Review of Danjou et al., AMTD:

In this manuscript, the authors discuss a methodology to estimate CO_2 emissions by cities in an automatic way to validate self-reported emissions by cities. For this, they use CO_2 concentrations simulated by a global model with an adaptive horizontal resolution enabling them to increase the resolution to about 23 km^2 locally around 31 cities with improved representation of the meteorology and potentially better comparability to satellite measurements around these targets. Their methodology consists of calculating hourly XCO_2 during local daytime from their simulation in a 150x150 km^2 region around the city and then derive criteria to select images which they conclude will be appropriate for satellites to observe the emission plume using instruments like OCO-3 or CO2M. While it is an important scientific topic the authors try to tackle, the descriptions in the manuscript are confusing and have many imprecisions making the manuscript hard to follow, which is why I suggest major revisions before the manuscript should be even thought of being published. In addition, it seems that the manuscript has been submitted in a preliminary stage because almost no abbreviations have been defined, some values are "xx" (e.g., L496) and there are many typos in the figures. My comments are separated in general and specific comments, followed by technical corrections.

General comments:

1. Citation Danjou et al. (20xx): While it is appreciated that it is made clear by the year number 20xx that it is a preprint under review, the whole manuscript is based on this non-published study (cited 40 times). In order to make the results of this current manuscript understandable for the general public, it is important to add a comprehensive overview with **all** the results from your previous study needed to understand the results in this manuscript. In addition, please add the journal the manuscript has been submitted to so that at least it may be found in the future. You could also consider to publish the preprint in a citable space and cite this here, which would make this manuscript here much more transparent at this stage.
2. It is never defined which is your reference determining the "error" which is mentioned at many points in the manuscript (e.g., L7, L91, L362 and many others). Unless the reference data of the city emission is not clear, the whole error analysis does not make sense, so please describe clearly what you use as a reference at some point of the manuscript. See also specific comment for L362. In addition, this also means that the motivation of this study is not clear: Are you investigating how well suited different methods are to determine emissions by cities (model as "true" emissions and method as uncertainty) or are you interested in which cities and meteorological conditions are best for analyzing city emissions with satellites, which is suggested by the title but where it is unclear what is your reference?
3. A major issue with satellite observations are clouds which will decrease the coverage. They are not mentioned in the methodology description (Sect. 4) in any way, but will probably be the main limitation to the emission estimates. I'm sure you can derive cloud information from your simulations which in my opinion has to be the first criterion to select whether emissions can be estimated using satellite measurements. This cloud screening would be something like a pre-selection of the images. Otherwise, the selection of the images is not "optimal" as suggested by the title.
4. It seems as if it is not accounted for detection limits of the satellites at any point in the manuscript and should be considered or at least mentioned somewhere. Can a satellite like OCO-3 or CO2M even distinguish emissions in the order of 2.1kt/h from the background?

5. All the results are based on simulations by the OLAM model, but the description of the model and of the performed simulations are not comprehensive, many things are missing making it hard to understand, see specific comments below for lines 95 to 138 and for Fig. 1.
6. You are mentioning at various points in the manuscript that you are interpolating and extrapolating without including information which methods are used and why you are using them. Do these interpolations to a 1x1 km grid influence the simulated emission fluxes? Are they mass-conserving? Why and where are you extrapolating the XCO_2 values (L173)? Common practice is to use the average of the background region. What is the reason that you use another procedure?
7. I am missing a description how you suggest to use your method for real measurements. How do you suggest to derive the variables needed for the analysis with your method with respect to real satellite measurements?
8. I would prefer reordering the methodology section because it is hard to follow in the current setup. First, I suggest putting the figure and table from Appendices A and B (Fig. A1 and Table B1) to their place in the main text where they are discussed as they have not any description in the appendices anyway. Second, it would be much better to move the descriptions from Sect. 5.2 to the decision criteria selection part in Sect. 4.2 and move the list of variables before this. This would improve the readability because otherwise, the reader is left with the methodology without the outcome in the current Sect. 4.2.
9. At some points you basically say that this is still work in progress and that your method is not applicable to all cities (L320, L324, L527) because it may depend on the surroundings of each individual city, which is okay but then the abstract should provide this information.

