Review of "Optimal selection of satellite XCO_2 images over cities for urban CO_2 emission monitoring"

Alexandre Danjou[1], Grégoire Broquet[1], Andrew Schuh[2], François-Marie Bréon[1], and Thomas Lauvaux[1,3]

[1]Laboratoire des Sciences du Climat et de l'Environnement (LSCE), IPSL, CEA-CNRS-UVSQ, 91191 Gif sur Yvette, France
[2]Cooperative Institute for Research in the Atmosphere (CIRA), Colorado State University, Fort Collins, USA
[3]Molecular and Atmospheric Spectrometry Group (GSMA) – UMR 7331, University of Reims Champagne Ardenne, 51687 Reims, France

Correspondence: Alexandre Danjou (alexandre.danjou@lsce.ipsl.fr)

I would like to thank the authors for their revisions, which improved the descriptions and the understandability of the study. However, some of the authors' responses to the reviews need more explanation and discussion in the manuscript which is why I still recommend major revisions before publication. The comments can be found below. The line numbers refer to that of the revised manuscript.

Thanks for those encouragement and compliments. We realize that we might have misunderstood some of the reviewer comments in the first review and thanks the reviewer for taking the time to come back on them.

Re their responses on the reviewers' comments

The authors asked in their responses where in the manuscript it is written that the method is not universally applicable to all cities and is still work in progress. The answer to this is e.g. in line 344: "However, despite the application of this criterion, the variability of the error distribution remains large across cities." and in the conclusions line 559: "These significant remaining biases raise the question of the current reliability of the results obtained on a single given city." So the comment by Reviewer 2 is still open: "However, [the manuscript] should also describe how the proposed method would be applied to real satellite measurements and critically assess, whether it's applicable at all within the error budget."

Thanks for the clarification. We didn't understood the first time what meant the reviewer but do now : indeed the methods are not sufficiently reliable to be applied blindly to actual data as things are. We add some sentences in the conclusion concerning the further work that we think necessary to develop this method and thus, we hope, answering the reviewer comment : "Future work should focus on determining the types of information that can be reliably derived considering the current error estimates (e.g. annual emissions budget, trend detection, ...) along with the required number of images/plumes following Kuhlmann et al. (2019). In parallel, applying this sensitivity analysis to actual satellite data, similar to the synthetic images used in our study (e.g. OCO-3 SAMs), would help to evaluate and to refine the criteria derived here."

Re response by the authors about the interpolation methods used: There is no "classic" interpolation as mentioned by the authors in their response to the reviewers' comments. So which method is used? Please add this information somewhere around line 140.

The information has been added by replacing the following sentence "The results of these simulations are then projected onto a regular grid at approximately $1\text{km} \times 1\text{km}$ *resolution. This is done to simplify the analysis of the model outputs." by the sentence "The fields are then horizontally regridded to* $1\text{km} \times 1\text{km}$ *using rasterization techniques where the center of the 1km by 1km grid cell is mapped onto the hexahedral grid average which contains it. While the mapping isn't strictly mass conserving, the errors should be relatively small and since this is a post-hoc operation, errors do not accumulate during the simulation."*

Around Fig. 4 and its results: The threshold of 2.1 ktCO2/h is only based on the city emissions from the inventory. But there is no discussion whether the current satellites are able to measure emissions in this order of magnitude. Since the manuscript is based on OCO-3 and CO2M (e.g. random noise according to the precision requirement of 0.7 ppm of CO2M) there has to be a discussion somewhere in the manuscript which is the lower limit of emissions that is observable by the satellite. Otherwise, this threshold is fine in theory but not at all applicable in practice. For instance, typical parameters could be used to convert the emission of 2.1ktCO2/h to XCO2 enhancement using your equations as done in the literature, e.g. [1], [2]. If this is larger than the noise of the instruments (0.7 ppm), then it is fine, if not, the emission corresponding to a 0.7 ppm enhancement has to be used as the emission threshold in the study.

