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)

Abstract. There is a growing interest in estimating urban CO_2 emission from space-borne imagery of CO_2 column-average dry air mole fraction (XCO_2). Emission estimation methods have been widely tested and applied to actual or synthetic images. However, there is still a lack of objective criteria for selecting images that are worth processing. This study analysis the performances of an automated method for estimating urban emissions as a function of the targeted cities and of the atmospheric conditions. It uses synthetic data experiments with synthetic truth and 9920 synthetic satellite images of XCO_2 over 31 of the largest cities across the world generated with the global adaptive mesh Ocean Land Atmospheric Model (OLAM) zoomed at high resolution over these cities. We use a decision tree learning method applied to this ensemble of synthetic images to define criteria on these emission and atmospheric conditions for the selection of satellite images.

We show that our automated method for the emission estimation, based on a Gaussian plume model, manages to produce estimates for 92% of the images when applied to our database of 9920 images. Our learning method identifies two criteria, the wind direction's spatial variability and the targeted city's emission budget, that discriminate images whose processing yield accurate emission estimates from those whose processing yield large errors. Images corresponding to a low spatial variability of wind direction (less than 12°) and to high urban emissions (greater than 12.1ktCO_2/h) account for 47% of our total number of images, and their processing yield relative errors on the emission estimates with a median value of -7% and a InterQuartile Range (IQR) of 56%. Despite such an efficient filtering, the accuracy of the estimates corresponding to this group of images varies widely from city to city. Images corresponding to a high spatial variability in wind direction or to low urban emissions account for 53% of our images and their processing yield relative errors on the emission estimates with a median value of -31% and a IQR of 99%.

Copyright statement. TEXT

1 Introduction

Many of the most emitting countries report their CO_2 emissions to the United Nations Framework Convention on Climate Change (UNFCC) annually (UNFCCC, 2013). However, despite this monitoring of emissions and the commitments made by nations to reduce them, the increase in CO_2 emissions continues year after year (Friedlingstein et al., 2022). Many cities worldwide have committed to reducing their emissions, notably through joint initiatives such as the Covenant of Mayors (https: //www.globalcovenantofmayors.org/) or the C40 cities (https://www.c40.org/). These cities carry out self-reported inventories (SRI) based on economic data to verify the effective reduction of their emissions. Gurney et al. (2021) compared SRIs of American cities to the Vulcan inventory (Gurney et al., 2020). This comparison shows large differences between the two datasets and highlighted the inaccuracy of the emissions estimates in most of the SRIs. Quantifying city emissions using satellite observations of CO_2 concentrations above cities could provide helpful information to decrease the uncertainty in such inventories.

Observations of CO_2 column-average dry air mole fraction (XCO_2) at the scale of a few square kilometers from the two current Orbiting Carbon Observatory missions (OCO-2 and OCO-3) have paved the way for quantifying emissions from large (a few $ktCO_2/h$) industrial (Chevallier et al., 2022; Nassar et al., 2017; Zheng et al., 2019) and urban (Lei et al., 2021; Reuter et al., 2019; Wu et al., 2018; Ye et al., 2020) sources of CO_2. Indeed, the accuracy of the observations (less than one ppm -part per million-, (Worden et al., 2017; Taylor et al., 2020)) is of the same order of magnitude as the XCO_2 enhancements of the plumes from these sources, and their fine resolution ($\approx 2 \times 2km$, (Eldering et al., 2017, 2019)) allows them to capture detailed transects or images of the plumes. The Snapshot Area Map (SAM) mode of OCO-3 provides "snapshot" images of about $80km \times 80km$ over the cities and thus a 2D coverage of the XCO_2 concentrations (Kiel et al., 2021), contrary to OCO-2 and to the nominal mode of OCO-3, which only samples XCO_2 over a fine swath ($\approx 10km$). Studies have used these SAMs to evaluate transport model simulations (Kiel et al., 2021) or to calculate local ratios between mole fractions of co-emitted species (Lei et al., 2022; Wu et al., 2022). First estimates of city emissions based on these SAMs have been presented in conferences. However, there is still a lack of systematic process of SAMs over cities to estimate the corresponding urban emissions.

Studies such as Broquet et al. (2018); Danjou et al. (2024); Pillai et al. (2016); Kuhlmann et al. (2020) have used synthetic data to evaluate the possibility of quantifying CO_2 urban emissions from 2D XCO_2 images, such as OCO-3 SAMs or simulated XCO_2 images from the future Copernicus Anthropogenic Carbon Dioxide Monitoring (CO2M (Sierk et al., 2021)) and Global Observing SATellite for Greenhouse gases and Water cycle (GOSAT-GW, https://www.nies.go.jp/soc/doc/IWGGMS-18/O/ 2-6_Hiroshi_Tanimoto.pdf) missions. The quantification relies on inverse modelling methods, some of which compare simulations from complex transport model to satellite observations to estimate emissions. However, Feng et al. (2016) and Lian et al. (2018) show that the Weather Research and Forecasting (WRF) model (used by Lei et al. (2021) and Ye et al. (2020) with OCO-2 data) simulate CO_2 transport poorly when the wind speed is low. Other emission estimation methods, called hereafter computationally-light methods, are based on simpler transport models (Gaussian plume (Krings et al., 2011)) mass balances (Integrated Mass Enhancement method (Frankenberg et al., 2016; Varon et al., 2019)) or direct flux estimation (Cross-Sectional method (Kuhlmann et al., 2020; Krings et al., 2011; Varon et al., 2019, 2020)). Danjou et al. (2024) evaluated these methods

and, again, showed that the quantification of emissions in low wind conditions bear large errors. However, no established procedures exist to properly select the cities and the satellite images for which the estimates are most accurate.

Schuh et al. (2021) use high-resolution simulations from a single global adaptive-mesh model, the Ocean Land Atmospheric Model (OLAM (Walko and Avissar, 2008a, b)), to rank the largest cities of the world as a function of the ratio between the average amplitude of the XCO_2 anthropogenic signals over the city and the variability of the background signal in the vicinity of the city. This classification provides insights, a priori, on the cities for which the emission estimates based on satellite images of their XCO_2 plume will likely be the most accurate. This analysis is made possible by OLAM's ability to represent both the plumes of cities around the world at high spatial resolution and the influence of large-scale variations in CO_2 concentrations on local variations in the background of these plumes. Danjou et al. (2024) investigate a set of computationally-light methods for estimating CO_2 emissions from a city using satellite images capturing most of the atmospheric CO_2 plume from this city. Their study compares existing computationally-light methods and their various parameterisation options at each step of the atmospheric plume detection and inversion process, using simulated satellite images (i.e. synthetic images) of XCO_2 concentration generated with a meteorological-atmospheric transport model over Paris. It identifies the most suitable methods and configurations, among those tested, for the estimation of Paris CO_2 emissions. In parallel, it quantifies the impact of the various sources of uncertainties associated with each method, at each step of the procedures. The error in the emission estimates is most sensitive to the meteorological conditions, and more specifically to (i) the spatial variability in the wind direction and (ii) to the homogeneity of the background concentration field. However, their study considers only one city, Paris, corresponding to a specific range of emissions, to a specific spatial extent and distribution of the urban emissions, to a single type of local topography, to a type of background concentration field, and to mid-latitude meteorological conditions. Therefore, there is a need to generalize these results, and to re-assess the distribution of the error on the emission estimate (bias, IQR) and the sensitivities of this error to the spatial variability in the wind direction and in the background concentration field by applying a similar analysis to multiple cities.

Wang et al. (2018) evaluate the ability to estimate emissions from a large ensemble of urban areas (≈ 5000, whose contours are defined on objective criteria) and power plants covering most of the global CO_2 emissions, based on synthetic XCO_2 images similar to those of the future CO2M mission, whose expected swath width is of $250 km$ and expected resolution of $2 km \times 2 km$ (Sierk et al., 2021). However, their quantification of the emission uncertainties does not account for the errors in atmospheric transport. Their study only addresses the sampling (swath, cloud cover loss, spatial resolution) and accuracy limitations of the XCO_2 imagery. However, the uncertainty in the shape and position of the plume (and thus the meteorology and the characteristics of the cities) can also influence the results and thus the ability to estimate the emissions of a city.

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 CO_2 emissions using satellite XCO_2 images. To do this, we use a little more than a month of simulations of local XCO_2 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.

Such an analysis can help to identify optimal targets for satellite targeting modes, for example, for OCO-3 SAMs, or, when processing large datasets from future imagers such as CO2M, help identify the portions of the data yielding images worth processing for plume detection and inversion. In addition, our analysis can help to robustly assess the errors associated with urban emission estimates as a function of city type and atmospheric or observational conditions. At least, this analysis is expected to support the development of tools to evaluate the reliability of the inversions.

Section 2 describes the derivation of the synthetic images and the definition of the cities boundaries. Section 3 describes the inversion method used to make the emission estimations for the main set of analysis in this study. The results with the three other automated methods described in Danjou et al. (2024) lead to similar conclusions and their analysis is thus summarized in Appendix B. Section 4 describes the learning method based on decision trees used to identify the best discrimination criteria. Section 5 analyses the sensitivities of the emission estimation error to the city type and atmospheric or observational condition and presents the results of the decision tree learning method. Section 6 discusses the limitations of the analysis conducted in this study and analyses the distribution of the discrimination criteria for cities in the world with more than one million inhabitants.

2 Simulations of XCO_2 images over multiple cities

Section 2.1 describes the OLAM model and Section 2.2 the configuration of the OLAM simulations used in this study. The derivation of the synthetic images from those simulations is described in Section 2.3. Section 2.4 details how we define the emissions zones we target for the inversions.

2.1 OLAM

The Ocean Land Atmospheric Model (OLAM) is a coupled ocean-atmosphere general circulation model (Walko and Avissar, 2008a, b) with a dynamical core that has been used in the Dynamical Core Model Intercomparison Project (DCMIP, (Ullrich et al., 2017)). The main feature of the OLAM model is its hexagonal grid whose size is adaptive (see illustration on Fig. 1), which makes it possible to bring high resolution to the zones of interest (non-hydrostatic mesoscale) while maintaining a coarse mesh over the rest of the globe (hydrostatic model). The adaptive horizontal grid allows, for example, areas with complex local dynamics, such as mountainous or coastal areas, to be modelled at high spatial resolution. In our case, it allows us to realistically represent the urban plumes of a large number of cities and the underlying large-scale variations in CO_2, while maintaining a global domain and an affordable computation time. This would not be feasible if using a global model with a

120 regular grid. Over the selected cities, the size of the mesh cells is of approximetely 9km^2 and progressively enlarges until it reaches around $4 \times 10^4 \text{km}^2$ (e.g. for the largest cells over the oceans).