Specific comments:

- L1: define XCO2
- L4: "using synthetic satellite images": Please give more information how these images are generated in the abstract since this is part of this study.
- L6: Success rate of what?
- L7: Which cities did you choose? Please elaborate a bit.
- L8: What do you mean by "error"? What are you comparing with?
- L8-10: The sentence starting with "Our learning method [...]" is clear: You already say that you reduce the error in the previous sentence. So I suggest to remove this sentence or rephrase.
- L12: Please define IQR
- L16: define UNFCC
- L16: The citation has errors. Please consider putting the author in additional curved brackets.
- L25: define OCO
- L28: please define ppm
- L30-33: Please add a citation for OCO-3 SAMs to this sentence.
- L38: define CO2M
- L38: define GOSAT-GW
- L41: Please define WRF abbreviation.
- L46: "Attempt" could imply that these studies were not scientifically justified. Please rephrase.
- L50: Sentinel 5-P does not measure CO_2. Please remove it.
- L50: Please add a citation for GOSAT-2
- L51: define OLAM

- L53: I'm not sure what you mean by "variability of the local background signal". Within the surrounding of a small area like a city, the natural signal should be the same everywhere which is used for verification of the satellite instruments by small area approximation e.g. described by Taylor et al., AMT, (2023) and references therein. Please explain what you mean by that.
- L57: Please remove "extensively" because this is a subjective rating.
- L58: Please define in more detail what you mean by "pseudo-image".
- L65: Please define IQR
- L76: I don't see any discussion of the uncertainty in the meteorology in this manuscript. Please explain.
- L80-81: Please elaborate which configurations in the framework of Paris you are talking about. It seems to me that you do not rely on the same values. Please clarify and refine your definition of configurations here.
- L87: Do you mean "evaluate" or "judge"?
- L88-89: If you want to mention the subsections here you should mention all subsections (Section 2.1 is missing). But I would prefer introducing the subsections at the very start of Section 2.
- L90: The configurations are not "recalled" in Sect. 3. There's just always a reference to the non-published paper. So, please provide a comprehensive summary of your study somewhere in this paper, summarizing all results needed to understand it.
- L91: Which "discrimination" do you mean? Please clarify.
- L99-105: It is not clear to me from this description how the model exactly is setup: My understanding from this is that the grid box sizes in this model are flexible and can be adapted for a region of interest, while being fully coupled to their neighbours. On the other hand, it seems as if the parameterizations are also adapted for each grid box and I'm wondering how the authors can ensure mass conservation and general consistency in this model, e.g. at which grid size do you decide to switch from hydrostatic to non-hydrostatic mode? And is the time step the same everywhere in the model, and if yes, how do you deal with the fact of wind speeds leading to motion across more than one grid box, especially in the higher resolution? A bit more details would be beneficial here.
- L96-105: In addition, you describe below that you simulate more than one month with the model. How do you achieve realistic meteorology in your model? Is it nudged towards an external dataset? Is it free-running?
- around L106: What is the top altitude of your model? Does it account for troposphere, stratosphere or only the boundary layer? As satellites measure the total column XCO2 this is an important information to be added. In addition, how many model levels does your model have and how well-represented is the lower atmosphere in the vertical, i.e. how many model levels are in the boundary layer and what is the vertical grid spacing there?
- L106-108: This is not correct. There are models that use altitude and surface following coordinates successfully. So please delete this sentence or at least rephrase.
- L108-109: I would prefer either to delete these general sentences about the model or move them above to the introduction or the general description of OLAM above in this subsection. Otherwise, it is confusing with respect to the actual setup of the model in your study.
- L109: What do you mean by "It [...] allows to reduce the representation of urban plumes"? Do you mean that you can simulate the plumes more accurately using your approach?
- L110: Comparing the mentioned mesh size of 3 km with Figure 1 and the statements at the beginning of Sect. 2.3, it seems as if there is one value per hexagon, meaning that the grid points represent the hexagons with a side length of 3 km (according to Sect. 2.3) which does

not mean that the mesh size is in this order. A more readable quantity would maybe an effective side length of a rectangle with the same area as the hexagon.