[1] https://amt.copernicus.org/articles/4/1735/2011/amt-4-1735-2011.pdf

[2] https://www.sciencedirect.com/science/article/pii/S0959652623006832

We still think that the notion of detection limit is irrelevant with our methods. But we think that this second mention of the reviewer to this shows some unclearness from our side. However, lacking any distance, we are unable to identify which points lack clarity. We thus try to expand our answer hoping that the explanation will make our point clearer and help us, with the help of the reviewer, identify the point that were missed. This (long) answer focus first on how we took into account instrumental noise in the threshold determination (1) , then on the questionable notion, in our particular case, of "detection limit" (2), before presenting the analysis asked by the reviewer (3). We would be interested to hear more from the reviewer on this issue, as it is a major point in our analysis.

(1) The threshold of 2.1 ktCO2/h is determined by the decision tree, which is based on the results of analysis of our synthetic images. These synthetic images include noise (our image is the sum of three "signals": the city plume, background concentrations -other anthropogenic sources, biogenic sources, mesoscale variations- and 0.7ppm noise) so this criterion is determined by taking instrumental noise into account. This value is thus a lower limit of emissions that are quantifiable with "acceptable" precision by the satellite.

(2) We would like to remind the reviewer that the issue when we target cities is not plume detection, but the capacity to perform emission quantification. Indeed, we do not need to detect the plume : we know where the city is and we know the wind direction, which means that we can know the position of the plume without looking at the image as shown in Danjou et al. (2024)'s sections on plume detection methods. Those sections shows that emission estimations done with an a priori defined plume are way better than those done with a plume detection algorithm. Thus, a close and more relevant notion would be "quantification limit" : limits under which we cannot properly quantify the emissions. Our analysis tends to show that a more

relevant criterion than emission budget for quantification limit is wind stability in the PBL (with a criterion on the spatial variability of wind direction, or wind speed), cf last paragraph of section 5.2.1.

(3) The analysis proposed by the reviewer with the Gaussian model is, from our point of view, a simplified version of what we have done here :

- *"typical parameters could be used to convert the emission of 2.1ktCO2/h to XCO2 enhancement using your equations as done in the literature, e.g. [1], [2]" : rather than using a Gaussian model, we do this with a transport model with complex meteorology, and for several emission values;*
- *"If this is larger than the noise of the instruments (0.7 ppm), then it is fine, if not, the emission corresponding to a 0.7 ppm enhancement has to be used as the emission threshold in the study." : rather than starting out with an a priori view of the result (the main limitation of quantification methods is instrumental noise, and we need to find a law to determine a threshold emission limit), we use a learning method to determine these criteria in a broader context by including several parameters in our analysis.*