The transport modelled between the different regions uses physical and dynamical schemes that vary according to the resolution, in particular for submesh convection. Turbulent diffusion is parameterised using the Smagorinsky model, which depends directly on the resolution of the mesh. For the submesh convection, the cumulus clouds, the precipitation and the

125 mixing are represented with a hybrid approach combining aspects of both Grell and Dévényi (2002) and Grell and Freitas (2014).
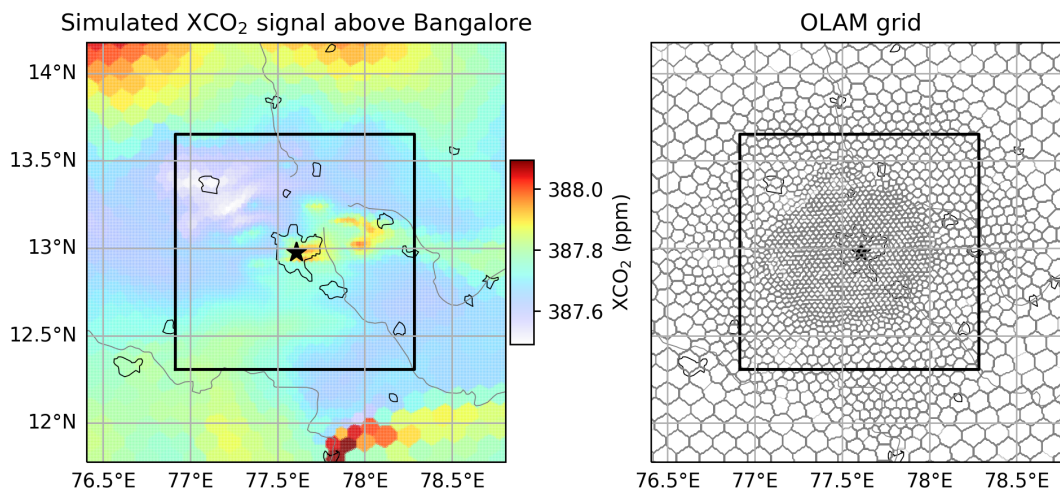
Figure 1. Illustration of the simulated XCO_2 signal above Bangalore on 8 August 2015 at 11a.m. (left panel) and the horizontal grid used for the simulation (right panel). The size of the illustration ($\approx 240\text{km} \times 240\text{km}$) is three times the size of the synthetic images used in the study ($\approx 150\text{km} \times 150\text{km}$). The boundaries of the Bangalore synthetic images are represented by the black line.

The model has 49 vertical levels (from 0masl to 37kmasl), twelve of them being in the first kilometre, which supports reliable simulations in the lower layers of the atmosphere, where the plumes are located. The vertical levels are at constant altitude and can therefore cross the surface. The fact that the levels can cross the surface helps avoiding gradient errors on steep

130 slopes that can be present in a pressure coordinate (or hybrid) grid (Ullrich et al., 2017).

2.2 Simulation of CO_2 transport

The OLAM atmospheric transport model is used to simulate the meteorological and CO_2 fields needed to build our synthetic images. The fluxes from the CarbonTracker 2017 global inversion system (Peters et al., 2007) are used as model input for the biogenic CO_2 surface fluxes. Anthropogenic emissions from the Open-Data Inventory for Anthropogenic Carbon dioxide

135 (ODIAC) spatialized inventory (Oda et al., 2018), are used to represent cities, industries, and powerplants. No temporal profile is applied to the ODIAC data, which means that the simulated anthropogenic emissions are constant over the month. From these data, the model will simulate on its hexahedral grid the wind, pressure, relative humidity and temperature fields (necessary for

the calculation of the PBL height, via the calculation of the potential temperature field and the Nielsen-Gammon et al. (2008) formula) and the CO_2 concentration fields in the atmosphere. The XCO_2 fields are calculated by the model on the hexahedral grid. The results of these simulations are then projected onto a regular grid at approximately $1km \times 1km$ resolution. This is done to simplify the analysis of the model outputs. The simulations are done for 31 cities. We retrieve the 2D fields of XCO_2 (i.e. vertical integration of the CO_2 profiles weighted by pressure levels) as well as the 3D fields of pressure, wind, relative humidity and temperature on the regular grid for all cities and their surroundings.

2.3 Genreation of the synthetic images

The model output resolution of $1km \times 1km$ is comparable to that of the OCO-3 SAMs ($1.25km \times 2.5km$, (Eldering et al., 2019)) and that planned for CO2M. This resolution is finer than the finest resolution of the model's adaptive native hexagonal grid (hexagons of $\approx 9km^2$). Therefore, the variations of the model variables (XCO_2 field, wind field,..) have a spatial resolution which is coarser than the $1km \times 1km$ resolution of the model output grid, on which the analysis will be conducted.

The simulations cover a little more than one month (08 August 2015 - 09 September 2015 (included)), providing hourly XCO_2 fields. For each day of the simulated period, we retain the hourly fields of XCO_2 between 10:00 and 17:00 local time for our synthetic images. This simulated database corresponds to a total of 9920 images interpolated at $1km \times 1km$ resolution. The spatial extension of the synthetic images is restricted to a $150km$ square whose axes follow the meridians and parallels and whose centre is the barycentre of the targeted city (in terms of CO_2 emissions). This size is halfway between that of the OCO-3 images and that expected for CO2M. Finally, a random noise of $0.7ppm$ standard deviation is added to the simulated XCO_2 field to simulate the satellite data. This value corresponds to the target accuracy for a single XCO_2 measurement from the CO2M mission, similar to the current precision of XCO_2 measurements from OCO-2 (Worden et al., 2017). We do not take clouds and the corresponding loss of XCO_2 retrievals into account when generating our synthetic images.

2.4 Defining the boundaries of the cities

The first task for urban emission estimation is to define the targeted emission zone. As the aim of our quantification is ultimately to help monitor actual emission reductions, we focus on the urban area corresponding to the most significant emissions rather than the actual administrative boundaries of the city. Consequently, the definition of the targeted emission zone is made regarding the size of plumes that can be detected in a SAM and by identifying the most emitting pixels from the spatialized inventories (using a similar concept but a different and more straightforward approach compared to Wang et al. (2019)). Because the typical size of a SAM is about $80km \times 80km$, we set the size of the targeted emission zone at roughly the size of a $20km$ radius disc. Thus, the emission zone we target occupies around 20% of a typical SAM and 6% of our synthetic images.

To define the boundaries of an emission zone, we first set its centre at the barycentre of anthropogenic emissions within the synthetic image. We then restrict the analysis to a disc of $50km$ radius around this centre. The size is arbitrarily fixed at 2.5 times the $20km$-long targeted emission zone radius. Within this $50km$-radius disc, we select only a fraction ($1/2.5^2$) of the most emitting pixels of the XCO_2 synthetic images. This fraction is explained by our choice to work with target areas of about $\pi \times 20^2km^2$, i.e. $1/2.5^2$ of the surface of the $50km$ radius disc in which this selection. is performed. In order to form a spatially

coherent set, we extend the selected area to all pixels within 5km of one of the pixels retained by this first selection. This enlargement allows us to avoid complex cuttings of the city and to obtain groups of pixels where emissions are statistically high. The last two steps include (i) the selection of the sole cluster of pixels located above the city centre and (ii) the addition of pixels not categorised as belonging to the target area but completely surrounded by the target area. The final target area covers an area between 1333km^2 (Lahore) and 2063km^2 (New-York) which corresponds to 6-9% of the spatial coverage of our syntethic images and 20-33% of the spatial coverage of most OCO-3 SAM images. We will call this targeted emission zone "the city" hereafter.

3 Inversion method

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.

The inversion method is based on the comparison of the urban plume detected in the image to a straight Gaussian plume. This comparison requires many preliminary steps. First of all (i) the definition of the boundaries of the urban area whose emissions we want to estimate. The method used here to define these boundaries is described in Section 2.4. Second, (ii) the plume boundaries are defined by the pixels located above the city and those in the cone downwind of the city within an angle of $45°$. The wind used to define the orientation of the cone is the average wind direction in the PBL over the entire image (from the OLAM simulation). Once the boundaries of the plume are known, we (iii) estimate the background concentrations, i.e. the XCO_2 signal in the plume which is not generated by the city emissions. This background concentration is extrapolated from the XCO_2 values of pixels outside the plume using a Gaussian kernel. The difference between the XCO_2 concentration in the synthetic image and the estimated XCO_2 background leads to an estimate of the plume enhancement generated by the city emission. We then (iv) calculate the central axis of the plume using a degree 5 polynomial regression using the pixels in the plume, weighted by the estimated XCO_2 signal from the city. Using this central axis of the plume, we (v) delineate the area of the plume that will be used for the Gaussian plume optimization (*analysis area*). This area is located between one times the approximate radius of the city ($\approx 20\text{km}$) and one and a half times the approximate radius of the city ($\approx 30\text{km}$) along the central axis of the plume (the justification for these distances is given later in the paragraph). At this stage, we have extracted the estimated XCO_2 signal from the city and we have determined the pixels that we will use for the optimization. We (vi) estimate the effective wind W, i.e. the wind driving the XCO_2 plume from the city, using the averaged wind within the PBL and within the analysis area. Finally, we estimate the emissions (vii) by inverting the following formula as defined by Krings et al. (2011):

$$\Delta\Omega_{gp}(x,y) = \frac{F}{\sqrt{2\pi} * |\boldsymbol{W}| * \sigma_y(x)} e^{-\frac{y^2}{2*\sigma_y(x)^2}} \tag{1}$$

where the x and y axes follow the directions parallel and perpendicular to the effective wind, F is the whole-city emissions estimate, and $\Delta\Omega_{gp}$ is the CO_2 mass enhancement of the plume in the atmospheric column per unit area. The term $\sigma_y(x)$ accounts for the horizontal extension of the source. We take $\sigma_y(x) = a * (x + (\frac{r}{4a})^{1/0.894})^{0.894}$ as Krings et al. (2011), where a is the Pasquill stability parameter (Pasquill, 1961) and r the city radius.

To estimate the emission budget, we perform a minimization of the mean square differences between the mass per unit area simulated by the Gaussian model ($\Delta\Omega_{gp}$) and the "observed" mass per unit area. The "observed" mass per unit is calculated from the XCO_2 signal from the city derived in step (iii) using : $\Delta\Omega(x,y) = \frac{M_{CO_2}}{M_{\text{dry air}}} * \Delta XCO_{2obs}(x,y) * 10^{-6} * \frac{P_{\text{s, dry air}}(x,y)}{g}$, where g is the Earth's gravity (in m/s^2), $P_{\text{s, dry air}}$ is the dry air surface pressure (in Pa), $M_{\text{dry air}}$ and M_{CO_2} the molar mass of dry air ($28.97g.mol^{-1}$) and CO_2 ($44.01g.mol^{-1}$) and ΔXCO_{2obs} the "observed" plume enhancement (in ppm).