- Sentence in L107-108: Does that mean that you use higher resolution at all coastlines and mountains on the globe in your configuration? Please clarify.
- L115: define ODIAC abbreviation
- L120: Which method is used to interpolate the results to the regular grid and which target resolution is used? Does the interpolation have an influence on the results?
- Figure 1: Where do the large XCO_2 values in the very south of the left panel come from?
- Figure 1: It would be helpful if you could include an example where the pseudo-images are located in this figure. In addition, "twice the size" is wrong because it relates to the square edges in both dimensions, therefore it should be "four times".
- L122: It seems from Fig. 1 that the mentioned regular grid is highly oversampling the native grid, but this should be noted somewhere here. In addition, it is not clear from the description whether XCO_2 is calculated on the native grid and then interpolated or vice versa. Figure 1 suggests calculation on the native grid and then interpolation, but the text describes it the other way. Please clarify.
- L125: The resolution should be moved upwards, including information how the data are interpolated. See comment above at L120.
- L128: As mentioned above, if there's one value per hexagon, the area of the hexagon is most relevant here, which is 23.4 km^2 for the model, 3.1 km^2 for OCO-3 SAMs and 6 km^2 for CO2M. Therefore, the hexagons at the highest resolution are still 4 to 7 times larger than the satellite footprints.
- L129-130: "[...] on which the analysis will be conducted" What is the mentioned spatial grid? Is it the actual satellite footprints?
- L129: It is not clear which "simulated patterns" you are talking about here. Please clarify.
- L130: Please provide the date notation in the standard format defined by Copernicus. In addition, these are 41 days, but obviously only 40 days are used to get the 9920 images. Please clarify.
- L132: Please clarify where the number of 9920 images comes from, because 8 hours per day x 41 days x 31 cities does not arrive at this number.
- L135: "that expected for CO2M" - Please mention the swath width of CO2M somewhere in the paper (maybe best in the introduction).
- Section 2.3: I understand from your description that you interpolate your simulation results to a 1x1 km grid and use this as a proxy for what the satellite sees. It would be much better if you used real orbit data from both satellites and account for the actual footprints of them (which you do anyway in your preprint Danjou et al., 20xx).
- L144-157: This procedure seems very complicated to me and also for the application to real satellite measurements. It would be great to have an illustration of the distribution of emission targets in a city and what is the benefit with respect to a simple circle of a specific size around the city center. I am sure there are reasons for you to do this procedure but it is not clear to me what these reasons are from the description in the text.
- L164: If the metropolitan areas from GRUMP are larger than a SAM what is your argument to decrease their size in your study? If the only reason is that you decrease it so you can study it, this would be a recursive argumentation and not scientific.
- L165: I am sure you can redefine these metropolitan areas in the model to your definition (why should it not be possible?). By that, this analysis would become feasible and possible.
- Sect. 3: Instead of always mentioning the non-published manuscript and comparing with it, it would be much better to describe your methodology here completely, so that the reader can

follow your steps (or as mentioned in the general comments provide a comprehensive summary of the other study at some point and refer to that).