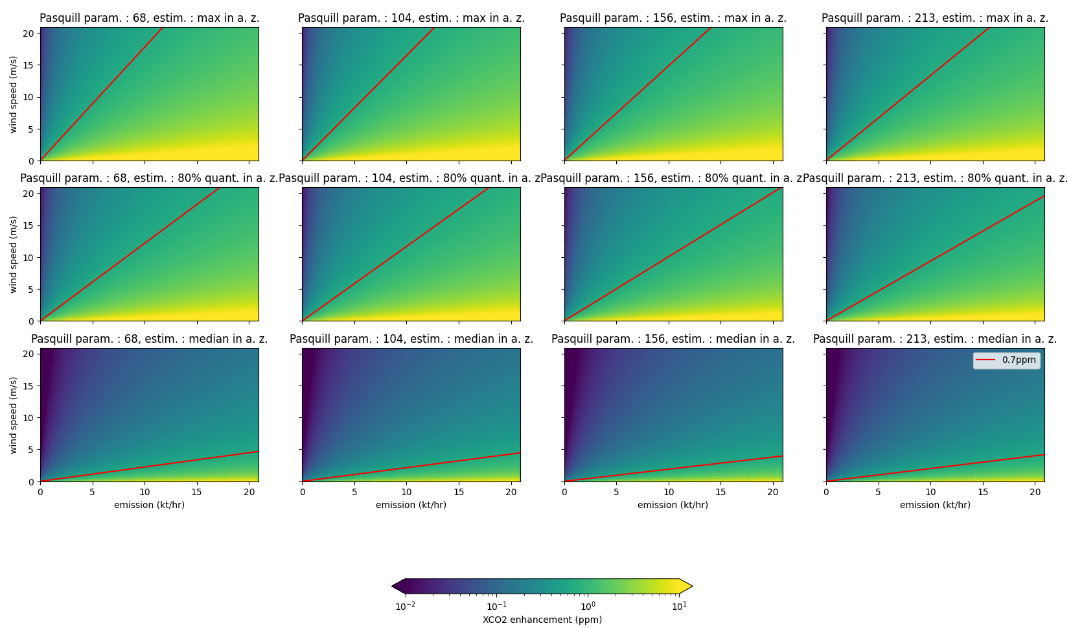

Figure 1. Maximum (first line), quantile 80% (middle line), median (third line) enhancement simulated by the Gaussian plume in the analysis area (see section 3) as a function of wind speed (W) and emissions (Q) for different Pasquill parameter values. The red line delimits the area where enhancement is greater than 0.7ppm.

Using the equation given as an example by the reviewer (used in the 2 articles but also by our inversion model) also shows that an enhancement bigger than 0.7ppm will be the result of a law involving emissions and wind speed ($XCO_2 \propto Q/|W|$, *see Section 3 for notation) and not just a threshold on sole emissions. An illustration of the law linking ppm increase and wind speed/emission according to the Gaussian model is shown in figure 1.*

We can confirm this first-order proportionality between the enhancement and the emission/wind speed ratio. The definition of detectable enhancement is, however, not obvious: should we base it on the median XCO2 signal in the plume, the maximum, or a specific quantile? We can see that the number of accepted cases would greatly vary from one metric to another.

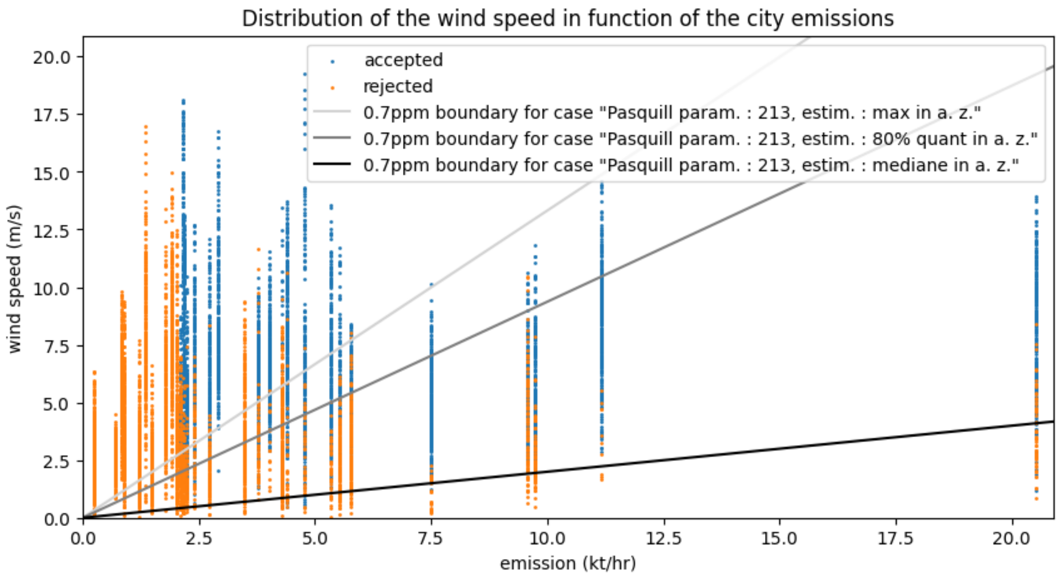

Figure 2. Distribution of synthetic images according to the average wind in the PBL at the time of the image and the city's emissions. A differentiation is made between cases accepted (blue) according to the criteria defined by the decision tree and those rejected (orange). The grey lines show demarcations at 0.7ppm for different metrics characterizing the enhancement and a Pasquill parameter of 213. These demarcations and axes are the same as those shown in figure 1.

Figure 2 shows the distribution of the 9,119 cases as a function of wind speed and emissions, with a differentiation between those accepted and those rejected according to our decision tree method. We can see that the decision tree criteria have no connection with the 0.7ppm lines of the previous graph (in grey on this figure).

