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)

We would like to thank both reviewers and the editor for the careful and detailed reviews and their patience during this process.

It seems to us while reading the reviews that the aim of the study and of the different analysis conducted were not sufficiently explained, which perturbed the reviewers and undermined the comprehension of the links between the different parts. We thus have made numerous changes on the introduction and took time to better introduce each sections. We have also changed the title as we found that it might give the impression that we used a model for our inversions, whereas we only use our model to generate our synthetic images. The description of the learning method has also been completely rewritten to make it clearer.

Multiple comments were in relation to the article Danjou et al. 2024. This article is now published and accessible (https://doi.org/10.1016/j.rse.2023.113900). Some of the references to it were unecessary and have been removed. We also tried as much as possible to gather the references and explain them when necessary. We hope that the article is now more self sufficient in regard to Danjou et al 2024.

1 Reviewer 2

The manuscript addresses an important scientific issue and presents an innovative approach to assess CO2 emissions from urban areas. However, many parts in the preprint suggest that it is work in progress. The study holds potential but necessitates substantial revisions to strengthen its clarity as well as scientific rigor.

We would like to thank the reviewer for this complimentary comment. We hope that the corrections made will address his concerns. We have rewritten the introduction and conclusion and added introduction parts to the section and hope that it had helped clarify the article.

1.1 General comments:

The manuscript should make a clear distinction that its primary objective is not the inversion process itself, but rather the assessment of the confidence in the inversion results via predictable and diagnostic variables. In the methodology section,

more emphasis could be put on how the two-step procedure via predictable and diagnostic variables contributes to the accuracy assessment of emission estimates. *Indeed, this was missing. We have developped the introduction of this section to make this point clearer.*

The manuscript heavily relies on Danjou et al. 2024, which is not accessible for verification. I think this is problematic and impedes a thorough evaluation in parts of this study.

It seems that there as been a problem. The manuscprit of Danjou et al. (2024) was provided to the editor so that it can be send to the reviewers. Please accept our apologize for this. The article has been published in RSE and is now accessible at https://doi.org/10.1016/j.rse.2023.113900.

The manuscript lacks clarity on what constitutes the error. The reference data used for determining the error is not explicitly defined. *A definition of the error as been added l.76 : "As we are working with synthetic data, the error in the emissions estimate is directly accessible by comparing the emissions estimated by the inversion method with the synthetic true emissions used in the OLAM simulations."*

The manuscript does not provide validation of its results with actual satellite data (e.g., OCO-3 SAM) or ground truth measurements. The absence of this validation makes the reliability of the simulated results vague to some extent.

The use of synthetic data is often used alone (Broquet 2018, Lespinas 2020, Kuhlmann 2019) to validate methods as it gives a good idea of the expected behaviour of these methods. There are indeed always differences between real and synthetic data, and differences in sensitivity will certainly be visible with real data, as there is still much to be understood in XCO_2 measurements at fine scale. Moreover, correlation unreprensented here can also appear that will highlight other criteria. Nevertheless, the criteria found here will still be pertinent with real data, are they will still be driving the error. From our point of view, comparing real and synthetic data (and their sensitivities) merits a study in its own right.

The variability of the error distribution remains significant across different cities. What are the implications for estimations from real satellite images? *It means, as stated in the conclusion, that there are still things that we do not understand, despite the progress made by defining these criteria. We have developed the conclusions to answer this question.*

The study should discuss the detection limits of current and future satellite missions and how those might impact the results. Is the purely random noise model imposed on XCO2 data in the study representative of real world atmospheric and environmental conditions? In actual scenarios, factors like surface reflectivity of different land types and the presence of aerosols can introduce more structured or systematic errors rather than purely random ones. Would incorporating more realistic, structured errors enhance the model's applicability and accuracy in real-world urban CO2 monitoring scenarios? Given this potential limitation, what implications does this assumption have for future research?

Understanding and simulating the error structure is an active field of research (see Bell 2023 https://amt.copernicus.org/articles/16/109/2023/, Taylor 2023 https://amt.copernicus.org/articles/16/3173/2023/amt-16-3173-2023.pdf, Worden 2017 https://amt.copernicus.org/articles/10/2759/2017/). We think that the complexity necessary to accurately take this into account (as can be seen in the papers cited) is out of scope of this study. Integrating this would only add complexity, increase the error and maybe add some new criteria, but the criteria that we find in this study should still be pertinent.