- L170: Also here, an illustration would be helpful. For instance, does this plume boundary account for the fact that the emission zone is an extended zone or does it just start in the center of the image?
- L171: "over the entire image": For point sources such as power plants, only the wind at the location of the emitting target is usually used to determine the direction of the plume (e.g. Nassar et al., 2022). Although this is not a point source, I think the average should be taken for the emission zone, only.
- L175: I think you mean an estimate of the enhancement in the plume.
- L176: Is the 5 degree polynomial used because of changes in the wind direction? Can you estimate the emission when the wind is so variable? There will be also mixing into the plume when the wind varies in its direction which is why usually only times are used where the wind can be assumed to be uniform.
- L182: "averaged wind" By that you already assume that the wind is uniform all over the analysis area and the PBL. Therefore, I do not understand the 5-degree polynomial mentioned earlier.
- L186: Is W the horizontal wind?
- L187: The value in brackets of sigma should be x.
- L187: "the mass of CO2 in the atm. column per unit area" I assume you mean the enhancement in the plume?
- L190: It is very confusing to talk about "modelled" here, because it could also mean some XCO_2 modelled by your OLAM simulations. Please rephrase and mention that it's the Gaussian plume model you're using here.
- L191: Similarly to the previous comment: "Observed" is very confusing because you're using simulated values everywhere. Please rephrase.
- L192: Delta XCO_2 in the equation should be dependent on x and y
- L192: As this equation seems to be a numerical value equation units should be given to the quantities. What is Delta XCO2 here exactly? And I assume the M's are molar masses?
- L193: I would call the surface pressure P_{s,dry air}(x,y) because it depends on the x and y direction.
- L194: The description would be much clearer if you would mention that r, a and F are free parameters in Eq. 1 which you want to fit here to get the best estimate in terms of the Gaussian plume model.
- L197: Where does the "average" radius of the city come from?
- L201: Why is the normalisation needed? And why are you using exactly this normalisation? Please add this information here.
- L202: "for clarity": Why did you choose these limits? Please explain.
- L203: The angle should be "Theta" and not "Theta_init" and the "Element" symbol is missing.
- L204: "defined as optimal by Danjou et al. (20xx)". Please provide a summary of the non-published manuscript somewhere in the paper with all the results needed to understand this study.
- around L210: The description would be much clearer if you would mention that the actual shape of Paris is nearly circular whereas this is not generally the case for all cities, which is why you had to adjust the radius of interest.
- L211: What does "4.3" mean? Please clarify.

- L215: I do not understand what you mean by "directly above the city". From my understanding, you have to estimate the emission downwind of the emission target. Please clarify.
- L218: As the IME abbreviation is not in the main text, it would be better if you defined it in the Appendix C where it is actually used.
- L221: appendix --> Appendix C (or actually Appendix A if the Appendices A and B are removed as suggested above)
- L222: I think you would like to do an error reduction of your analysis, so please rephrase the title accordingly
- L222: Please explain first what you mean by the "sensitvities" and what is the general purpose of this section before going into the details.
- L229: You have not mentioned how the error on the emission estimate is calculated before. Please add this information to the previous section about the methodology.
- L231: As mentioned above, it is not clear from your description why you are doing this sensitivity analysis. A motivation for this is needed.
- L233-234: "a way to define ...": This is the motivation. Please move this to the front of the section.
- Sect. 4.1: I think you're saying here that you are binning your images according to percentile thresholds in 5% steps. But it is not clear from your description if you do this for each city separately.
- L263: "variance reduction": So your error mentioned earlier is the "variance"? Please explain.
- L266: "the depth is set to 2": You do not mention here, which criteria are used in the end, which is confusing. It would be better to combine this discussion with Sections 5.2.1 and 5.2.2 where the criteria are discussed, also because the second choice ("diagnostic" variables) depends on the first choice.
- Sect. 4.2.2: It seems as if your decision tree is a two-step procedure: First, you do it for the predictable variables, then you do it for the remaining images using the diagnostic variables. Please include this information somewhere.
- L285-286: The sum of the number of variables in brackets do not match the ten diagnostic variables.
- L286: I think it would be appropriate to put the table here as part of the main text. Otherwise, the reader has to go to the end of the paper to understand what you are referring to here.
- L290: There are many peer-reviewed publications highlighting the importance of wind in the calculation of emissions from emission targets, so please use another publication here.
- L294-298: So your suggestion is to apply exactly this method with the same thresholds to real satellite data for these cities?
- L299: You have already characterised the background XCO_2 in the sections before. Do you mean you want to characterise the variability of the background because this could lead to errors?
- L300: It is clear that the background is crucial for the analysis of the emission plume. There are many publications highlighting this, so please use another citation for this.
- L303-304: It is your choice to define a fixed size of the images. You could easily include a variability in the size of your images to analyse this effect, e.g. by adding a random parameter to the edge size of the images. Because you're saying this could be important this would be worth doing.
- L314: This is the very first mentioning of GP2. Please define what you mean by that.