Figure 3 is much the same as figure 2, except that the x-axis is now in logarithmic scale (for greater visibility) and the color of the dots now indicates the error on the emission estimate (in %). We can see that cases whose enhancement is greater than 0.7ppm according to the Gaussian plume model present errors of more than 60% (red ellipse). Conversely, cases with enhancement of less than 0.7ppm according to the Gaussian plume model show errors of less than 40% (majority of cases in the blue ellipses). These zones correspond roughly to those given by the criteria of our decision tree method.

The aim of this article is precisely to delve a little deeper and go beyond the general idea of an urban plume detection threshold. For us, this comment shows that our approach may not have been fully understood, and therefore requires further

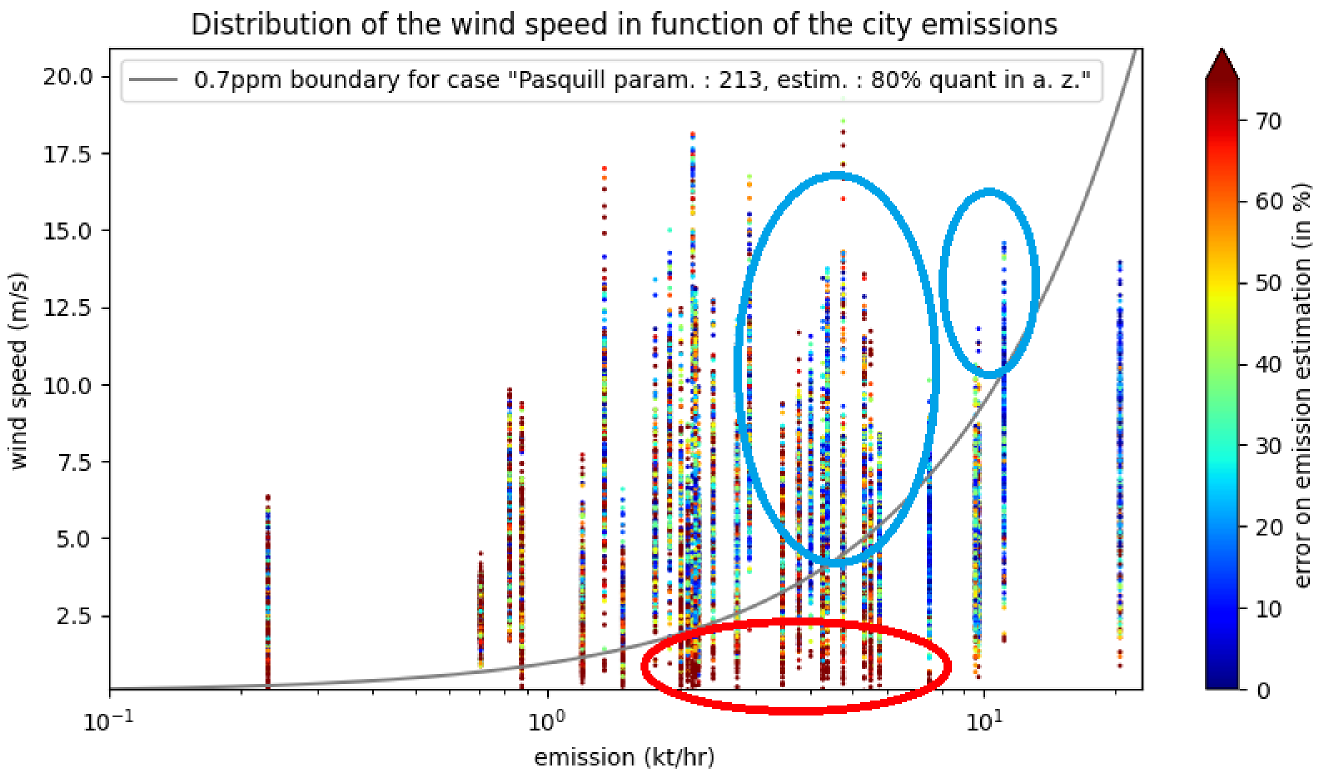

Figure 3. Emission estimation error for the 9,119 synthetic images in function of the average wind in the PBL at the time of the image and in function of the city's emissions. The grey lines show one example of demarcation at 0.7ppm from figure 1 and 2.

explanation. We thus have added a paragraph on the notion of "detection limit" at the end of section 6.3 and a sentence in the ante-penultimate paragraph of the introduction .

