pass over" to "passes over"*

  Done

- *P3, L68: Change "analyses" to "analysis" (singular form needed here)*

  Done

- *P3, L87: The ACT swath seems much closer to 30 km than to 40 km, and similarly for the 3-sweep mode seems closer to 45 km than to 60 km.*

  We revised the texts to explain more about this issue as follows:

  "We study two configurations in the MicroCarb satellite city-scan observing mode (Figure 2): 1) the two-sweep configuration has only forward and backward observations, leading to a ~40 km ACT swath (2 x 13.5 km including the off-nadir pointing footprint deformation); and 2) the three-sweep configuration has forward, nadir, and backward observations, leading to a ~60 km ACT swath. These are typical values including off-nadir centred city modes. For a city mode centred at nadir, the ACT swath is 34 km for the two-sweep mode, 52 km for the three-sweep mode. The allowed duration of each sweep mode leads to an ALT swath of ~40 km."

- *P4, L98: Change "where g defined" to "where g is defined"*

  Done

- *P4, L110: I was confused by the usage of the term "scene", which apparently refers to a single observation/pixel rather than a full "scene" of a given area.*

  We have revised this problem and replaced "scene" with "observation"

- *P5, L143: The high correlation coefficient of 0.98 gives a false impression of accuracy. It only tells us that the diurnal cycle is on average well represented, which seems comparatively trivial as it follows the mean diurnal cycle of radiation. The more important message is that the fluxes are in a realistic range.*

  Thanks, we agree the high correlation is based on the mean diurnal cycle of radiation. We have removed the correlation coefficient and revised the texts as follows:

  "The seasonal-mean NEEs simulated by SMUrF in July 2018 agree reasonably with the eddy-covariance flux data (slope=1.07 for the ensemble mean of all sites and the cumulative seasonal-mean NEE is -12 g m^{-2} day^{-1} for the simulated flux and -10 g m^{-2} day^{-1} for the flux data) indicating that the simulated biogenic CO_2 fluxes are in a realistic range."

- *P7, L187: Please provide more details on the "realistic noise model". Which parameters determine the noise and how was it applied in the present study? Do you explicitly consider, for example, surface reflectance, and if so, based on which data set?*

  <<MICROCARB NOISE MODEL>>

  Thanks for identifying this point. We revised the texts to explain this issue as follows:

  "We then add realistic observation noise to each cloud-free scene (determined by the cloud coverage) based on estimates from detailed simulations of the MicroCarb measurement error budget. These simulations transfer the instrumental performances (signal to noise ratio, spectral sampling, and spectral band positions) to random measurement error using the gain matrix formalism from Rodgers (2000). Only the radiometric noise has been included here. The measurement random error is expressed as the standard deviation of a Gaussian probability distribution law, which is computed for each footprint given its Sentinel-2 L1C reflectance, solar zenith angles, and atmospheric profiles for temperature, H_2O, and CO_2. The random measurement errors range between 0.4–1.6 ppm with a mean value of 0.93 ppm for both the two-sweep and three-sweep configurations over Paris, and mean values of 0.88 ppm (two-sweep) and 0.91 ppm (three-sweep) over London (Figure A5)."

- *P7, L198: What is the assumption of a 25% random error for the biological fluxes based on? I can't track this value from the material presented in the paper or in Figure A2. Is this* *uncertainty assumed to be purely random and spatially uncorrelated?*

The assumption of a 25% random error is an approximately temporal mean value of the standard deviation of the seasonal means of the simulated biogenic CO_2 fluxes, which has been clarified in the manuscript. This is a very simplifying assumption without considering the spatial correlation (i.e., spatially uncorrelated). We did not try to make an in-depth estimation of the random error for the biological fluxes, which is a little beyond the scope of this study.

- *P7, L200: The eigenvalue decomposition sounds interesting. Please provide more information on how it works.*

  We revised the texts to explain more about this issue as follows:

  "We use an eigenvalue decomposition method to generate *a priori* flux noise from the flux error covariance matrix (B). The vector of flux noise is calculated by multiplying the eigenvector of the flux error covariance matrix (after considering the spatial correlation of errors) with a normal distribution vector characterizing the systematic and random flux errors."