The selection criteria for the size of the target emission zone radius, are not comprehensively described. I can't find a clear rationale for the chosen size of the emission zone radius. Any potential to estimate this radius through an inversion approach? *Indeed, a clear explanation was missing. We have reshaped the first paragraph of this subsection and introduced a rationale.*

The influence of cloud coverage on satellite observations is a significant factor that the manuscript should address. *When it comes to determining which cities are most suitable for measurement, cloud cover is indeed a major issue, and the frequency of cloud cover is an important criterion for city selection. We examined this point in section 6.2.*

We have reserved this analysis for discussion, however, because realistically adding the effect of clouds in images (filtered pixels and contamination of neighbouring pixels) and trying to objectively quantify their impact on error is of a complexity that would merit a dedicated study to do it more properly than what is proposed in the discussion.

We also believe that, once an image is partly contaminated by clouds, it is not worth processing. Indeed, the effect of cloud presence on the measurement is very complex, due to the presence of 3D cloud radiative effects in OCO-2 retrievals (https://doi.org/10.5194/amt-14-1475-2021). Given the size of these effects (of the order of a few kilometers) and the size of our images (a few tens of kilometers), we doubt the value of trying to obtain an estimate from an image contaminated by clouds. However, this remains to be demonstrated, and the complexity of such a task requires, in our opinion, a complete study and has no place here.

Adding clouds would not alter the conclusions we have reached, the criteria identified should still be important even in the presence of clouds. Thus, adding cloud coverage would only increase the error and add another criterion.

The authors of the manuscript should revise Section 3 to summarize only the key aspects of the Danjou et al. 2024 study that are directly relevant to their current research. As mentioned above, it is problematic that the Danjou et al. 2024 paper is not yet available and that the study heavily relies on it. *Danjou et al. 2024 is now published. We tried to remove the unecessary references to this paper and regroup the ones that were unevitable. We hope that this article is now easier to read and less repetitive from that point of view.*

The description of the OLAM model and its simulations should be more comprehensive in order to better understand the simulation results. *Changes have been made following the specific comments of reviewer 1.*

The manuscript acknowledges that the method is not universally applicable to all cities and is still a work in progress. However, it should also describe how the proposed method would be applied to real satellite measurements and critically assess, whether its applicable at all within the error budget. *We do not share the conclusion of the reviewer. Could the reviewer indicate which sentences give this impression? Indeed a gaussian plume is applicable to all images and the criteria are also valid for every cities.*

The conclusion should reiterate the study's focus on developing a procedure for accuracy estimation, summarizing how the study advances this goal and its implications for future research and practical applications. *We have rewritten the conclusion, keeping in mind the reviewer advices.*

1.2 Specific comments

The manuscript frequently uses abbreviations and technical terms without defining them (e.g., XCO2, UNFCC, OCO, ppm, GOSAT-GW, WRF, OLAM, IQR). *definitions added in the relevant places.*

3: The phrase "selecting images to be processed" should be more clearly defined. *changed to "selecting the images to be processed"*

Fig 1: Identifying the factors behind the peak XCO2 values in the simulation domain deyond the city boundaries. *The two small peaks on the east of the city may correspond to an XCO2 enhancement due to the city that as accumulated earlier in the day and is now ventilated. The peak on the bottom of the images seems to be coming from the city at the edge of the image when we look at a more broader view.*

46: The Danjou et al. (2024) should be made available to the reviewers due to its significant relevance to the research. *It should have been. Sorry for this issue.*

53: Clarify local background signal (not clear to me, what is meant by this term). *it is the background signal in the vicinity of the city (please see Schuh 2021 for a more accurate description).*

56, 58: “This study” can be misleading. Write “The study”. *Changed to "Their study"*

48-50: Consider moving the sentence up to line 25, before you start describing the OCOs. *The mentionned sentences have been removed as they weren’t bring new informations compared to the ones at l.25.*

261: Set by the user? Do all produce the same value? *The description of the decision tree as be rewritten and we hope that it is clearer now:*

"4.2 Analysis with the decision tree learning algorithm

In this study, we seek to better understand the relationship between the input variables (predictable/diagnostic variables) and the reliability of an emissions estimate. For this, we train an explainable machine learning algorithm to predict the relative error of the emission estimate given some input variables (described in Section 4.3), like the variability of the wind direction or the emissions budget, and then study which variables are determined to be relevant by the algorithm. We choose a regression decision tree for this, as they work by learning simple decision rules and therefore are highly interpretable while able to find non-linear relationships between the inputs and the target variable.