Sect. 5.1: Since this analysis motivates the use of the machine learning algorithm described in Sect. 4.2, I still think that it would be better to move this between Sect. 4.1 and 4.2. In addition, the reader has to always move back and forth between 4.1 and 5.1 to understand all the details mentioned in both sections. I understand the authors' argument that it is common practice to describe methodology first, but I think it is different when the methodology depends on previous results. And also from the point of view of the reader it would make the manuscript better understandable.

We still disagree with the reviewer on the solution to this issue. A third of section 4 (l.234-246, l.304-338) is common to both analysis. Separating this would require a huge amount of reshaping : move 5.1 between 4.1 and 4.2 as suggested by the reviewer, but also move 4.3, and rewrite introduction parts of sections 4 and 5 (half a page each).

Mostly, we fear that the development it would need to make the preliminary analysis stand alone would have to be important and make the preliminary analysis take a lot of place compared to its interest. For now the decision tree analysis occupies 6 pages, the preliminary one 2 pages and the common parts 2 pages. Reshaping would need to move the common parts in the *preliminary analysis and thus have 4 pages vs 6. Moreover, we think that then it will also be harder for the reader to make the separation between what is common to both analysis and what is different, making it thus less clear than it is now.*

We spent a lot of thinking on this both while writing the article and during the review. We agree with the reviewer that this choice of ordering is not perfect and make some points hard to understand, but we don't have any better solution. If the reviewer still disagree, we will make the changes in order to move on.

Another solution might simply be to remove the preliminary analysis which was merely here to introduce the concept of our analysis and make the decision tree purpose easier to understand. The preliminary analysis could then be put in annex under the title "Illustration of the sensibilities of the error on the emission estimation to different variables characterising the observation conditions and the inversion". But we do think that this preliminary analysis helps the understanding of the decision tree analysis and removing it from the main text would also hampers the clarity of the paper.

Fig. A1: I still think that because the description of this figure is in the main text, it should appear there (Sect. 2.3 or 2.4). Or at least there should be a text describing the results in Appendix A. Currently, it is just the figure without any description.

See answer to specific comment.

1 Specific Comments

- L13: In the main text it is always 2.1, not 12.1 ktCO2/h. *Typo corrected.*
- L79: Please add "about" before 2km x 2 km because it varies with the distance to nadir and the ratio will not be exactly a square. *Correction made.*
- L88: It is not clear what "realistic" means here. Is realistic in the sense of the OLAM model applied? Is it the grid spacing which is similar to the expected satellite footprint size? *Both. We added a parenthesis to precise our point : "(as obtained from a global non-hydrostatic atmospheric model with a maximum resolution of a few km)"*
- L138: PBL not defined. *We replaced "PBL" by " Planet Boundary Layer (PBL)" on l. 138.*
- Sect. 2: There has to be a note somewhere that your 40-day simulation is free-running because then the forecast skill after 14 days is small and the meteorology is not realistic (but of course consistent within the model). *We added a sentence ("The simulations are free running.") in section 2.2.*
- L149: Why are the dates different from the previous manuscript? Are the authors sure that the simulation did not start at 01 August? That would also be consistent with 8 images per day for 31 cities. *this is a typo, the correct begin date is indeed 1 August.*
- L154: "that expected for CO2M": This is only true if the city is in the center of the nadir swath of CO2M. Please clarify and rephrase. *Following the reviewer comment, we add extended the sentence to clarify : "This size is halfway between that of the OCO-3 images and the expected swath of CO2M in nadir mode."*

- L169: "most emitting pixels": As pixels cannot emit themselves I think the authors mean the pixels with largest emission sources. Please clarify and rephrase. *The sentence was rephrased into : "Within this* $50\mathrm{km}$*-radius disc, we select only a fraction (1/2.5*^{2}*) of the pixels, keeping those for which the emissions are the highest."*