References:

Mallia, Derek V., Adam K. Kochanski, Shawn P. Urbanski, and John C. Lin. "Optimizing smoke and plume rise modeling approaches at local scales." *Atmosphere* 9, no. 5 (2018): 166.

Maier, F., Gerbig, C., Levin, I., Super, I., Marshall, J., and Hammer, S.: Effects of point source emission heights in WRF–STILT: a step towards exploiting nocturnal observations in models, Geosci. Model Dev., 15, 5391–5406, https://doi.org/10.5194/gmd-15-5391-2022, 2022.

Lei R, Feng S, Danjou A, et al. Fossil fuel CO2 emissions over metropolitan areas from space: A multi-model analysis of OCO-2 data over Lahore, Pakistan[J]. Remote Sensing of Environment, 2021, 264: 112625.

Jouglet D., Landiech P., Bréon F.M. et al. : MicroCarb, first European program for CO_2 monitoring: nearing development conclusion before launch, IWGGMS-17, 2021.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding, WORLD SCIENTIFIC., 2000.

Reviewer #2

GENERAL COMMENTS

This study evaluates the ability of the upcoming MicroCarb satellite to retrieve anthropogenic CO_2 emissions based on synthetic measurements. The study follows a well-established approach for such observation system simulation experiments: generating perfect observations based on flux inventories and a model of atmospheric transport, perturbing them based on a model of instrument error, using the perturbed observations to retrieve fluxes in an inverse model of atmospheric transport, and analyzing these fluxes.

While the modeling and analysis methods are sound, I have some concerns on the setup of the experiments and, consequently, applicability of the results to real measurements. In addition, I believe that the interpretation of a key result is inaccurate.

The manuscript is mostly easy to read. A few paragraphs, especially on methods and experimental setup, need clarifications and additional information to better understand the results.

In summary, the manuscript is valuable for the CO2 flux estimation community to understand the capabilities of the upcoming MicroCarb satellite. However, especially the major issues summarized above need to be addressed before publication, which might require additional simulations or better communication of the limitations in abstract and conclusions.

We appreciate your comments and suggestions that have helped to improve the manuscript. We have revised the manuscript based on these suggestions and comments.

SPECIFIC COMMENTS

In my opinion, the major issues to address are:

- The authors omit large point sources in their analyses. Therefore, I'm not sure how the results relate to real measurements, which of course see the integrated signal of all sources (see comments on Sect. 2.3).

Thanks for pointing this out. Please see our response above to Reviewer 1 <<LARGE POINT SOURCES>> that includes the revised text.

- I believe that the authors' interpretation of their results on biogenic fluxes is not accurate (see comments on Sect. 3.3).

- Information on the optimization method is missing, namely how the state vector is set up and, related, how the posterior anthropogenic flux component is obtained (see comments on Sect. 2.7)

- The manuscript needs more information on the observation error model (see comments on Sect. 2.7)

Thanks, we addressed the above comments on specific sections, as described below.

Below I elaborate on these major and some other points in detailed comments arranged by section.

<u>Abstract or Introduction</u>:

State somewhere early on that the study analyses only one out of many sources of uncertainty in emission estimation with MicroCarb, i.e. random measurement errors (as discussed in Sect. 4).

We revised the manuscript to claim this point in the introduction.

"We quantify the impacts of different observing modes, random measurement errors, cloud cover, and biogenic fluxes on inferring urban fossil fuel CO_2 emissions using synthetic MicroCarb observations."

<u>2.2. Cloud screening</u>

This is probably a minor issue, but please comment:

The authors decide cloudiness with a random number generator while in reality there should be some spatial correlation when clouds are bigger than satellite pixels. So e.g. for cloud cover 50%, the analysis is valid for clouds that are smaller or similar in size as the satellite pixel. I don't know whether this is a realistic assumption or what impact larger clouds would have.

Thanks for identifying this point. We agree that there should be spatial correlation when clouds are bigger than satellite pixels. We would like to clarify that cloud cover in this study is not totally treated in a random fashion. Please see our response to Reviewer 1 labeled <<CLOUD DISTRIBUTIONS>> that includes the amended text.

