Responses to Referee#1 comments:

General remarks

The paper compares 5 data-driven methods to estimate local CO_2 emissions from satellite images as an extension of earlier studies (e.g. Kuhlmann et al. 2021). The images including clouds were generated using a regional atmospheric transport model and known emissions from power plants and a city with output sampled as for the satellite in construction. In general the paper is well written and contains a lot of details.

We thank Referee#1 for this positive feedback, for his insightful comments and for thoroughly reading the manuscript, which allowed errors to be corrected.

Concerning the cloud effects the language is sometimes not clear or too lengthy. It should be clearly said that in the 'cloud-free' scenario (section 2.3) modeled clouds are ignored and all pixels used for analysis, right? It might be interesting to include a remark what will happen for the accuracy of the methods if a cloudiness threshold of 2 or 5% is selected which might be more typical for real images.

We agree with the referee/s remark and rephrased the sentence describing the cloud-free scenario and add a new paragraph in the text to describe the impact of the cloudiness threshold on the results:

Old sentence:

The most optimistic or ideal scenario considers that inversions are performed with CO_2 and NO_2 cloud-free data using directly the winds from the COSMO-GHG simulations (SMARTCARB winds). It is the ideal case because 1) with the inclusion of NO_2 data, the data constraints on the estimates are stronger than when using CO_2 data only; 2) the absence of clouds maximizes the number and quality of the estimates, and 3) the winds are perfectly consistent with the data as they were used to simulate the XCO_2 and NO_2 fields.

New sentence:

The most optimistic or ideal scenario corresponds to the application of inversions to CO_2 and NO_2 images without the removal of pixels associated to cloud-cover (ignoring the clouds modelled with the COSMO-GHG model; we label such inversions "cloud-free" hereafter), and with a perfect knowledge of the wind field (i.e. using directly the winds from the COSMO-GHG model, denoted SMARTCARB winds). It is the ideal case because 1) the joint analysis of NO_2 and CO_2 images strengthen the estimates compared to the analysis of CO_2 images only; 2) ignoring the potential loss of data due to cloud cover in the CO_2 and NO_2 images yield full images, whose analysis is more robust than that of partial images, and thus provides a higher number and precision of estimates.

Concerning the impact on the performance of the methods of increasing the cloudiness threshold to 2% or 5%, we rewrote the last paragraph of section 3.3. (Impact of the cloud cover) and we added the figure A3 to illustrate this impact on the LSCF method

New paragraph:

Furthermore, the filtering of data removing those with a significant cloud cover not only affects the number of estimates but also impacts the performance of the methods, although to a much lesser extent. When comparing results obtained from the same images, cloud-free inversions produce slightly better results than cloud-filtered inversions (Fig A2). This is because, in images partially masked by cloud cover, some pixels containing useful information are likely removed, which can lead to less accurate determination of emissions. Consistently, if the threshold of cloud cover above which XCO_2 pixels of the image are discarded for the analysis is increased to 2% or 5%, the performance of the methods does not increase significantly, unlike the number of estimates, which can increase, e.g. by 12% and 29% respectively when using the LCSF method (Fig. A3).

New figure A3:

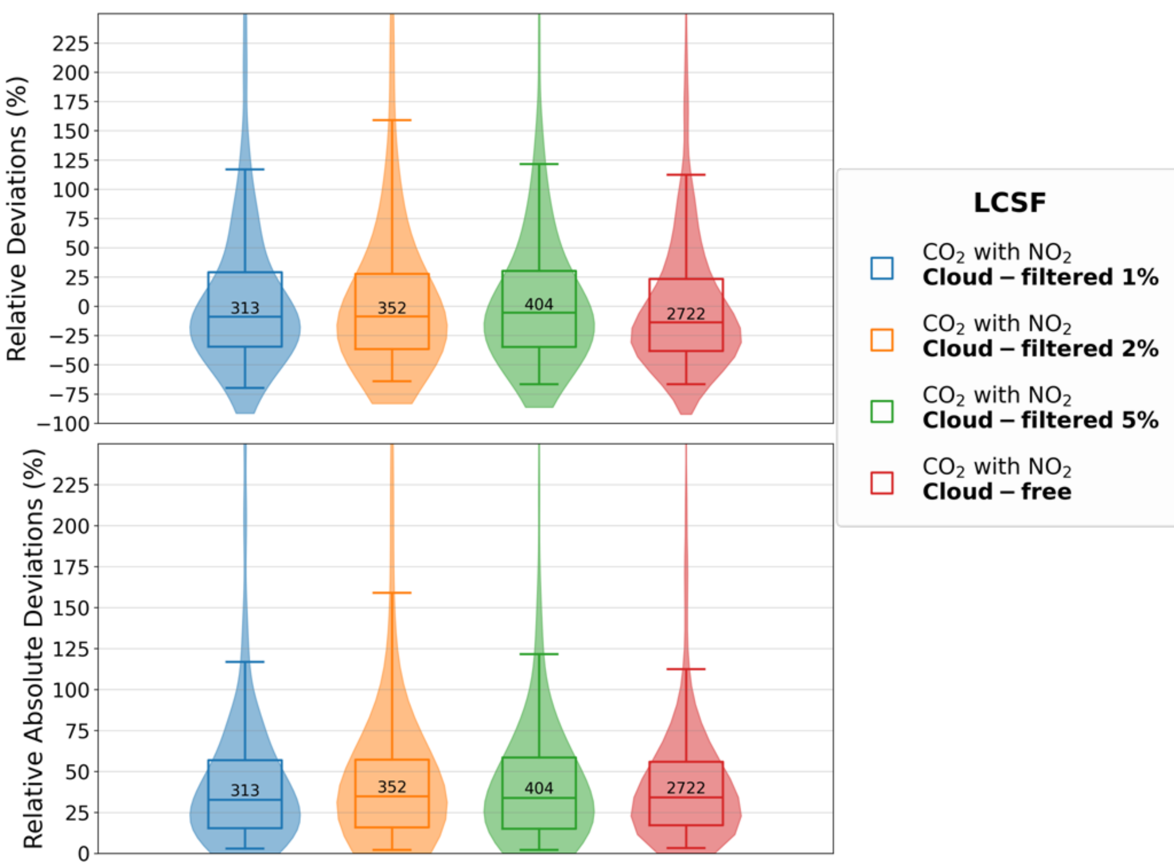

Figure A3: Performance of the LCSF method when estimating emissions from single images of CO2 and NO2 without considering clouds (in red) and for different cloudiness thresholds: 1% (in blue), 2% (in orange) and 5% (in green). Distributions of the relative deviations (top panel) and relative absolute deviations (bottom panel) are illustrated using violin plots. Boxes are the inter-quartiles of the distributions, the whiskers are the 5^{th} and 95^{th} percentiles, and the lines

within boxes are the medians. Numbers in the inter-quartile boxes are the number of estimates for each benchmarking scenario.

A companion paper in preparation (Kuhlmann et al., 2023) is not a proper reference. Include only a short remark in parentheses on that in the text or refer to the code repository.

The paper has been recently published. We updated the reference accordingly:

Kuhlmann, G., Koene, E. F. M., Meier, S., Santaren, D., Broquet, G., Chevallier, F., Hakkarainen, J., Nurmela, J., Amorós, L., Tamminen, J., and Brunner, D.: The *ddeq* Python library for point source quantification from remote sensing images (Version 1.0). Geoscientific Model Development, 17(12), 4773-4789, https://doi.org/10.5194/gmd-17-4773-2024, 2024.

In the section "code and data availability", we give also the address of the code repository (`https://gitlab.com/empa503/remote-sensing/ddeq`)

Some acronyms are defined several times in abstract and text, including section headers. Usually it should be defined only at the first occurrence.