The emission budget F, the Pasquill parameter a, the city radius r and the orientation of the axis (i.e. the wind angle) are free parameters in equation 1 that are optimized during this minimization. The initial values are: for a, the value given by the Pasquill (1961) table corresponding to the meteorological conditions at the time of the image acquisition; for the orientation of the reference frame, the direction of the average wind in the PBL (noted θ^{init}); for the radius of the city, the average radius of the city (noted r^{init}) defined as the square root of the city surface divided by π. The choice of the initial value of the emission budget (noted F^{init}) is more critical. Indeed, setting an initial value close to the exact value (let alone the exact value) might artificially improve our results. Instead, we take a random number from a beta distribution (with α=1.35, β=2.5 and a scaling factor of 5) multiplied by the actual emission of the central urban area. We normalise the variables for the optimisation as follows: $X = \left(\frac{F-F^{init}}{F^{init}}, \quad \frac{a}{120}, \quad \frac{\theta-\theta^{init}}{\pi/4}, \quad \frac{r-r^{init}}{r^{init}/2} \right)^T$. We further impose bounds on these variables during optimization (the bounds are shown without the normalization for clarity) $F \in [-F^{init}; +\infty]$, $a \in [0; 240]$, $\theta \in [\theta^{init} - \pi/4; \theta^{init} + \pi/4]$ and $r \in [0.5 * r^{init}; 1.5 * r^{init}]$. This bounds are fixed to avoid unrealistic results (e.g. detected plume direction perpendicular to the wind, high CO_2 uptake from the city,...).

The methods used for steps (ii) to (vii) are those defined as optimal by Danjou et al. (2024). Step (i) have been redefined in Section 2.4, and step (v) has been slightly adapted. We choose to make the analysis area (step v) closer (and smaller) than them. The new analysis area is located between the edge of the city ($\approx 20km$) and 1.5 times the radius of the city ($\approx 30km$) along the plume centerline, while it was located between the edge of the city ($\approx 20km$) and 2 times the radius of the city ($\approx 40km$) along the plume centerline in Danjou et al. (2024). The conclusions on the sensitivity of the inversions to the analysis area indicate in Danjou et al. (2024) that the closer the analysis area is to the city, the better the estimate.

4   Analysis of the sensitivities of the emission estimation error to observation conditions : general principles

To identify the main criteria of classification of the images based on the performances of the emission estimation, we analyze the sensitivity of the emission estimation error to the different variables characterising the observation conditions and the inversion. We thus can see which variables are influencing the most the emission estimation error, and define criteria, based on those variables, determining whether a synthetic image is suitable for emission estimation or not.

We test here two types of variables: (i) predictable variables, used to determine the most favorable conditions for the inversion, which aggregate information about weather conditions and city characteristics; and (ii) diagnostic variables, used to evaluate the inversion results, which aggregate image diagnostics and inversion diagnostics. The sensitivity of the emission estimation error to the predictable variables in the first instance, and to the diagnostic variables in the second instance, are analysed separately and in the same way. The two types of variables are analysed separately as they can answer to two different questions. The predictable variables can be used before the inversion to determine if an image will give a reliable emission estimate and is thus worth acquiring and inverting. The diagnostic variables are accessible only after the acquirement of the image and the inversion, and can thus just give an indication on the reliability of the emission estimate. The analysis described below is therefore applied to each of the two groups.

As a starting point, we examine separately the relationship between each variable of the chosen group (predictable or diagnostic) and the error on the emission estimate. This preliminary analysis provides a first overview of the variables to which the error is sensitive. After this first step, we analyse all the relationships between the variables and the error to identify the one or two variables to which the error is most sensitive. This identification is performed using a decision tree, the depth of which determines the number of variables identified. The decision tree directly defines thresholds on these variables: following a strict interpretation of the algorithm, these thresholds can be used in a binary way to define whether a synthetic image is suitable for emission estimation or not. In a more general way, these thresholds can be used as an indicative criteria to evaluate the synthetic images and the corresponding urban emission estimation. These identified variables, together with their respective thresholds, can be used to indicate the level of error of an estimate obtained during an inversion.

4.1 Preliminary analysis

To quantify the sensitivity of the error on the emission estimate to a specific meteorological variable, or a variable diagnosed by image processing or by inversion, we order our synthetic images according to the values of the variable. For the analysis with predictable (respectively diagnostic) variables, we separate our set of 9,920 synthetic images (resp. 4,259, cf. Section 5.2.1) thus ordered into subsets of 496 (resp. 213) synthetic images (5% of the total number). For example, when considering the mean wind in the PBL (i.e. a predictable variable), the first subset will include 496 synthetic images corresponding to the 496 smallest values of the mean wind in the PBL. The second subset will be composed of the 496 images corresponding to the values of the mean wind in the PBL ranked between the 497th and the 992nd position. The last group of images will include 496 synthetic images corresponding to the 496 largest values of the mean wind in the PBL. We then plot the error distribution for these subsets as a function of their rank to see if a significant trend is observed.

The simulations we use are based on an inventory of anthropogenic emissions with no temporal variations. As a result, the variables related to the emissions and the shape or topographical environment of the city have no temporal variability and therefore take only 31 values. Our study of the sensitivity of the error to these variables is therefore based directly on the analysis of the error distribution as a function of the value taken by the variable of interest.

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.

4.2.2 Description of our method for determining the decision criteria

A simple approach is to use the total set of synthetic images (9920 synthetic images in the case of the analysis of predictable variables, ≈4259 synthetic images in the case of diagnostic variables) as the input set for the decision tree learning method. We thus obtain at most 4 subsets and select the one with the smallest Mean Absolute Error (MAE) on the emission estimate. This subset is considered as the subset of synthetic images best suited for emission estimation and the rest of the synthetic images as less suited ones. We then study the distribution of the error for this set as well as the pair of criteria that led to this partition. In doing so, we have no information on the stability of the criteria and thresholds with respect to the starting set. This is problematic, especially since the city features can only take 31 values for the 9920 images, which increases the risk of overfitting.

To overcome this problem and to get an idea of the stability of the criteria, we create 100 sets of synthetic images each composed of random samples of 10% of our total set of synthetic images. We apply the learning algorithm to each of the 100 sets. We look at the subsets corresponding to each leaf and select the one with the smallest MAE. The decision path that leads to this leaf gives us the pair of criteria that we retain. This gives us 100 pairs of criteria. We analyse the redundancy of the criteria across these 100 pairs and the stability of the threshold values of the pair with the highest occurrence. The different threshold values found for the pair with the highest occurrence are applied to determine, for each pair, a subset of synthetic images for which the emission estimates are accurate. The distribution of the criterions obtained with the 100 sets of images,
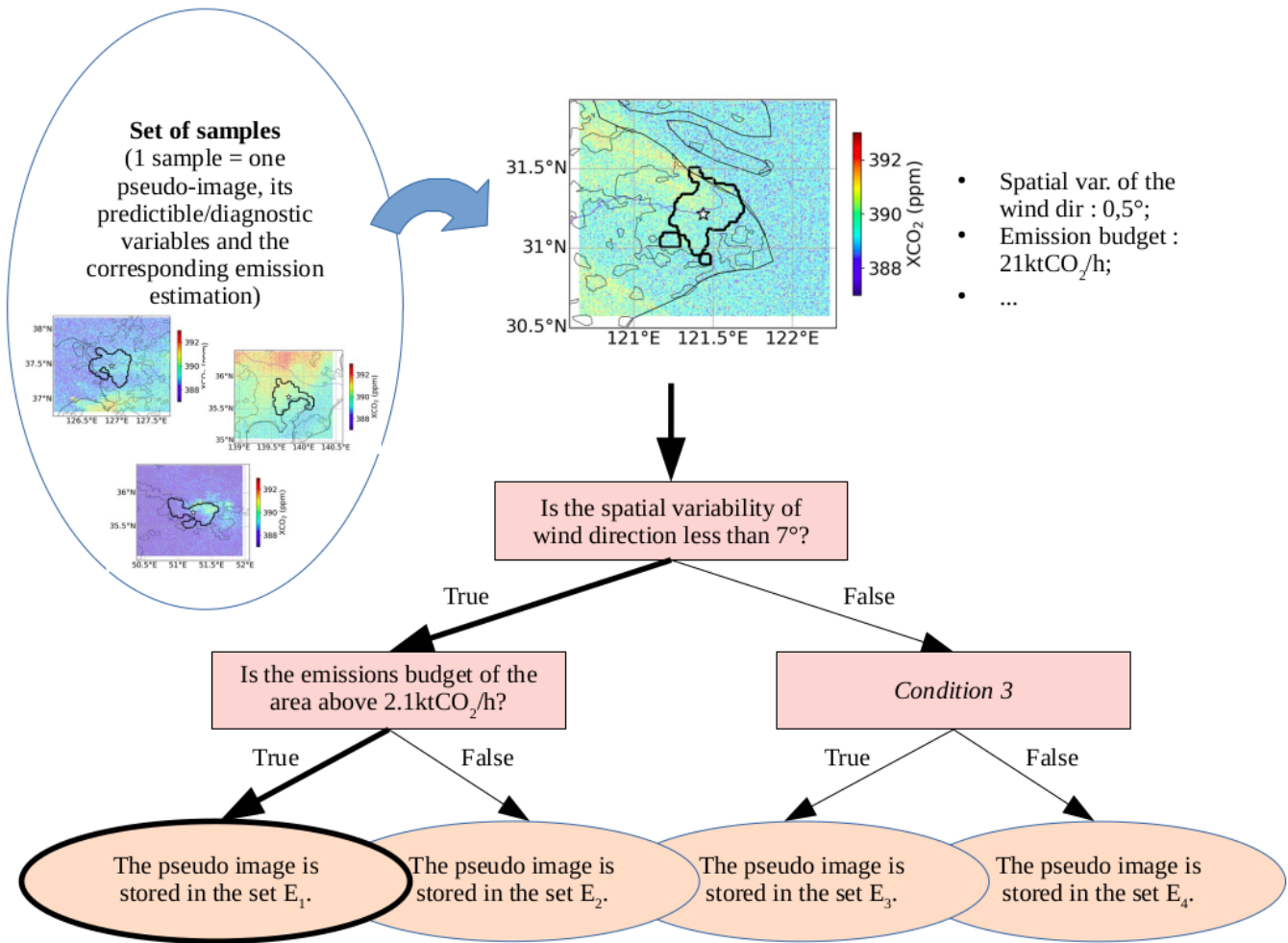
Figure 2. Illustration of how a decision tree works. The decisions (orange ellipses) and the conditions (pink rectangles) are called *tree nodes*. The first condition, through which the tree is entered, is called the *root node*; the terminal nodes (the decisions, represented by orange ellipses) are the *tree leaves*. Nodes that are not leaves are called *internal nodes*. The path followed by an individual (i.e. the conditions that have been tested for that individual) to arrive at a leaf is called a *decision path*. The length of a decision path is equal to the number of conditions tested on that path (i.e. number of internal nodes traversed). The *tree depth* is the length of the longest decision path. The input synthetic image will follow the decision path in bold and will be classified in the bold leaf.

as well as the error distributions of these subsets of synthetic images are studied to determine the reliability of the criterion threshold values.