<u>2.3 Anthropogenic emissions</u>

The authors exclude point sources from the analysis, citing smaller uncertainty than the areal anthropogenic sources they do include. I don't understand that rationale because anthropogenic fluxes are the main focus of the study and the real measurements will of course include all sources. I think that this omission is problematic, because it alters some results of the study:

Thanks, we addressed a similar comment in Page 5 <<LARGE POINT SOURCES>> that includes the revised text.

- *As mentioned by the authors, the prior uncertainty of the areally diffuse sources is higher than that of the point sources. Therefore, the random prior uncertainty used in this study is higher than that of the total anthropogenic emissions. Therefore, the reduction of random uncertainty (the "skill" of the observation system) of city-wide integrated anthropogenic emissions is overestimated.*

Yes, the prior uncertainties of those extremely strong point sources (e.g., power plants) are generally smaller than conventional areal and point sources, presumably because there is

more information about them included in the inventory. We do not include them into the inversion system of this study because inventories (such as ODIAC) estimate them using a relatively accurate and precise approach than the other conventional (areal and point) emissions sources that are the focus of this study. Prior emission uncertainties used in this study are smaller than the total anthropogenic emissions. For example, Fig. 4b shows that the random prior uncertainty is 0.89 tCO_2 s^{-1} that is about 34% of the total anthropogenic emissions (2.6 tCO_2 s^{-1}) in Paris. Thus, the reduction of random uncertainty is not overestimated.

- The atmospheric signal is reduced due to the omission of point sources, which could imply that (relative) biases in posterior emissions are overestimated in this study.

Yes, we agree the atmospheric signal could be reduced compared to real-data experiments depending on whether the satellite can sample a plume of a power plant with little diffusion on the specific sampling day, which is related to wind direction and wind speed on the sampling day and the location of the satellite. Since the true emissions do not include those extremely strong emission sources (nine pixels in Paris and 22 pixels in London), the truth, prior and posterior emissions and the synthetic data (generated from the truth) consistently do not include those strong emissions points. Thus, we believe the biases in posterior emissions are not overestimated.

- Omitting point sources means that this study is about an artificial problem that includes only areally diffuse anthropogenic emissions, whereas the abstract includes statements on "total emissions".

Every OSSE underestimates the skill of an observation system because they cannot account for/quantify all sources of uncertainty, and one may argue that it is fine to only focus on the harder flux component. However, as explained above, some uncertainties are overestimated while some are underestimated w.r.t. total anthropogenic emissions, so how the results relate to the "real" problem of integrated signals is complex. One way out could be to argue that point source plumes may be removed from the real MicroCarb signal (i.e. reducing the real problem to the one studied here), but this would introduce new errors, and their analysis would fill its own dedicated paper.

I think that the best way forward would be to repeat the analyses but with the strong point sources included. Alternatively, the limitations arising from the omission need to be clearly stated and discussed in the manuscript (but in my opinion this would diminish the value of the manuscript because only a partial problem is studied).

Thanks for pointing this out. We addressed a similar comment in Page 5 <<LARGE POINT SOURCES>> that includes the revised text. We have revised the manuscript to discuss more about this issue.

<u>*2.7 Experiments*</u>

The section on measurement errors needs to be expanded: please provide a reference or add a dedicated, if short, section on synthetic measurement errors, including how they were estimated and possible limitations (e.g. if there are error sources that were not considered).

Thanks for identifying this point. Please see our response to Reviewer 1 labeled <<MICROCARB NOISE MODEL>> that includes our amended text.

Please clarify:

Are the mean prior fluxes the true fluxes + bias or the true fluxes + bias + random flux error realization? The text sounds like the latter, but in figures 4 and 5, the variations in the prior look very close to those of the true fluxes. Given that the random error in the fossil fluxes is the same as the bias, and the correlation length is small compared to the satellite swath/sweep, I would expect to see differences in the structures more clearly, i.e. on the order of the bias between prior and truth.

Yes, it is the latter (true fluxes + bias + random flux error). Since prior flux noise is proportional to the flux signal of the true emission, there is more noise in the city centre and smaller errors decreasing towards the edges of the centre. The structures of flux errors are showed in Figures A6a and A7a. Since most pixels have very low true emission (less than 1 µmol m^{-2} s^{-1}), their random flux errors, which are scaled by the emission signals, are smaller. That is why the variations in the prior look very close to those of the true fluxes in Figures. 4 and 5.