We assume that this comment mainly refers to the definition of acronyms both in the abstract and in the introduction. However, we think that the main text cannot rely on definitions made in the abstract, e.g. see the point 4. in https://www.lithoguru.com/scientist/litho_papers/JM3%20editorial%202012%20q4_Acronyms.pdf.

Therefore, we prefer to keep these definitions in both the abstract and introduction in this new version of the manuscript, and let the editor rectify this if needed.

Specific remarks

Line 13 and 47: Conflicting definitions of an acronym.

We keep the definition of CO2M as "Copernicus CO2 monitoring mission" (Line 13). We then remove the definition at line 47 and only keep the acronym.

Line 16: Insert 'mole fractions' or 'volume mixing ratio'.

As suggested, we insert in the text a proper definition of XCO2: "To support the development of the operational processing of satellite *column-averaged CO_2 dry air mole fraction* (XCO_2) and NO_2 imagery.

Line 17: Is that 'tropospheric column NO_2' like later in the text? Please be consistent.

Yes, for sake of simplicity, tropospheric column NO2 is referred to later in the text as NO2.

This simplification is now mentioned in the introduction:

Each satellite will carry an imaging spectrometer providing images of XCO_2 and of NO_2 tropospheric column densities ***(referred to as NO_2 hereinafter)*** *along a 250 km wide swath with a resolution of 2 km × 2 km (Sierk et al., 2019).*

Line 29ff: This long sentence is difficult to understand and should be split and improved for clarity.

We rephrase this sentence as:

The GP and the LCSF methods generate the most accurate estimates from individual images. The deviations between the emission estimates and the true emissions from these two methods have similar Interquartile Ranges (IQR): between ~20% and ~60% depending on the scenarios.

Line 55 and 108: Define acronyms.

TANGO acronym is defined as "Twin ANthropogenic Greenhouse Gas Observers" and ESA acronym as "European Space Agency".

Line 199: This is in strong contrast to the swath of CO2M (line 14). Add a remark please.

As suggested we add a remark at the end of the sentence:

"It is derived from the method originally developed by Zheng et al. (2020). This original method was designed to estimate the CO_2 emissions of cities and industrial areas in China that produce atmospheric plumes clearly detectable in the relatively narrow transects of OCO-2 observations. These observations are characterized by a resolution of few km^2 but by a swath about 10 km wide, which is almost 25 times narrower than the ~250 km wide swath of the CO2M instruments."

Line 220: Because of typical stack heights?

The answer is a bit complex. In case of power plants, the effective injection height of the plumes (accounting for the stack heights and plume rise) would tend to be higher than 100m. However, for the different sources in the cities, it ranges from the surface to such heights, and on average, it would tend to be lower than 100m (Brunner et al., 2024, https://acp.copernicus.org/articles/19/4541/2019/). We would thus keep this choice of the 100m height as a pragmatic choice based on preliminary sensitivity tests (actually, it raises results that are very close to those when using 200m or 300m height). We have now added to the text the following indication

"This result may be reflecting a trade-off between the need to account for emission injection heights higher than 100m when considering isolated power plants, and lower than 100m when considering the mix of sources within cities, whose emissions are not dominated by large power plants (Brunner et al., 2024). The automatic process of sources limits the ability to derive a case by case selection of the height for the wind extraction, but a finer option for future analysis might be to discriminate this selection as a function of the type of target (considering, at least, isolated power plants vs. urban areas)."

Line 255ff, 275 and 280: Better use equation style with separate lines, don't repeat parts of an equation.

The text was reformatted as suggested.

Line 257: This differs from the recommendation for LCSF.

Yes, indeed. The practical derivation of the effective winds can be different between the specific versions of the different methods. Varon et al. derived an estimate of the effective wind based on 10 m wind speeds probably because they generalized cases where the analysis are supported by local meteorological stations providing near surface measurements for such a variable. As explained above, the choice made for LCSF arose from the assumption that the analysis would rely only on meteorological reanalysis, so that the effective wind should be extracted at a higher altitude, and from a pragmatic analysis of preliminary sensitivity tests. Finally, here, the effective wind is derived using the same practical computation in the version of the IME and CSF methods, i.e. using vertically GNFR-A weighted averages. This is mentioned a few lines after in the text. To emphasize the fact that the choice of an effective wind based on 10 m wind speeds is specific to the version of the IME method in Varon et al., we added "*For example*" at the beginning of the sentence.