4.3   List of variables

We tested 15 predictable variables (8 characterizing the weather and 7 characterizing the city) and 10 diagnostic variables (1 being an image diagnostic and 9 being inversion diagnostics). The detailed list is provided in table 1. Example and justification for our choice are provided in the following

To characterize the meteorological conditions, we have, for example, retained the wind speed in the PBL and the spatial variability of the wind direction (calculated as the circular variance of the 3D wind field in the PBL at the observation time), 2 variables whose influence on the accuracy of the emission estimation has been highlighted in previous studies Danjou et al. (2024); Feng et al. (2016). We have also looked at commonly-used quantities characterizing the wind (divergence, vorticity,.. of the wind in the PBL). To characterize the city properties, we looked at spatial variables (its size, the topographic variability in the surroundings, its symmetry) and variables representing the characteristics of the urban emissions (emission budget given by the inventory, emission density). In our synthetic data experiments, the analysis is based on values of the predictable variables that are extracted from the model, i.e. on the "true" values for all predictable variables. When using real satellite images (which is out of scope of this study), meteorological variables can be derived from weather products such as ERA-5 (Hersbach et al., 2018). City characteristics can, as in this study, be calculated from gridded inventories sur as ODIAC, and from database on urban land cover and population/socio-economic activities such as GRUMP (Center For International Earth Science Information Network-CIESIN-Columbia University and International Food Policy Research Institute-IFPRI and The World Bank and Centro Internacional De Agricultura Tropical-CIAT, 2011). The analysis will then rely on estimates bearing uncertainties, which could decrease the potential to identify suitable observation conditions. We note here that during our evaluations, the thresholds given in Section 5.2 will be compared to crude estimates when dealing with actual satellite data, a possible source of errors in the classification.

To characterise the complexity of the background XCO_2 field in the image, we use the spatial variability of the XCO_2 concentration. This variable has been highlighted by Danjou et al. (2024) as being correlated to the error on the emission estimation. Indeed, a high variability of the background leads to an estimation of the background concentration (step (iii) of the inversion method) less accurate and thus an error in the plume enhancement estimation, and thus in the emission estimation. This is the only variable diagnosed directly in the image among the list of diagnostic variables investigated here. With real data, the size of the image and its spatial coverage may have an influence on the accuracy of the emissions estimate. In this case, including this size in the list of diagnostic variables would make sense. However, this is not the case as we are working with synthetic data and all our images have the same size. Reproducing this variability in the coverage of real data is outside the scope of this study. The diagnostics of the inversion robustness include the size of the plume, the residual error after the optimisation with the Gaussian plume, the curvature of the central axis of the plume, the ratio between the estimated amplitude of the city signal and the variability of the signal outside the plume,... Unlike predictable variables, the calculated values for the diagnosed variables are directly inferred from the observations with real data. Therefore we will not have classification errors

analysis group	name	expression	type
predictable variables	mean wind in the PBL	$\frac{\sum_{\text{PBL}} \boldsymbol{W} * dP_{\text{air}}}{\sum_{\text{PBL}} dP_{\text{air}}}$	meteorological conditions
	spatial variability of the wind speed	$\sigma(\|\boldsymbol{W}\|)_{\text{PBL}}$	
	spatial variability of the wind direction	$\sigma(\theta_{\boldsymbol{W}})_{\text{PBL}}$	
	PBL height	$< z_{\text{PBL}}^{\text{a.g.l}} >_{\text{synth. image}}$	
	mean divergence of the 2D wind	$< \frac{\partial u^{\text{2D}}}{\partial x} + \frac{\partial v^{\text{2D}}}{\partial y} >_{\text{synth. image}}$	
	mean vorticity of the 2D wind	$< \frac{\partial v^{\text{2D}}}{\partial x} - \frac{\partial u^{\text{2D}}}{\partial y} >_{\text{synth. image}}$	
	mean stress of the 2D wind	$< \frac{\partial u^{\text{2D}}}{\partial x} - \frac{\partial v^{\text{2D}}}{\partial y} >_{\text{synth. image}}$	
	mean shear of the 2D wind	$< \frac{\partial v^{\text{2D}}}{\partial x} + \frac{\partial u^{\text{2D}}}{\partial y} >_{\text{synth. image}}$	
	small axis of the ellipse encompassing the city	ellipse defined with the library *opencv*	characteristics of the city (take only 31 distinct values)
	big axis of the ellipse encompassing the city		
	variance of the city emissions	$\sigma(\text{emiss})_{\text{city}}$	
	city area	$A_{\text{city}} = \sum_{\text{city}} 1$	
	emission density	$\sum_{\text{city}} \text{emiss}/A_{\text{city}}$	
	city topographic variability	$q_{90}\left(z_{\text{surf}}^{\text{a.s.l}}(x,y)\right)_{\sqrt{x^2+y^2}<50\text{km}}$ $- q_{10}\left(z_{\text{surf}}^{\text{a.s.l}}(x,y)\right)_{\sqrt{x^2+y^2}<50\text{km}}$	
	emission budget	$\sum_{\text{city}} \text{emiss}$	
diagnostic variables	spatial variability of the concentration in XCO$_2$	$\sigma(XCO_2^{\text{obs}})_{\text{synth. image}}$	image diagnostics
	estimated effective wind speed	$\|\boldsymbol{W}_{\text{eff}}\|$	inversion diagnostics
	residual mismatch after optimisation	$\frac{\sqrt{<(\delta\Omega^{\text{model}} - \delta\Omega^{\text{obs}})^2>_{\text{panache}}}}{<\delta\Omega^{\text{obs}}>_{\text{panache}}}$	
	optimized Pasquill parameter	/	
	optimized city radius	/	
	optimized plume direction	/	
	plume size	$\sum_{\text{panache}} 1$	
	ratio "average anthropogenic signal" vs "background signal variability"	$\frac{<XCO_2^{\text{obs}} - XCO_2^{\text{bckg,calc}}>_{\text{a.z.}}}{\sigma(XCO_2^{\text{obs}})_{\overline{\text{panache}}}}$	
	spatial variability of the XCO$_2$ outside of the plume	$\sigma(XCO_2^{\text{obs}})_{\overline{\text{panache}}}$	
	curvature of the plume centerline	average distance between the centres of the cross sections and the linear plume centreline	

Table 1. List of variables of interest divided by analysis group and type. The operators $< X >_E$, $\sigma(X)_E$, $q_{90}(X)_E$ and $q_{10}(X)_E$ denote respectively the mean, the standard deviation, the ninth decile and the first decile of the variable X on the set E. The sets *a.z.*, *plume*, $\overline{plume}$, *city* and *synth. image* designate respectively the set of pixels in the analysis zone, in the plume, out of the plume, above the city and all the pixels of the synthetic image. The reference frame used for the calculation of divergence, vorticity, stress, wind shear and topographic variability is the orthogonal frame with centre the centre of the city and with horizontal axis the mean wind direction in the PBL. $\left(u^{\text{2D}}, \quad v^{\text{2D}}\right)^T$ is the vertical average of the wind in the PBL, weighted by the dry air mass.

due to this. However, the values taken by the variables might have incorrect distributions in this theoretical study. For example, the distribution we use to simulate the measurement noise in our simulations is much simpler than actual measurement errors.

5 Results

5.1 Preliminary analysis

When we apply our inversion method to our 9920 synthetic images, we obtain an emission estimate in 92% of the cases (i.e. for 9119 synthetic images): in 8% of the cases, the optimisation does not converge. The bias (median) of the error distribution on the emissions estimate is -16% of the emissions, and the spread of this distribution (IQR) is 78% of the emissions. Reducing the bias and spread of this distribution is essential in order to obtain usable emissions estimates. Danjou et al. (2024), in their synthetic data study on the city of Paris, defined an image discrimination criterion based on the spatial variability of the wind direction, with a threshold of 7° (empirically defined). When we apply this filter, we reject 46% of the 9119 synthetic images and obtain a much less biased distribution of the error (5% of the emissions) and slightly less spread out (64% of the emissions). However, despite the application of this criterion, the variability of the error distribution remains large across cities. After filtering, the error distribution for the city of Lahore (largest MAE on the emissions estimate) shows a bias of -21% and a spread of 154% of the emissions, while that for Moscow (smallest MAE on the emissions estimate) shows a bias of -3% and a spread of 26%. This confirms that, although the criterion defined in Danjou et al. (2024) is relevant, our filtering step does not seem to be sufficient to select the synthetic images. The strong disparity of the error distributions between cities suggests that the error on the emission estimation is sensitive to the city characteristics (topography or city-specific atmospheric conditions) and/or to the city emissions (spatial distribution, magnitude,..).

Emissions are strongly underestimated when the wind is weak or when the spatial variability of the wind direction is strong (see Fig. 3). These 2 variables are also strongly correlated here (spearman correlation of -0.75). The results are more accurate (lower bias and IQR) when the meteorological conditions favour the ventilation of the emitted CO_2 in a narrow and straight plume, i.e. with a high wind speed and a low variability of the wind direction; but when the emitted CO_2 accumulates above and in the vicinity of the city in a diffuse plume with high values of XCO_2, or forms a plume with a complex structure, the results of the emission estimation show an important bias (see Fig. 3).

The error in the emission estimation also shows sensitivities to other variables characterizing the observation conditions: sensitivities to the emission budget, to the ratio between the average anthropogenic signal and the variability of the background signal, or to the difference between the optimised inversion angle and the average wind direction in the PBL are also visible (see Supp.Mat. B2).

The error in estimating emissions therefore shows sensitivities, sometimes complex, to several variables, some being related, again in complex ways. Because of those intricated sensitivities, the simple analysis conducted in this subsection is insufficient to determine the optimal set of variables and thresholds for defining the most optimal discrimination criteria for the synthetic images. This justifies the use of a more complex learning method. The supervised learning method described in Section 4.2 will enable us to determine the discrimination criteria more objectively, despite the covariances among the variables.
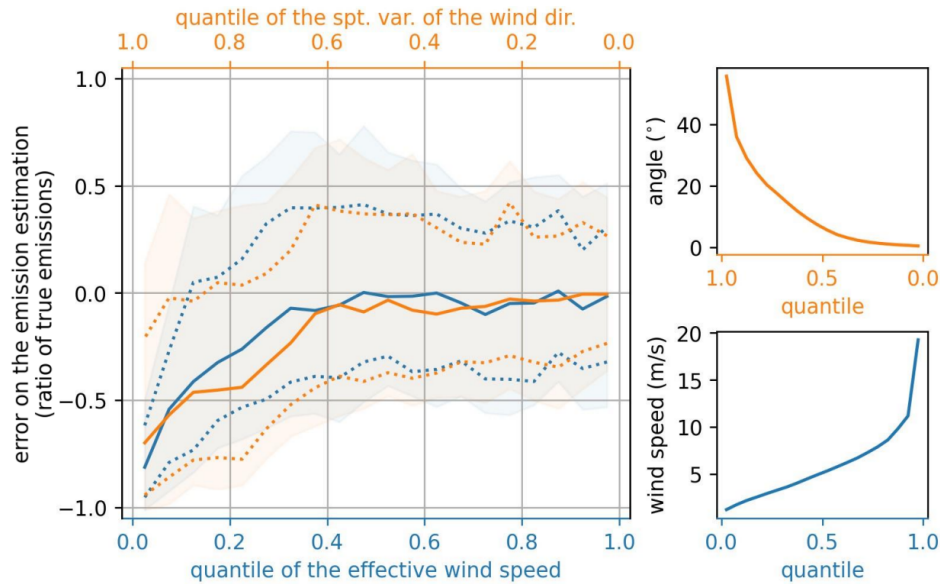
Figure 3. Sensitivity of the emission estimation error to two discriminant variables: the estimated effective wind speed (blue) and the spatial variability of the wind direction (orange). The left panel shows the evolution of the error distribution as a function of the quantile of the variable of interest: 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 right-hand panels show the values taken by the variables of interest for the different quantiles.