4.2.1 Description of the decision tree learning algorithm

A decision tree is constructed following a recursive process: at each step, the algorithm splits the data into two subsets following a binary rule on a single variable, finding the split that best reduces a particular loss function on the target variable. Each subset is split further into two until some stopping condition is reached (see Fig. 2 for illustration). This algorithm therefore splits the input space into regions, where each region corresponds to a similar value of the target variable (i.e. the error on the emission estimation in our case). We use the regression tree implementation from the scikit-learn library (Pedregosa et al., 2011) with a squared error loss, and impose conditions on the algorithm to prevent overfitting (creating over-complex trees that don’t generalise well): we set the maximum depth of the tree to 2 (i.e. two levels of binary splits) and

we impose that the leaves must contain at least 10% of the training set. The training set (at the root node) is described in the following paragraph."

405-406: Please clarify why Paris exhibits such a low emission rate? Is this due to the absence of significant point sources? *Indeed there are no significant point sources in Paris. However we do not think that discussing the low emissions of Paris is relevant in this study, are we are working with synthetic data and are not interested in city in particular.*

418: What are the implications or consequences for the study if the dependencies of errors are not completely comprehended, even when using synthetic data. *This study brings objective criteria to evaluate the accuracy of an emission estimation, criteria that have not been evaluated as precisely until now to our knowledge. The error distribution of the emission estimation has not yet been (and cannot be?) fully characterized. Thus we will still have a non negligeable range of uncertainty on the bias given by an emission estimation for a certain city. However, the criteria found here remain pertinent and will still be pertinent when applied to real data. We have adapted the conclusion to emphasise this.*

Fig. 5: Typo in y-label (thrue). *ok*

316: The GP2 inversion method is presumed to be a variation of the Gauss Plume approach. However, at this point it is unclear for me what the '2' in 'GP2' represents. Please provide further details or clarification. *It is a typo remaining from an older version of the draft. There was a GP1 (a variation) that we discarded. We thought we had removed this naming in the main body. It is now done.*

520: Was the 12° threshold for wind variability found empirically or is there a rationale behind choosing it? Is this higher variability threshold, compared to what's mentioned in Danjou et al. 2024, a result of the increased resolution in the model? *The threshold of 12° is the one found in section 5.2 of this article with the tree learning method. We have rewritten this paragraph of the conclusion and , among other things, replaced l.518 "We were also able to determine precise and objective thresholds on these criteria to select the images." by "This analysis with a learning method also provides precise and objective thresholds to these criteria supporting the selection of images."*

The threshold that the reviewer refers to may be the 7° threshold defined in Danjou et al. 2024. It was empirically defined, which explains the difference. Addition of "(empirically defined)" l.317

578: The error appears to be highly sensitive to the city's radius. Could you clarify what "pseudo-image filtering" specifically entails? I don't have access to the Danjou et al. 2024 source for reference. Additionally, referring to line 550 and following, I guess it implies filtering out scenes with variability above a certain threshold? *This minimum seems to disappear after filtering and is not judge as an important critera by the decision tree. The values for the optimized radius seem to be driven by two things : the spatial variability of the wind direction and the shape of the city. Given the small number of cities, we think that this pic is due to a weird effect of the relationship between the two. Moreover, the statistically-speaking low number of cities does not help to have a smooth curve and increase the chance of having outliers. We rephrase l.551 and l. 578 to clarify and make the statements more self sufficient in regard to Danjou et al. 2024.*

58: What are pseudo-images? Synthetic 2D CO2 concentration images, I guess? *yes. Precise definition added l.51 : "synthetic satellite images (i.e. pseudo-images) of XCO_2 concentration generated with a meteorological model". We have replaced every occurences of "pseudo-images" with "synthetic images" for clarity.*