More information on the state vector is needed:

- How are anthropogenic emissions inferred from the posterior fluxes? Are individual flux components (e.g. anthropogenic, GPP, respiration) optimized as separate state vector elements? Or are total emissions optimized and then partitioned (how?) into the individual flux components?

- On what resolution are fluxes estimated, i.e. what is the size of the state vector?

- Is there a temporal resolution?

- A new section dedicated to the state vector could be helpful for structuring this additional information.

Thanks for pointing these out. We have revised the manuscript to clarify these points:

"We optimize individual flux components as separate state vector elements. The spatial resolution of fluxes is 1 km × 1 km and the temporal resolution is monthly. In Paris, the size of the state vector is 10080*1 for only fossil component and 20160*1 for including fossil and biogenic fluxes. The introduction of biogenic fluxes doubles the size of the state vector because we optimize them separately."

3.3 Biogenic fluxes

Upon first reading the manuscript, only here did I realize that results so far only considered the anthropogenic flux component. In hindsight it's obvious, because figures 4 and 5 don't show negative fluxes/enhancements. Nonetheless, this approach is unexpected and needs to be made *clearer throughout the manuscript, e.g. in the introductory paragraph of Sect. 3 (for example, there are 18 scenarios, not 9), in subsection headings, and in figure captions (and perhaps in Sect. 2.7).*

Thanks for identifying this point. We have revised the manuscript to add this clarification:

"We firstly only include the fossil component and focus on estimating fossil fuel CO_2 emissions (Sections 3.1 and 3.2), and then extend the inversion system to include the biogenic component (Sections 3.3 and 3.4)."

I believe that the interpretations of the results in this paragraph (lines 253ff) are not accurate. The Bayesian posterior random flux error is independent of enhancements (see Eq. (4)), so smaller enhancements do not, as the authors state, limit the ability to reduce the random flux errors (note that the posterior flux is of course not independent of the random error or enhancements). What actually limits the ability to retrieve the anthropogenic component should be uncertainty in the biogenic fluxes, not, as the authors state, the fact that they are negative in summer. Since the authors set the uncertainty of the biogenic fluxes at 25% of the flux, it's higher in summer (according to Fig. A2), so the inversion dumps more corrections into this component then (assuming the different flux components are optimized separately, which, as mentioned above, is not clear). This should explain why the posterior anthropogenic fluxes get closer to the truth (Fig. 8) and why the random uncertainty reduction is better without bio fluxes (Fig. 9), especially in summer.

We agree that the random flux error is independent of urban enhancement signal that is not affected by the introduction of biogenic fluxes. However, in our study we divided the flux errors into systematic errors (bias) and random errors. The flux bias of the fossil component is not independent to the enhancements. The bias correction of the fossil component is related to negative biogenic fluxes, while the random flux error of the fossil component is only related to biogenic flux uncertainty. For the random error correction of the fossil fuel CO_2 emissions (Figure 9), the random error of biogenic flux is important, and the impact of negative fluxes on urban enhancement is independent with this part. For the bias correction of the fossil fuel CO_2 emissions (figure 8), the presence of negative biogenic fluxes is important since it reduces the enhancements and further reduce the signal-noise ratio of urban enhancements. In short, both the random error of biogenic fluxes and their uptake effect on urban enhancements can make impact on the accuracy and precision of anthropogenic CO_2 emissions, both can degrade our ability to infer urban CO_2 emissions.

We have clarified this issue in the revised manuscript and added the above explanations.

"Although the random error correction of the fossil component is only related to the random error of the biogenic component, we find that biospheric uptake reduces the signal to noise ratio in satellite measurements (Figure 7d), which limits the ability to reduce bias in the *a priori* anthropogenic CO_2 emissions. The degree of the influence from biosphere is subject to the uncertainty of *a priori* biogenic fluxes, the separation of anthropogenic signals from net CO_2 measurements, and the scale of inverted fluxes."

TECHNICAL CORRECTIONS

Recurring

- Please include the DOI in all references that have one
Not sure if we need to do this.