5.2 Application of the decision tree method..

5.2.1 ... for predictable variables

This first analysis, using the decision tree learning method described in Section 4.2.1, is based on the results of the inversion of the 9119 synthetic images produced by the GP2 method. We focus on the discrimination criteria given by our learning method

370   with 100 different samples, as described in Section 4.2.2. For 82 of the 100 samples, the pair of criteria given by our learning method is the spatial variability of the wind direction and the emission budget, i.e. favoring large emissions and low variability of the wind direction. For the remaining 18 samples, wind direction variability appears 9 times in the criterion pair and emission budget 7 times. The other variables appearing in the criteria pairs for the 18 samples are spatial variability of emissions in the city (5 occurrences), spatial variability of the emissions in the city (4 occurrences), mean PBL height (2 occurrences) and length

375   of minor axis of ellipse (2 occurrences). For 6 samples, the pair of criteria is in fact a singleton indicating that one variable is significantly more important than all the remaining variables. The spatial variability of the wind direction and the emissions within the city thus stand out very strongly.

We will now study in detail the threshold values taken by the spatial variability of the wind direction and the city's emission budget for these 82 pairs of criteria. The distribution of the threshold on the spatial variability of the wind direction is char-

380   acterised by a median of 12° and an IQR of 5°. 10% of the inversions are found between the bounds formed by the quartiles
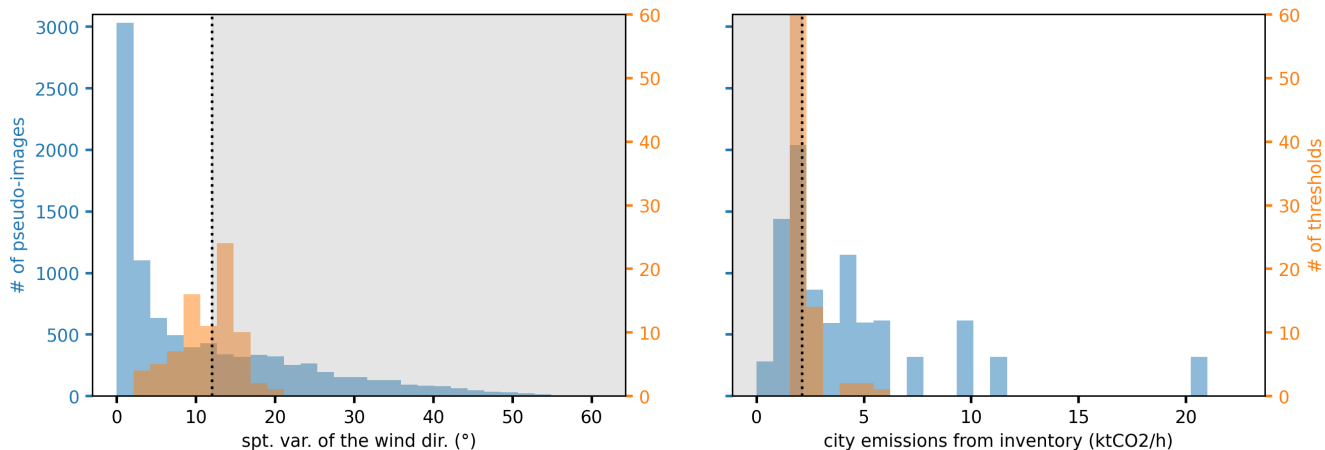
Figure 4. Distributions of spatial variability of wind direction (left) and city emissions (right). In orange the distribution is that of the criterion values and in blue that of the simulations. The black dotted line indicates the median threshold found and the grey area the discarded subset.

of this distribution (9° and 14°). The distribution of the threshold on the emission budget is characterised by a median of 2.1ktCO$_2$/h=5.1MtC/yr and an IQR of 0.7ktCO$_2$/h. 22% of the situations fall between the bounds formed by the quartiles of this distribution (2.6ktCO$_2$/h and 1.9ktCO$_2$/h). The distribution of the thresholds are therefore spread out. For a given pair of criteria among the 82 retained, the subset giving the lowest error is that formed by images whose spatial variability of wind

385 direction is below the threshold given by the decision tree and whose emission budget is above the threshold given by the decision tree. The 82 subsets are homogeneous in terms of the median of the error distribution (-7% [-6%,-8%]) and the IQR (55% [52%,58%]). This is less the case for the subsets size (45% [36%,52%] of the 9119 synthetic images). For comparison, other studies such as Wang et al. (2019) or Lespinas et al. (2020) have found lower thresholds on the emission budget (respectively 2MtC/year and 0.5MtC/year) leading to more precise estimates (uncertainties of less than 20%). But these studies, which both

390 follow the same formalism, include fewer sources of error in their framework (perfectly known background concentration, simplistic simulations of the urban plumes), which explains our higher threshold and uncertainties.

For the following analysis, we take the medians of the threshold distributions of our 82 retained pairs as the thresholds for these two criteria. The subset formed by the synthetic images respecting these two criteria is characterised by a median error of the estimated emissions of -7% of the city's emissions, an IQR of 56% and includes 47% of the 9119 synthetic images. The

395 subset formed by the synthetic images that do not respect these two criteria is characterised by a median error of the estimated emissions of -31% of the city's emissions, an IQR of 99% and includes 53% of the 9119 synthetic images. The criteria therefore allow us to isolate the synthetic images that are most suitable for inversion, as the synthetic images that do not pass the criteria give highly biased estimates.

The discrimination criterion based on the spatial variability of the wind direction reduces the bias and the IQR of the error

400 distribution, while the criterion based on the emission budget only reduces the IQR. Indeed, by applying only the discrimination criterion based on the spatial variability of the wind direction, we obtain for the subset passing the criterion a bias of the error

of -5% and an IQR of 68% (-31% and 99% respectively for the synthetic images not passing the criterion). Applying only the discrimination criterion based on emissions balance gives us, for the subset of synthetic images passing the criterion, a bias of -16% and an IQR of 66% (-17% and 110% respectively for the synthetic images not passing the criterion). Thus the criterion based on the spatial variability of wind direction is a selection criterion (the synthetic images that do not pass the criterion are considered unusable), and the criterion based on the emission budget is a discrimination criterion (the synthetic images that do not pass the criterion will give a less accurate emission estimate).

5.2.2   ... for diagnostic variables

In this section, the set of synthetic images used for the analysis is the set of synthetic images (47% of our previous set) passing the criteria on the spatial variability of the wind direction variability and on the emission budget defined in Section 5.2.1.

The pair with the highest occurrence (42 out of the 100 pairs) is the "ratio between the average anthropogenic signal and the variability of the background signal" and "spatial variability of the XCO_2 concentration outside the plume". For the other samples we obtain 20 different pairs. The spatial variability of the XCO_2 concentration outside the plume is also used in the calculation of the estimated signal to background ratio. The two variables have a correlation of 0.34. We thus choose to reduce the tree depth to 1 and remove the ratio between the average anthropogenic signal and the variability of the background signal from our list of variables of interest. The choice of which variable to remove between the 2 is made on the number of occurrences across the pairs (54 for the ratio between the average anthropogenic signal and the variability of the background signal, 77 for the XCO_2 signal variability).

In this new configuration, 72 samples out of 100 give the variability of the XCO_2 signal as a criterion. The distribution of threshold values found for this criterion has a median equal to 0.72ppm and an IQR of 0.02ppm. 19% of the synthetic images in the test set fall within the bounds formed by the quartiles of this distribution. By taking the median of this distribution as the discrimination criterion, we obtain two subsets which contain respectively 30% and 70% of the tested set and are characterised by biases on the estimation of emissions of -6% and -7%, and IQRs of 74% and 50%. This discrimination criterion reduces the IQR but not the bias. However, the accuracy of this criterion is questionable: 50% of the values taken by the signal variability outside the plume are between 0.70 (which corresponds to the measurement noise) and 0.73ppm. A slight variation (0.01ppm) in this separation criterion has a strong impact on the error distributions of the two subsets. Moreover, the modelisation of the instrument noise (which have an important impact on the signal variability outside of the plume) is over simplistic in our work. We therefore choose not to retain this criterion.