Line 397: Also more combustion of fossil fuels for heating (cities).

As suggested, we rephrased the end of the sentence:
…emissions are higher during winter due to increased fossil fuel consumption associated with electricity and heat production.

Line 608: Refer to Fig.4.

Added as suggested.

Line 804: Filtered for clouds or what?

No, we meant the estimates filtered based on their uncertainties. We clarify by changing: “*when estimates are not filtered*” by “*when estimates are not selected based on their uncertainty*”.

Table 2: 'Cloud fraction' is confusing here. In meteorology a cloud fraction of 100% means overcast sky and not clear sky or cloud free (scenario 1, 5, 0%?). If you like to say with a 'threshold of 100%' that model simulated clouds are ignored as in Kuhlmann et al. (2021) or line 324 please say that in caption or footnote.

As suggested, we added to the caption the sentence:

Note that a cloud fraction threshold of x% corresponds to the rejection of data pixels if the cloud cover exceeds x%, so that a cloud fraction of 100% yields full images without a loss of data pixels.

Technical corrections

Line 342: typo.

Corrected as suggested: covers -> cover

Line 360ff, Fig. A3: The unit is 'ms-1' (or 'm/s'), remove '.'.

Corrected as suggested.

Line 371: 'considering or ignoring the cloud cover'

Corrected as suggested.

References: Use consistent style for the year of publication.

Indeed. We put the year of publication at the end of all the references.

Line 974ff: Is there a preprint (URL) available? Status?

This paper has been recently published, we updated the reference accordingly:

Hakkarainen, J., Kuhlmann, G., Koene, E., Santaren, D., Meier, S., Krol, M.C., van Stratum, B.J.H, Ialongo, I., Chevallier, F., Tamminen, J., Brunner, D., Broquet, G. Analyzing nitrogen dioxide to nitrogen oxide scaling factors for data-driven satellite-based emission estimation methods: a case study of Matimba/Medupi power stations in South Africa, Atmospheric Pollution Research, Volume 15, Issue 7, 2024, 102171, ISSN 1309-1042, https://doi.org/10.1016/j.apr.2024.102171, 2024.

Skip lines 991 to 994 (twice!).

Corrected as suggested

Line 1018f: In the doi of this important paper '689838' is missing at the end.

Corrected as suggested

Line 1020ff: This is not a reference. Status?

The paper has been recently published, we updated the reference accordingly:

Kuhlmann, G., Koene, E. F. M., Meier, S., Santaren, D., Broquet, G., Chevallier, F., Hakkarainen, J., Nurmela, J., Amorós, L., Tamminen, J., and Brunner, D.: The ddeq Python library for point source quantification from remote sensing images (Version 1.0). Geoscientific Model Development, 17(12), 4773-4789, https://doi.org/10.5194/gmd-17-4773-2024, 2024.

Figure 3: Shouldn't it be [ ] in the legends?

Corrected as suggested

Figure 4, 9: Mention the meaning of the numbers in the inter-quartile boxes in the caption (number of estimates?).

Yes. We have updated the captions accordingly: …***Numbers in the inter-quartile boxes are the number of estimates for each benchmarking scenario and inversion method.***

Figure 5, 6, 7, A4, A7: Remove '.' in units at label, better write 'CO_2 emissions (Mt yr^{-1})'.

Corrected as suggested. Captions were also corrected.

Figure 6, 7, A3, A4, A7, A8: Jänschwalde! Correct spelling!

Corrected as suggested

Figure 8, A5, A6: Legend and caption inconsistent concerning percentiles.

Captions corrected: 90th percentiles -> 95th percentiles

Table 1: What is 'mn'? Should it be 'min' for minute?

Corrected as suggested.