- L286: "We thus obtain at most 4 subsets": It's not really clear where the number 4 is coming from. Is it because the decision tree has a depth of two and the whole set of 9119 images is split into two subsets in each step? A reference to Fig. 2 would be helpful here, I guess. *We change the sentence following the reviewer suggestions :"As the maximum depth of the tree is two, we obtain at most 4 subsets (see illustration of that case on Fig. 2) and select the one with the smallest Mean Absolute Error (MAE) on the emission estimate."*

- Fig. 2: "pseudo image" should be replaced by "synthetic image" here as well for consistency. *Correction made.*

- L315-317: This citation looks weird although I realize that this the recommended citation. Maybe this can be abbreviated in some way? *As pointed out by the reviewer, it is the recommnded citation. We didn't find any other ways of citing it in the literature.*

- L338: "the optimisation does not converge": Do the authors mean the minimization process described in Sect. 3? What is the criterion for convergence? *We replaced this bit by : "the optimizer used for the minimization described in Section 3 does not converge." The criterion for convergence is arbitrarily fixed. We set a tolerance of* 10^{-5} *times the root-mean-square-difference between the mass per unit area simulated by the Gaussian model (with the prior guess) and the "observed" mass per unit area. According to the optimizer documentation : "The iteration will stop when* $\max|\mathrm{proj}g_i| i = 1, ..., n <= \mathrm{gtol}$ *where* $\mathrm{proj}g_i$ *is the i-th component of the projected gradient" and* gtol *our arbitrarily fixed tolerance.*

- L369: What is GP2? *This is a remnant of the first review. Change to "our inversion method".*

- Fig. 4: replace pseudo-images by synthetic images *Correction made.*

- L443: It is not clear what is meant by "accuracy" here. Is it the bias or the spread of the distributions shown in Fig. 5? Please clarify in the manuscript. *Indeed, the term was not clear, we changed the sentence to : "We can see (figure 5) that the spread of the error on the emission estimation generally increases with decreasing emissions budgets."*

- Fig. 6: panel (a) still saying 11 degrees as the threshold. *Correction made.*

- L515: I think these percentages are the fraction of cities that pass all the criterions. I don't agree with the statement that their emissions are "easier to quantify than cities on other continents" because this is not relative to other continents. Please rephrase. *Indeed, it was misleading. We changed the sentence to "Asia and Australia stand out, with 37% (102 cities) and 40% (2 cities) of cities passing the criteria. Indeed, those cities, according to ODIAC dataset, are more likely to have emissions above our threshold."*

– L518: 0.7 ppm is the precision requirement for CO2M, not accuracy of OCO-3. Please clarify and rephrase or provide references here. *We changed the sentence to make it refer to CO2M. In the previous version, we had the article of Worden et al. (2017) in mind, which demonstrated a precision and accuracy of around 0.7 ppm for OCO-2 (which is equipped with the same instrument as OCO-3). But we agree that this is not indeed directly applicable to OCO-3 and changed the reference.*

– L528: This sentence still suggests that ERA-5 has only 37 vertical levels, which is not true. The authors should add here that they are using the ERA-5 product on constant pressure levels. By the way, it would be better to use the native resolution of the reanalysis with 137 vertical levels to ensure that the best resolution within the boundary layer is used, which is most relevant for this study. *The sentence was changed to clarify which product we are referring to. We also clarified that our main point is about the horizontal resolution of the data : "Indeed, the horizontal resolution of the weather product used here is very high around the cities of interest ($\approx 3\text{km}$ horizontally), higher than that of, for example, the ECMWF "ERA5 hourly data on pressure levels" product ($\approx 25\text{km}$). The vertical resolution is of the same order here and in the above-mentionned ERA-5 product (49 and 37 vertical levels)." About using the native resolution of 137 levels, it may indeed increase the resolution in the boundary layer, but our main point (and issue) is the horizontal resolution.*

– L585: It is confusing that 7 degrees are used here whereas in the main text, 12 degrees was found to be the threshold to be used. Is this an analysis that preceded the analysis of the main text? It would be helpful then if this is mentioned in the text. *A descritpion of the subsections was added at the beginning of the section. "Section B1 describes the inversion methods and the differences with the one described in the main text. Section B2 and B3 are construct in the same model as Sections 5.1 and 5.2 with for the first subsection a preliminary analysis (independent of the decision tree) and for the second the analysis of the decision tree method results."*