5.3   Study of the results by city

Of the 31 cities, five (Bogota, Lima, Los Angeles, Mexico City and Tehran) have more than 90% of the synthetic images that do not pass the selection criterion based on the spatial variability of wind direction. We therefore have less than 30 images passing the selection criterion for these cities and choose to set them aside. All these cities are located in basins or at the foot of high mountain ranges, which explains the high spatial variability of wind direction for the vast majority of observations. Of the remaining 26 cities, 7 have their emission budget below the threshold of the emission budget criterion (see Fig. 5) and
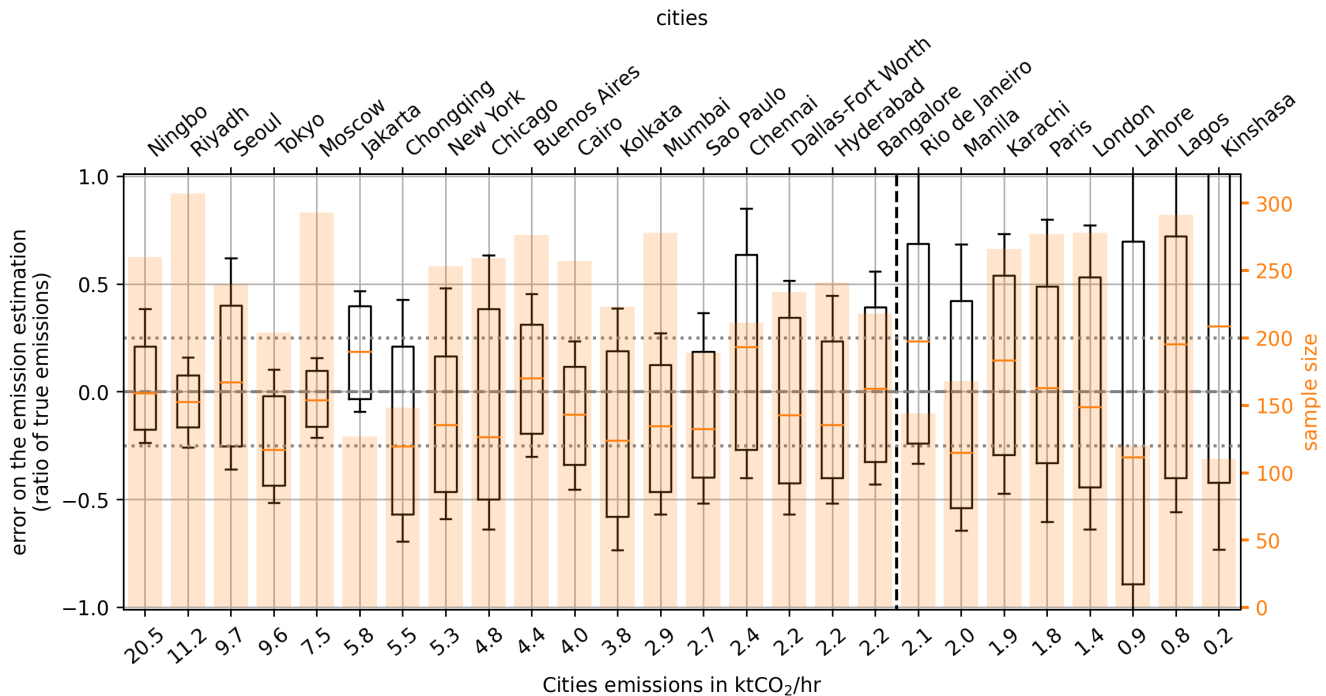
Figure 5. Distribution of the error on the emission estimates (boxplot) for the synthetic images passing the selection criterion on the spatial variability of the wind direction ($> 12°$). The orange bars show the number of synthetic images used. The dotted line separates the cities according to the discrimination criterion on the city's emissions budget with on the left the cities passing the criterion (emissions $<2.1\text{ktCO}_2/\text{h}$) and on the right the cities not passing it. The cities are ranked in descending order according to their emissions budget.

435 should therefore have low accuracy estimates. Paris is one of these cities in our simulations, with emissions of $1.8\text{ktCO}_2/\text{h}$ for the target area. The error distribution of the emission estimate for the city of Paris has a bias of 2% and an IQR of 83% for the synthetic images passing the selection criterion on the spatial variability of the wind direction (86% of the synthetic images). These results are close to those obtained in Danjou et al. (2024) with the synthetic images of Paris generated by WRF: the distribution of the error on the emission estimate had a bias of 4%, an IQR of 74% and 57% of the synthetic images passed

440 the criterion then defined. The IQR is larger in this study, and the number of images passing the criterion is higher. This can be explained by the fact that the criterion was stricter in Danjou et al. (2024) ($<7°$, $<12°$ in this study), and that the selected months are not the same (December-April, August in this study).

We can see (figure 5) that the accuracy of the estimate decreases with decreasing emissions budgets. However, this criterion alone is not sufficient to classify the cities. In particular, the bias varies considerably from one city to another, even when their

445 emissions are similar. Only 8 cities (Bangalore, Buenos Aires, London, Moscow, Ningbo, Paris, Riyadh and Seoul) have a distribution of error on their emissions estimates with a bias of less than 10%. Our selection allowed us to roughly filter out the worst situations for estimating emissions with our method, but not yet to fully understand the error dependencies. We want

to point out that these errors are significant, even with many images (≈ 320) per city and our filtering. Future studies should consider how best to use the emission estimates provided by satellite image analysis.

6 Discussion

6.1 Limitations of the study

Some potential sources of error not considered here (complexity of measurement error, loss of data due to - among other things - cloud cover and aerosols) have already been discussed in Danjou et al. (2024) and are therefore not repeated here.

A major difference between the simulations in this study and those in Danjou et al. (2024) is the lack of temporal variability of the emissions used. In reality, the plume is generated by the emissions that occurred up to a few hours before the satellite overpass, and inventories show significant daily cycles, in particular related to the traffic and industrial activity. When analyzing real data, our analysis zone may correspond to emissions that occurred, for example, three hours before the acquisition time, and comparing the emission estimate to the emissions at the acquisition time of the synthetic image introduces an additional error. Additional uncertainties due to the emissions temporal variability might affect real-case studies. In practice, Danjou et al. (2024) showed that the analysis zones correspond to emissions that are very recent (less than 2 hours) in most cases. Thus, carrying out this study with variable emissions should not significantly alter our results, assuming that the emissions of a given city remain similar within two hours in the middle of the day (no morning and evening traffic rush hours). However, the issue of temporal variation of emissions arises with real data when we compare our emission estimates with inventories. For cities without hourly emission budgets (or if the comparison is made with an inventory that does not vary on an hourly basis), we will have an additional source of error, this time coming from the estimated emission budget of the inventory.

We also note here that one of the criteria is based on the city's emission budget, which may be problematic when using real data. Indeed, including an a priori value from an inventory to rank the cities means that ranking errors might result in additional uncertainties if city's inventory estimates are incorrect.

6.2 Distribution of the discrimination criteria for the cities with more than 1M inhabitants

This Section focuses on the number of cities passing the discrimination criteria for a non-negligible part of the year. We are interested in cities with more than one million inhabitants in 2018, according to UNDESA (2018).

The spatial variability of wind direction is calculated as the pressure-weighted circular variance of wind in the PBL in a 150km square centred on the city centre given by UNDESA (2018). For the analysis conducted in this subsection, the meteorological data (3D wind field, pressure field, PBL height) come from the ECMWF ERA-5 product (Hersbach et al. (2018)) at $0.25°$ resolution for the year 2020. We calculate this variability for each day at 10 am, 1 pm and 4 pm local time. These different times are chosen to sample possible times of OCO-3 overpasses. We calculate for each city the proportion of these "observations" for which the spatial variability is above $12°$. We can see in Fig. 6 (a) that for the vast majority of cities (93%), this proportion is above 50%. The distribution of the spatial variability of the wind direction is different from the
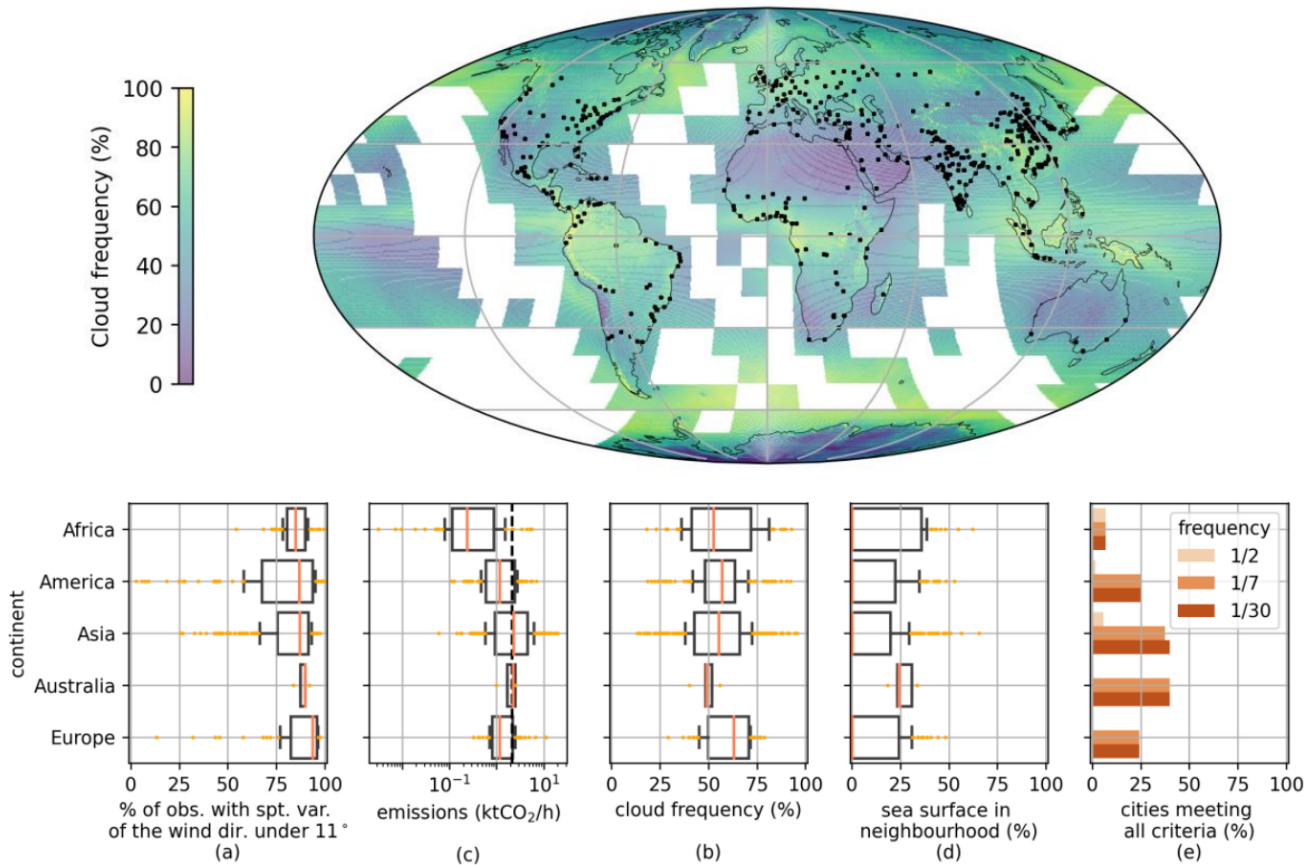
Figure 6. Average annual cloud frequency over the period 2000-2014 derived from satellites Terra and Acqua (Wilson and Jetz, 2016) and location of cities with more than one million inhabitants (top panel). Distributions (4 panels from left to bottom) for cities of different continents of the frequency of the synthetic observation with spatial variability of wind direction less than $12°$ (a), of the mean annual cloud frequency (b), of the emissions (c) and of the proportion of sea surface in the neighborhood of the city (d), defined as the area within $30km$ of the city. The panel (e) shows the percentage of cities per continent passing the emission budget criterion, having less than 25% of their neighborhood covered by the sea, and for which the frequency of observations allowed a priori by the cloud cover and the threshold on the spatial variability of the wind direction is one day out of two (light orange), one day per week (orange), one day per month (dark orange).

one we have with the OLAM model, where more cases were rejected. An explanation may be the much lower sampling of
480 ERA-5 (around $25km$ against around $3km$ for OLAM in the neighbourhood of the cities of interest), which smoothes the wind direction variations and thus leads to smaller values of the spatial variability of wind direction. Nevertheless, we can see that, based on this variable, the least suitable cities for emissions monitoring are located in Asia and America.

City emissions are evaluated with the ODIAC product for the year 2019 and using the definition of city boundaries defined in Section 2.4. Among cities with more than one million inhabitants, only 40% pass this criterion (Fig. 6 (b)).