- L314: It is unclear from the description where this number of 92% comes from. Please extend your explanation.
- L315-316: Does that mean that the uncertainty of your method is 78% which would be huge and mean that we basically cannot infer much information from that? Please clarify. An illustration of this would be helpful, too.
- L320: "the variability of the error distribution remains large across cities": Again, an illustration would be very helpful here to understand what is meant by error distribution.
- L324: Are you saying here that your method can only applied to single cities?
- L331: "significant": Did you check for statistical significance here? If not, please replace by "increased".
- Fig. 3 caption last line: left-hand --> right-hand
- L335: Since Appendix C has many subsections, please refer to the correct subsection here.
- L337: "impairing our ability to determine the optimal set of thresholds": If the variables are correlated they do not provide additional information, so you should choose independent variables.
- Section 5.2: These results should be moved to a much earlier place, because otherwise the manuscript is hard to follow.
- L352-357: I think what you want to say here is that you calculate the median threshold and remove all data that are beyond the threshold where the error increases. Please clarify.
- L354-355 and L356-357: Why is it important how many images are between the bounds? I do not see the importance of these sentences.
- Fig. 5 caption: It would be good to add the absolute emissions of the 31 cities you investigated at some point of the manuscript. Do they differ by many orders of magnitude?
- Fig. 5: From the description in the main text, I think the values on the y-axis are not given in percent but in ratios to the emission.
- L362: Since it is not explained at all in this study up to now, where the "error" comes from, this statement cannot be validated. Please explain in detail at some point how you define and calculate the error in your analysis.
- L424: Please repeat these here or somewhere else in the paper since your previous study has not been published yet.
- Fig. 6: Please add numbering to the panels and refer to them in the text. Otherwise, it is not clear which panel you are referring to.
- L448: The orbit of OCO-3 is not really predictable so that the overpass can happen at any local time during daytime. Please clarify.
- L449: Why are you using 11° here and not 12°?
- L451: It is clear that a lower resolution will result in smoothed and more homogeneous wind speeds/directions.
- L452: As mentioned above, the resolution of your model is not 3 km but rather something around 5 km.
- L453: “are located in Asia and America”: Which is surprising because these cities have supposedly the highest emissions. How do you explain this?
- L456-460: This should be mentioned much earlier. Please mention at your earliest convenience that you are interested in clear-sky conditions only here.
- L461-462: "as current instruments cannot make measurements mover both water and land in a limited time and space interval.": I think it would be better to say that the signal-to-noise-ratio is lower over water making measurements more challenging.
- L463: I thought it was 20 km. Why is it 30 km here??

- L465-468: I think this should be part of your methodology and not of your discussion of the results because this makes the connection between your model simulations and the real satellite measurements.
- L468: Why 11 degrees? Above, you derived a threshold of 12 degrees.
- L471: As can be seen for the cities at the west coast of middle Africa, there are regions with large cloud frequency in Africa. Please rephrase.
- L474: "no more observable cities": Please rephrase because this means that there are no cities that can be observed. You could e.g. write something like "the number of cities does not increase..."
- L476: "if observed daily": maybe better "if there were daily SAMs or overpasses"
- L477: Please move this sentence above to around L466 where you describe the general procedure.
- L480 and L484: "stand[s] out" is not a scientific notion. Please remove and just write the facts, e.g. "In Australia, only five cities..."
- L481: "For this contintent": Maybe better talk about the five cities instead of the contintent.
- L484-488: I think it would be much better here to talk about absolute number of cites instead of the relative number since the number varies a lot between contintents (e.g. 5 in Australia and 273 in Asia).
- L496: "xx vertical levels" Please include the number. In addition, the vertical resolution in the boundary layer is of relevance here.
- L497: Do you mean 137 levels?
- L500: "is less sensitive to sampling": I do not understand what you mean by that. For a given grid spacing, you get variability only for a certain resolution for all variables. Please clarify.
- L518: "precise" What do you mean by precise? How representative are these thresholds for real applications?
- L532: But one of the strengths of your method is that you can give some indication which cities can be used for satellite observations. So, which cities come out of your study whose emissions can be observed from space? This conclusion is missing here.
- Code and data availability: I don't think that this is conform to AMT policy. Please add your data to a repository to be publicly available.
- Appendices A and B: Actually, I do not see the reason to put these figures into the appendix of the paper. They can be part of the main text. Their discussion is done in the main text only anyway at the moment.
- Fig. A1: Please add information to the caption how the emissions shown in the panels are calculated.
- Fig. A1 caption: "OLAM boundaries": Please clarify: Is this your estimated area of the city or is this the GRUMP product mentioned in Sect. 2.4? If it's the GRUMP product, I would prefer that you show the boundary of the city you use in your analysis.
- Table B1: As mentioned earlier, I'm missing the cloudiness as a parameter here which will be very important to select times suitable for satellite measurements.
- L547: "in over 98%" Don't these methods vary in this number?
- Fig. C1: The quantile for the emission budget is not in the range between 0 and 1.
- Fig. C1: The maximum true ratio in panel (e) is 5 whereas it is 3 for the retrieved ratio in panel (d). How can you explain these differences?
- Fig. C1: panel (f) please add units of the standard deviation
- Fig. C1 panel g: Where does the local minimum at the 0.5 quantile in both methods come from?
- Fig. C1 caption: Please add to the caption why there are only 2 lines in panel (g).