Line 10: I think it would be clearer if the structure of the sentence were switched around, i.e.:

The three-sweep observing strategy, which generally outperforms the two-sweep mode by virtue of its wider scan area that typically yields more cloud-free scenes, can retrieve the total emissions of the truth within 7% over Paris and 21% over London.
Done

Line 14: See comments on Sect. 3.3. I think it's the random error, not the fact that the biospheric signal is negative.
Thanks. For the random error correction of the fossil component, it is related to the random error of the biogenic component and not the negative flux signals. For bias correction in the fossil component, it is related to the uptake effect of biogenic fluxes. We have revised the manuscript to clarify this point.

Line 19: Is there a better peer-reviewed reference for this statement? Bulkeley 2013 is not peer-reviewed.
Done

Line 29: I'd remove "integrated".
Done

Line 35: "subject to" doesn't really fit here - perhaps "but they are subject to"
Done

Line 45: "region thereby" -> "region, thereby"
Done

Line 52: "continuity of collecting ..." -> "continuity with other satellites collecting ..."
Done

Line 61: I'd explicitly write out "TCCON"
Done

Line 68: "analyses" -> "analysis"
Done

Line 92: "still exist" -> maybe you mean "may not be predicted"?
Done

Line 98: "is defined as"

Done

Lines 90-106: Add somewhere in the beginning something along the lines of "We follow the approach by Palmer et al. (2011) and briefly outline the method here" - the section is copied (partly verbatim) from there.

Thanks, we have revised the texts.

Line 143: In the comparison of Eddy data and SMUrF results, please include the cumulative NEE, over one year or perhaps during the months analyzed in the study, alongside r and slope.
Done, we have added the following text:

"The seasonal-mean NEEs simulated by SMUrF in July 2018 agree reasonably with the eddy-covariance flux data (slope=1.07 for the ensemble mean of all sites and the cumulative seasonal-mean NEE is -12 g m^{-2} day^{-1} for the simulated flux and -10 g m^{-2} day^{-1} for the flux data) indicating that the simulated biogenic CO_2 fluxes are in a realistic range."

Lines 150ff: State how long back in time the backtrajectories are.
24 hours, which has been clarified in the manuscript.

Line 189: "Realistic 20% random error": I don't find this number in a quick scan of either of the cited references. Please elaborate a bit on how you arrive at 20% to represent transport error.
This is an educated guess based on previous studies led by us and others.

Line 197: Please also provide the bias in terms of percentages of the fluxes.
Yes, we have included it in the revised manuscript.

Line 200: "closer to one" -> "close to one"
Done

Line 202: Similar to line 189: Please explain briefly the reasons behind the choice of 10km as correlation length. A quick look at the cited references might not make it obvious to the reader.
It is an educated guess. The estimation method is related to these references.

Line 215: "Figure 4" -> "Figures 4a, b, e and f"
Done

Lines 281f: Would the results be better if the position of the city scan were adjusted to the wind direction? I.e. scan further South when the wind is from the North? You could add this to the conclusions.
It is more about where the wind comes from on the sampling day (the emissions on those areas can be optimized by the data), not related to the scan direction.

Line 282: See comments on Sect. 3.3. I think it's the random error, not the fact that the biospheric signal is negative.
We addressed this comment on Sect. 3.3 and have revised the manuscript to clarify this issue.

Lines 285ff: This should be a major conclusion, alongside the call for measuring additional species.
Agree.

Lines 296f: "Quantifying sub-city scale emissions requires improving the accuracy and precision of satellite-based CO2 measurements." - Does this refer to the MicroCarb skill in this study, or is it a general statement?
It is a general statement for all satellite measurements.

Fig. 2:

- Indicate the overpass direction
Done, we have added it in the caption. "The satellite moves from northeast to southwest and wind comes from southwest on the day."

- Caption: "halve the cloud cover from one to 0.5" is unclear to me. Sounds like cloud cover up to 50% gets a pass, which shouldn't be the case
- Update to this comment after reading a bit further: It seems that this phrase refers to the artificial halving described in line 109. If so, the phrase is confusing in the caption of Fig. 2 and I would either leave it out or refer the reader to Sect. 2.2.
Done, we have deleted it.

Fig. 5:

- Just a comment - the figure is really helpful!

Fig. A1:

- Caption: "red star marks" -> "red stars mark"
Done

Fig. A6:

- Do these figures only include anthropogenic or also biogenic fluxes?
Only include anthropogenic fluxes, which has been revised in the updated manuscript.