229: What is the primary source of error in the emission estimates? Does it originate from the tree model or the GP2 inversion method? *The tree model is not used to estimate the emission but to classify the emission estimates. The main sources of error are the estimations of the background estimation and the wind speed (cf Danjou et al. 2024).*

47: Improve the English language in this sentence. It encapsulates the primary motivation of the study and thus should be distinctly emphasized and articulated. *We do not understand what is wrong with this sentence as it seems to us grammatically correct.*

76: The frequent citation of Danjou et al. 2024 for all the "light" methods seems inappropriate. It would be more suitable to refer to the original papers specific to each approach. *It seems that the line given by the reviewer is not the good one. On line 74, we only cite Danjou et al. 2024 as we say that we are using the same methods as therein. We assume that the reviewer is refering to the paragraph between l.36 and 51 or the annex. We removed the unnecessary references to Danjou et al. 2024 in this paragraph and we added the relevant citations in appendix c1.*

74-79: The logic presented in this paragraph is unclear. It would be beneficial to revise and clarify the content for better understanding. *We have completely reshapen and rewritten the paragraph, with some additions to make it clearer : "The objective of our study is to resume the series of analysis from Wang et al. (2018); Schuh et al. (2021); Danjou et al. (2024) and deepen the evaluation of the conditions corresponding to reliable estimates of urban CO2 emissions using satellite XCO2 images. To do this, we use a little more than a month of simulations of local XCO2 scenes over large cities. This simulations are generated with the global OLAM model and evaluated by Schuh et al. (2021). We use these simulations to generate synthetic satellite images for the selected cities, and estimate their emissions by applying one of the automated and computationally-light inversion methods implemented, tested and optimized by Danjou et al. (2024). By using realistic simulations to derive the synthetic image and using a method independent of the model used for the simulations to estimate the emissions, we take into account realistically the uncertainty in the meteorology, atmospheric transport and background. As we are working with synthetic data, the error in the emissions estimate is directly accessible by comparing the emissions estimated by the inversion method with the synthetic true emissions used in the OLAM simulations. The study of the emission estimation error for different cities and weather conditions aim to support the identification of criteria for discriminating between images, separating those whose processing yields statistically reliable estimates from those whose processing is statistically unreliable."*

Sec. 2: Mention that ECMWF ERA-5 product is used for meteorological data. It is only mentioned in line 446. *ERA-5 is only used for the analysis in the subsection 6.2, not before (and therefore not in section 2). l 445 "For this analysis" changed to "For the analysis conducted in this subsection".*

130: What is the methodological rationale behind assuming constant emission rates? Do you expect that incorporating dynamic emission rates would significantly alter the study's results or conclusions? *We did it for the sake of simplicity. We don't think it will altar the results and conclusions, as mentionned in the discussion (l.425-437)*

217-221: Does the observation that all methods yield similar results suggest that the assumptions in the model used to generate the synthetic data might be overly simplistic? *No. It was already the case in Danjou et al. 2024. It is explained by the fact that all method share most of their steps : same way of estimating the plume boundaries, the background concentration.*

The final step (using a gaussian plume or a mass balance method to estimate the emission) has a minor impact on the overall error (as shown in Danjou et al. 2024).

Sec. 3: Does this paragraph solely describe the work done in Danjou et al. 2024? Certain statements, like in line 160 "The method used here...", are ambiguous in the current context. Does "here" refer to this study or to Danjou et al. 2024? It may be beneficial to condense the paragraph preceding Eq. (1) for clarity. *Addition at the beginning of section 3 of "The complete description of the inversion method and the details and justifications for its specific configuration and implementation can be found in Danjou et al. (2024). We make the assumption that the configurations chosen in the framework of their study remain optimal for other cities. This assumption seems justified, as the chosen methods for each steps differ from the discarded methods on objective criteria. This section only gives an overview of the different steps and the adaptations (compared to the reference configuration from Danjou et al. (2024)) that were made in the context of this study." We didn't shorten the part before equation (1) as we think that the informations given here are necessary to have an idea of the inversion method used and understand the changes made compared to Danjou et al. 2024.*

217-221: Consider clarifying this point earlier in the document, perhaps where the Gaussian plume inversion (Eq. (1)) is introduced, for better coherence and understanding. *Indeed. It comes much too late. We rewrote the introduction of the article and of the subsection to make this point clear from the start.*