- Caption of Fig. C1: "error" with respect to what?
- L567-568: This sentence "Cities with emissions..." does not make sense to me. Please rephrase.
- L571: Please remove "very"
- L572: Please add more information at which quantiles they are similar.
- L573: Where does the "real" anthropogenic signal come from?
- L578: Please remove "very" or define its dependence e.g. by numbers.
- L583-586: I don't understand these arguments. From my understanding, if the variables are correlated, only one of them should be included in this analysis in the first place.
- L590: There is no section 5.3.4.2
- L590: There is no table 5.2.
- L592: Why is the number lower for CS and IME?
- L593: Remove "very" here.
- L598: Again, there is no table 5.2. Please update.
- L608 and L609: I do not understand what you mean by "standing out". Please rephrase and explain.
- L611: Why do you now use the same thresholds as for GP2? The thresholds will depend on the method you use. In the previous section, you showed that the thresholds are different for each method and to use the optimal one for each method would be the way to go.
- L611: The description in section 5.2.1 has never been referred to as GP2 method. Please add this to this section.
- Fig. C2: Please add the unit to the y-axis
- Fig. C2 caption: Do you mean emissions > 2.1 ktCO2/h?
- L614: What do you mean by "accuracy"? Is it the median or the spread or both?

Technical corrections:

- L28: Please use \citep instead of \citet
- L29, L37, L38: XCO2: subscript for number 2
- L30: \citep instead of \citet
- L59: remove second "the"
- L61: "is" --> are
- L83: help to identify
- L84: help to identify
- L88: Remove "The" at the beginning of the line.
- L88: move "used" before OLAM simulations
- L97: Use \citep for Ullrich citation
- L116: I think, it should be "power plants"
- L145: a --> an
- L154: 2063 km2 (squared is missing)
- L156: Remove comma between "Note that" and "the"
- L161: Use \citet for the citation
- L163: citation is wrong
- L201: It would be better to include commas between the vector elements.
- L214: annex --> Appendix
- Figs. 2 and 3: Please switch the order of the figures to match the mentioning in the text.
- Sect. 4.2.1: Please convert the verbs from future to present, e.g. "will separate" --> separates

- Fig. 2 caption: will classified --> will be classified
- L294: "are based" --> is based
- Fig. 4: black line --> black dotted line
- Fig. 4 caption: the the --> the
- L397: modelisation --> simulation
- Fig. 5: typo in y-axis caption: % of the true emissions
- Fig. 6: right panel x-axis: Better "meeting" instead of "combining".
- L476: "an order of magnitude of" --> approximately
- L484: "of cities of" --> of cities with
- L504: "5.2.2" --> "Sect. 5.2.2"
- L511: "XCO2" add subscript
- Fig. A1: Please add more space for the Ningbo panel.
- Fig. A1: Please reverse the color scale because it is confusing that red means lowest emissions.
- Caption of Fig. A1: Add "Sect." to 2.3
- Caption of Table B1: 1first --> first
- Caption Table B1: W2D --> W (which actually occurs in the table)
- L538: "optimized" --> "applied" or "investigated"
- L539: "across" --> for
- Caption of Fig. C1: remove "according"
- L568: Remove "emissions." at the end
- L598: close --> similar