– Fig. A1 is not referenced anywhere in the text. Please add a reference somewhere and put the figure at this place. *A reference to the annex has been added in section 2.4 . We have also extended the annex (and put a reference in it to figure A1) with a more mathematical description of the process as it was subject of numerous comments in all reviews.*

– Fig. B1: Why is the GP2 line in panel b different from Fig. 3? They should be the same. If I understand it correctly, this figure is the same as Fig. 3 in the main text. If not, please explain what is the difference. *In figure 3, the x-axis for the spatial varability of the wind direction is inverted (goes from 1 to 0), whereas it is in the classic direction in the annex. We choose to invert the direction to put into light the closeness between the two curves. A sentence has been added in figure 3 label to highlight this : "Note that the x-axis is plotted in the direction of decreasing spatial variability of wind direction (i.e. inverse axis) and increasing wind speed."*

– Fig. B1: The graphs inside each panel overlap with them quite significantly which makes them difficult to read. There must be another way to plot this, e.g. by adding another figure or decreasing the size of the small graphs and placing them better in each panel. *We decrease the size of the incrusted panels, replaced them and make the dotted lines behind them semi-transparent. We think that the information are now more readable.*

– Fig. B1: The y-axis of the small graphs does not have a label so that it is not clear what is illustrated there. It is also not described in the figure caption. *We extend the figure caption and hope that it will be sufficient.*

– Fig. B1: Panel c has an x-axis different from the other panels. Please explain and add the correct label for the x axis. *We extend the figure caption to deal with this.*

– Fig. B1: The caption does not describe what is actually shown in the figure. Please add this information or at least a reference to the similar figure in the main text. *New caption : "Sensitivity of the emission estimation error to different variables of interest. For each subfigure, the main panel shows the evolution of the error distribution as a function of the quantile of the variable of interest: the solid line indicates the median, the dotted lines the 1st and 3rd quartiles, the highlighted area the quantiles at 15.9% and 84.1%. The small incrusted panel shows the values taken by the variables of interest for the different quantiles. Subfigure (c) is an exception : as we have only one value of emission budget per city, we plot the evolution of the error distribution as a function of the rank of the city regarding the variable of interest. The optimized radius shown in panel (g) is a parameter of the gaussian plume models (see section 4.3) and is therefore not calculated for the other methods."*

– L603 and L606: Please remove "very" as this it is not clear what makes it "very" instead of just similar / different. *Correction made.*

– L612: same comment as above: Why 7 degrees? *see previous answer.*

2 Technical corrections

– L120: typo: approximately *Corrected*

– L144: typo Generation *Corrected*

– L170: remove dot between "selection" and "is" *Corrected*

– L175: remove dash between New and York *Corrected*

– L224: have –> has *Corrected*

– L314: sur –> such *Corrected*

– L360: Supp.Mat. –> Appendix *Corrected*

– L535: Annex –> Appendix *Corrected*

- L544: enabless –> enables *Corrected*
- L579: add a "to" between "referred" and "as": referred to as GP2 *Corrected*

References

Danjou, A., Broquet, G., Lian, J., Bréon, F.-M., and Lauvaux, T.: Evaluation of light atmospheric plume inversion methods using synthetic XCO2 satellite images to compute Paris CO2 emissions, Remote Sensing of Environment, 305, 113 900, https://doi.org/10.1016/j.rse.2023.113900, 2024.

Kuhlmann, G., Broquet, G., Marshall, J., Clément, V., Löscher, A., Meijer, Y., and Brunner, D.: Detectability of CO_2 emission plumes of cities and power plants with the Copernicus Anthropogenic CO_2 Monitoring (CO2M) mission, Atmospheric Measurement Techniques Discussions, pp. 1–35, https://doi.org/10.5194/amt-2019-180, 2019.

Worden, R. J., Doran, G., Kulawik, S., Eldering, A., Crisp, D., Frankenberg, C., O'Dell, C., and Bowman, K. W.: Evaluation and attribution of OCO-2 XCO_2 uncertainties, Atmospheric Measurement Techniques, 10, 2759–2771, https://doi.org/10.5194/amt-10-2759-2017, 2017.