Cloud cover is also a factor limiting the number of images that can be acquired, which is not considered in this study. Wilson and Jetz (2016) is a database giving the frequency at which clouds cover a point on the globe. This dataset integrates 15 years of twice-daily remote sensing-derived cloud observations at 1km resolution. We are interested in the annual average of this frequency to have an order of magnitude of the days not observable due to clouds. We can see in Fig. 6 (c) that the cloud frequency is, for most cities, between 40 and 80%, whatever the continent. The seasonal distributions of cloud cover and spatial variability of wind direction are not taken into account in this analysis.

Finally, the proportion of the water surface in the vicinity is also important for our measurements. The difference in reflectivity between terrestrial and aqueous surfaces results in very heterogeneous measurement quality. OCO-3 SAMs partially overlooking aqueous surfaces (e.g. coastal cities) include a large fraction of excluded pixels. For the analysis of this subsection, we define a city neighbourhood as the area within 30km of the city edge as defined with the method described in Section 2.4. For most cities (77%), this proportion is less than 25% (Fig. 6 (d)).

To give an idea of the current ability to quantify urban CO_2 emissions using satellite imagery, we look at the distribution of cities with emissions greater than 2.1ktCO_2/h and with less than 25% of the sea surface in their vicinity. We add an index of how often we can measure them by multiplying the proportion of cloud-free days by the proportion of days where the spatial variability of wind direction is greater than 12°. We can see in Fig. 6 (e) the proportion of cities per continent that pass the criteria and can be measured on average every other day, one day per week and one day per month (1 day/30). Very few African cities (4 out of 57) pass our criteria, mainly due to their low emissions. The proportion of cities passing all three criteria (emissions, sea in the vicinity, frequency of observation) does not change with the frequency threshold. Indeed, for those cities the emission budget is the discrimination criteria, and not the cloud cover nor the spatial variability of the wind direction. America and Europe show similar results: most cities are rejected by our emission criterion, and the high cloud cover (often more than 50%) does not allow for observations at least every other day. On the other hand, the number of observable cities does not increase when the threshold on the frequency of observation is raised from one day per week to one day per month. The observable cities in America and Europe (30 cities out of 119 and 14 cities out of 58) can provide approximately one observation per week if there were daily overpasses. Asian cities, due to higher emissions, show a higher proportion of cities passing the criteria. Very few cities (16 out of 273) are observable on average every other day. Again, the proportion of cities passing the criteria varies little between a threshold of one day per week and one day per month (101 and 109 cities out of 273, respectively). Australia stands out: only five cities have more than 1 million inhabitants. For this continent, the distribution of the variables we are interested in is fairly homogeneous, which places them at the limit of observability with the criteria on emissions and the proportion of sea surface in the vicinity (all the cities are coastal).

Asia and Australia stand out, with 37% (102 cities) and 40% (2 cities) of cities with more than one million inhabitants for which CO_2 emissions are theoretically relatively easier to quantify than cities on other continents. They are followed by America and Europe, with 25% of cities for both (i.e. 30 and 14 cities). Due to their lower emissions compared to other continents, African cities seem more difficult to monitor (only 7% pass our criteria, i.e. 4 cities). These conclusions remain valid for satellite imagers with characteristics close to those of the OCO-3 instrument (2km × 2km resolution, 0.7ppm accuracy) and should be revisited for future satellites with different viewing geometry or ground track.

6.3 Other potential criteria

Wind speed is often cited as having an impact on the magnitude of error while quantifying greenhouse gases emissions of local sources using satellite imagery (Varon et al., 2018, 2020; Nassar et al., 2022). As we have seen, using a criterion based on wind speed is relevant, as low wind speeds are often associated with high spatial variability in wind direction. These situations give rise to poorly ventilated plumes with complex structures whose corresponding emissions are difficult to calculate. This study's decision tree learning method indicates that the criterion on spatial variability of wind direction is more accurate than a criterion on wind speed with the set of images used here. However, this might be different when using real data. Indeed, the resolution of the weather product used here is very high around the cities of interest ($\approx 3\mathrm{km}$, horizontally, 49 vertical levels), higher than that of, for example, the ERA-5 product ($\approx 25\mathrm{km}$, 37 vertical levels). With wind data at a resolution comparable to that of ERA-5, the spatial variability of wind direction will be underestimated when the typical size of the horizontal variations is between 3 and $25\mathrm{km}$, and the accuracy of the criterion will be lower. A criterion based on wind speed might then be more relevant, as this variable is less sensitive to the resolution.

Another criterion often associated a priori with error in emissions estimation is the ratio between the average anthropogenic signal and the variability of the background signal (Schuh et al., 2021). This ratio quantifies the visibility of the plume and indicates how easy it is to quantify the emissions. We have seen that the error on emission estimation shows a high sensitivity to this variable (Section 5.1 and Annex B) and is apparent in our decision tree analysis for diagnostic variables (Section 5.2.2). However, this dependence of the error on the ratio "average anthropogenic signal" - "background variability" is slightly less important in our analysis than the dependence on the background variability. The relevance of the background variability as a criterion has already been discussed in Section 5.2.2. A priori, we might have expected the error's dependence on this ratio to be greater than its dependence on background variability. However, this dependence has already been partly filtered out by our analysis of the predictable variables, with the criterion on the emission budget.

7 Conclusions

This study analyses the performance of an automatic process for estimating urban emissions from XCO$_2$ satellite images. This process is independent of the targeted cities: it is applied identically to all of them. The methods used are low in computation time (on the order of a minute to process an image) and flexible, which enabless to process a database of around 10,000 images with a low failure rate (8% of the image). This study, therefore, contributes to the development of standard and automated methods for the operational monitoring of urban emissions with satellite observations.

Our analysis, using a decision tree learning method, of the variations of the error on the emission estimation as a function of the targeted cities and atmospheric conditions shows that the spatial variability of the wind direction and the city's emission budget are the two main criteria, among those tested, to select the most suitable images for city emission estimates based on XCO$_2$ satellite imagery. This analysis with a learning method also provides precise and objective thresholds to these criteria supporting the selection of images.

The threshold, of 12°, on the variability of the wind direction within the image area allows to reduce both the bias and the spread of the distibution of the emission estimation error, which reflects the uncertainties which should be encountered when tackling actual images. The threshold of 2.1ktCO$_2$/h on the emission budget reduces the spread of the error on the emission estimate. The application of these two criteria simultaneously allows us to separate the synthetic images into two sets: the first, grouping 47% of the synthetic images, for which the distribution of the error on the estimation of emissions has a bias (median) of -7% of the emissions and a spread (IQR) of 56%; and the second for which the distribution of the error has a bias of -31% and a spread of 99% of the emissions. However, some of the subset of results from individual cities show biases on emission estimates of over 10%, despite our filters. These significant remaining biases raise the question of the current reliability of the results obtained on a single given city.

This study provides objective criteria for selecting the most suitable satellite images for urban plume inversion. However, these criteria are derived from experiments with synthetic data, based on atmospheric model simulations and inventories. Even though the realism of these simulations and inventories has been previously evaluated against actual observations, there is a need to confirm the robustness of these criteria and of the corresponding threshold values with applications to real satellite images. Our study, nevertheless, directly supports the interpretation of future inversion results using XCO$_2$ satellite images such as the OCO-3 SAMs.

Code and data availability. Code and Data are available upon requests.

A Illustration of the boundaries of the target areas

B Extension of the study to other inversion methods

B1 Inversion methods

Three other inversion methods have been investigated by Danjou et al. (2024) and tested here: one based on the optimization of a rotating Gaussian plume model (denoted GP3), one based on flux estimates of plume cross sections (denoted CS) and one based on a CO_2 mass balance in the plume (Integrated Mass Enhancement method, denoted IME). Details of these methods can be found in Danjou et al. (2024). Concerning the pre-processing steps (i to vi, cf. Section 3), they are the same for the method based on a rotating plume model as those described in Section 3 for the straight plume model. For the other two inversion methods (CS and IME), the steps of defining the analysis area and estimating the effective wind (steps v and vi) are different: the analysis area is the plume area within one time the radius of the city along the central axis of the plume, and the effective wind is estimated with the wind tangent to the central axis of the plume in the analysis area according to Danjou et al. (2024). The gaussian plume method used in the main body of the article will be now referred as GP2 for clarity.

B2 Preliminary analysis

When we apply the CS, IME and GP3 inversion methods to our 9920 images, we get a result in over 98% of the cases for each of the three methods. At first sight, the error distribution on the emission estimate seems less biased for CS and IME (bias less than 10%) than for GP2 and GP3 (bias between -13 and -16%). The IQRs of the error distributions are however larger for CS and IME (90-91%) than for GP2 and GP3 (78-86%). When we discard synthetic images for which the spatial variability of the wind direction is above 7° (as prescribed in Danjou et al. (2024)), the underestimation of the emissions by the GP2 and GP3 methods disappears: the error distributions have bias between -5 and 7% for the inversions based on GP2 and GP3, as well as for those based on CS and IME. The IQR of the distributions also decreases: it is 75-76% for the inversions based on CS and IME and between 64 and 67% for those based on GP2 and GP3. After filtering, we are left with 53% of the images. For CS, IME and GP3, like the results presented for GP2 in Section 5.1, the relative error on the emissions shows strong disparities between the cities, even after applying the Danjou et al. (2024) filtering based on the spatial variability of the wind direction.

Here we will detail the sensitivities of the error on the emissions estimate to the variables of interest. This qualitative study is much reduced in the Section 5.1 to keep the message concise and clear in the main body of the article.

The emissions are strongly underestimated when the wind is weak or when the spatial variability of the wind direction is strong (see Fig. B1 (a) and (b)): the results are more accurate (lower bias and IQR) when the meteorological conditions favour the ventilation of the emitted CO_2 in a not very diffuse and straight plume, i.e. with high wind speed and low variability of the wind direction; but when the emitted CO_2 accumulates over and in the vicinity of the city in a diffuse plume with high XCO_2 values or forms a plume with a complex structure, the results contain significant errors.
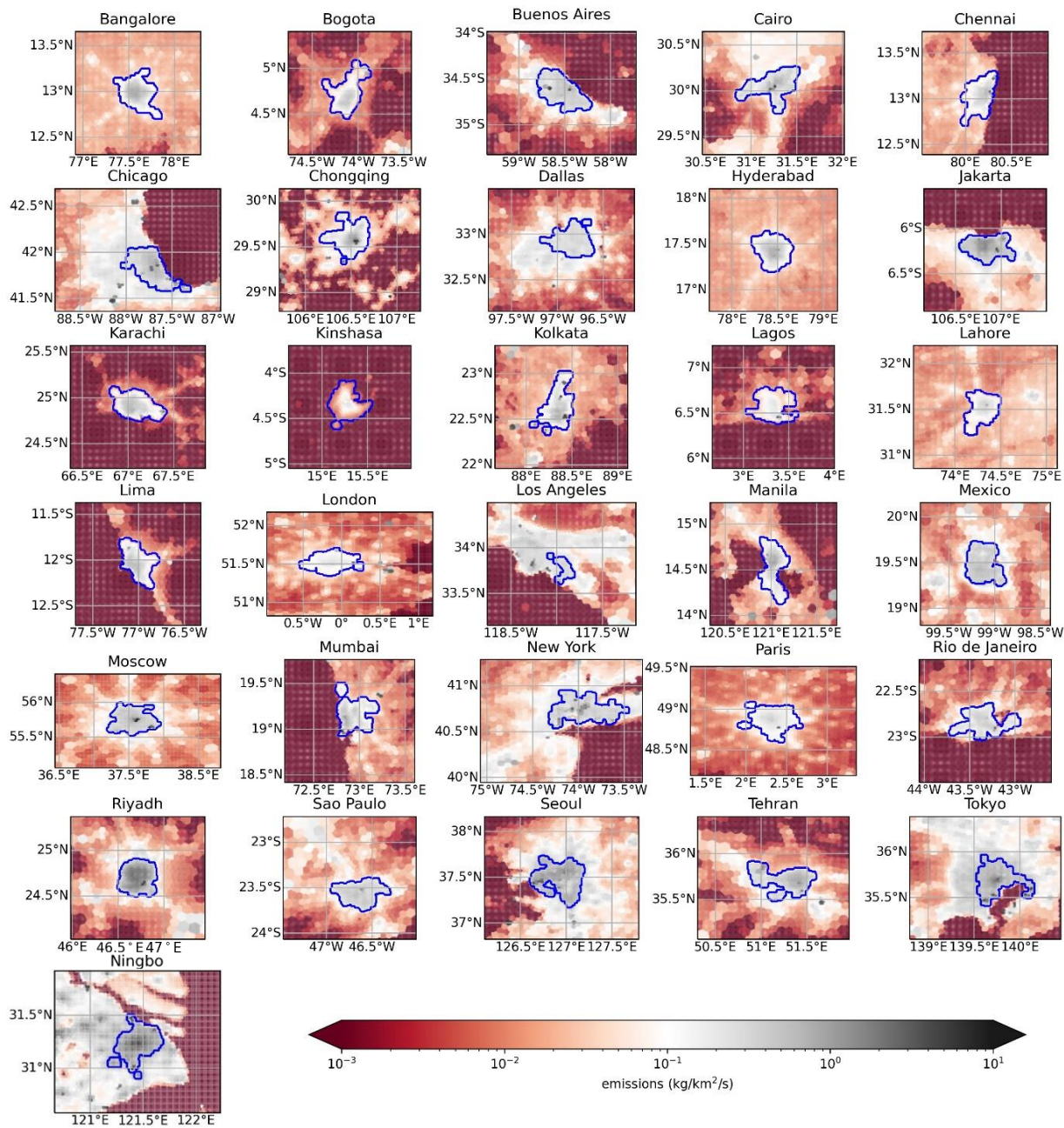
Figure A1. Illustration of the target areas selected for the inversions. The emissions maps are taken from ODIAC. In blue are the boundaries of the areas we are targeting. The frame of the figures coincides with the area covered by the 150km square synthetic images described in Section 2.3
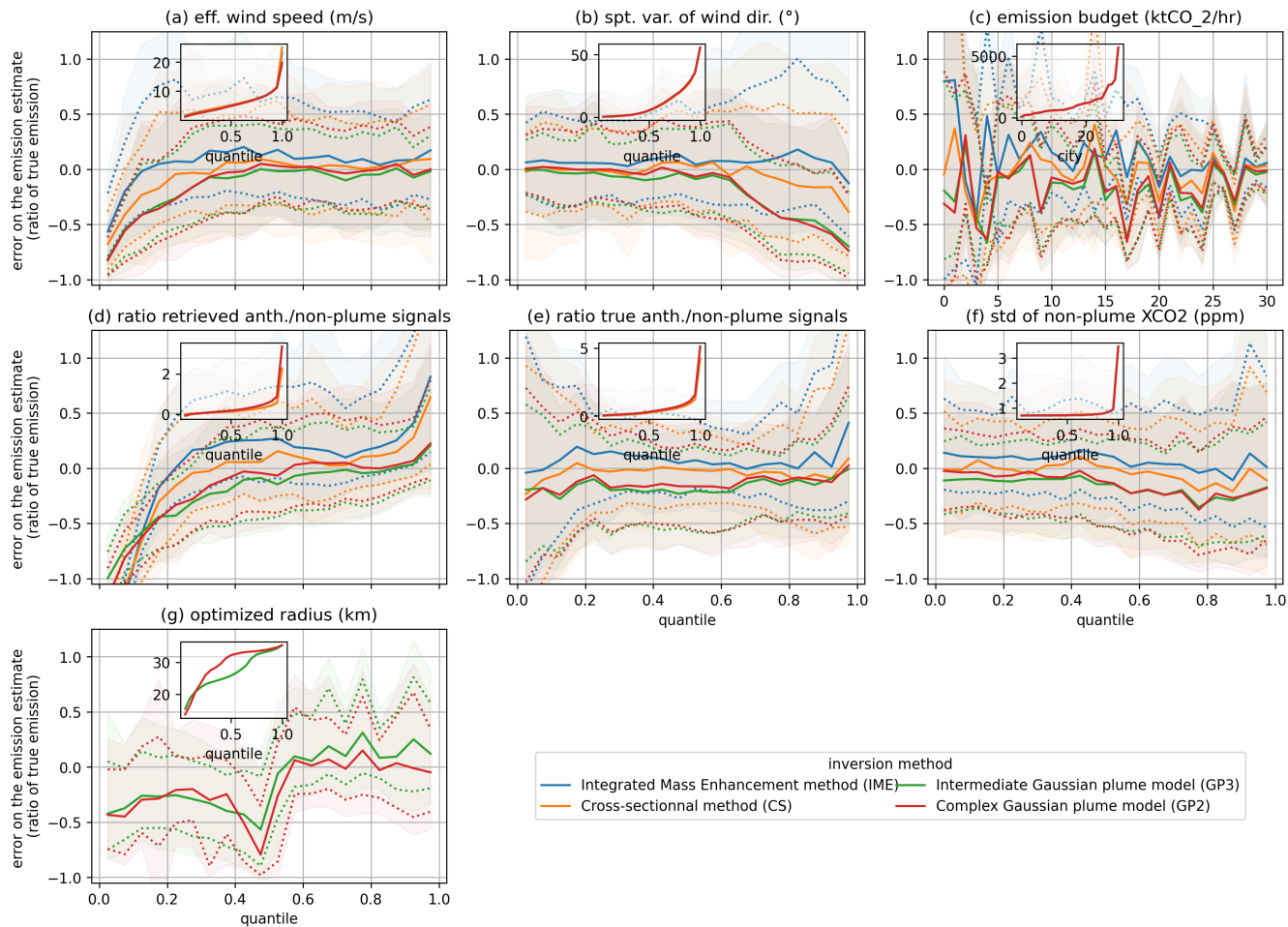
Figure B1. Distribution of the error on the emission estimation in function of different variables of interest. The optimized radius shown in panel (g) is a parameter of the gaussian plume models (see section 4.3) and is therefore not calculated for the other methods.

Figure B1 (c) shows a sensitivity of the error on the emission estimate to the actual emission budget. Despite noise, we can see that the IQR of the error decreases when the emissions increase. Cities with important emissions have a plume that stands 600 out more strongly from the background signal and allow a more accurate emission estimation.

The sensitivity of the error to the ratio of the average anthropogenic signal to the variability of the background signal is shown in Fig. B1 (d) for the estimated anthropogenic signal and (e) for the actual anthropogenic signal. This actual signal-to-background ratio is close to that used by Schuh et al. (2021). The error sensitivities to these two ratios are very similar when this ratio is high, i.e. when the signal from the city differs most strongly from the variability of the background signal. In this 605 case, the estimated anthropogenic signal is close to the real anthropogenic signal. The sensitivities of the error to these two ratios are however very different when these ratios are low. This can be explained by different reasons for the low ratios. For

the estimated background ratio, poorly defined plume boundaries lead to an overestimation of the background signal and thus to a low estimated anthropogenic signal and an underestimation of emissions. For the actual background ratio, low emissions result in a weak anthropogenic signal that is difficult to discern, and thus a higher uncertainty in the emission estimate.

610 Finally, the error on the emission estimate is very sensitive to the radius of the city optimised during the inversion for the inversions based on a Gaussian plume (GP2 and GP3, see Fig. B1 (g)). However, when we discard synthetic images for which the spatial variability of the wind direction is above 7° as prescribed in Danjou et al. (2024), this sensitivity almost disappears. Indeed, when the spatial variability of the wind direction is large, a dome, or at least a very diffuse plume, forms over the city and disturbs the optimisation of the city radius.

615 The error on the emission estimate thus shows sensitivities to several variables, some of which are correlated. These sensitivities can be complex and it is difficult at this stage to determine on the basis of these sensitivities which of the variables are the most discriminating regarding the error on the emission estimation, and thus to determine the optimal criteria for discriminating the synthetic images.

B3 Application of the decision tree method

620 #### B3.1 ... for predictable variables

The application of our learning tree method to inversions with GP3 gives very similar results to those described for GP2 similar to those described for GP2 in Section 4.2. The pair of criteria that emerge is the same (spatial variability of wind direction and emission balance) with a slightly higher number of occurrences (95 for GP3, 82 for GP2). For the inversions with CS and IME, the same pair of criteria is also found, but with a lower number of occurrences (53 and 63 respectively).

625 The distributions of threshold values for the criteria are similar for all methods. The medians of the thresholds found for the spatial variability of wind direction are 10° for the GP3 method, 10° for the CS method and 11° for the IME method. For the emissions bugdet, they are 1.9ktCO$_2$/h for the GP3 method, 2.0ktCO$_2$/h for the CS method and 1.9ktCO$_2$/h for the IME method.

The bias ($<$10%) and IQR (between 52% and 70%) on the emission estimate for the subsets of synthetic images passing the 630 criteria are similar for the different inversion configurations, as well as the size of these subsets (between 36% and 55%). The subsets that do not pass the discrimination criteria show differences depending on the inversion configuration. The results with GP3 are similar to those with GP2 in terms of bias, IQR of the error on the emission estimate and plume size; but for CS and IME the biases are smaller (between 6 and 12% in absolute value for CS and IME, between -25 and -37% for GP2 and GP3) and the IQRs are larger (higher than 120% for CS and IME, lower than 107% for GP2 and GP3).

635 #### B3.2 ... for diagnostic variables

In this section, the set of synthetic images used for the analysis is the one formed by the synthetic images passing the criteria on wind direction variability and on the emission balance. The inversion methods (GP3, IME, CS) are tested separately. For all these methods, no pair of criteria has more than 40 occurrences when the tree depth is set to 2. We therefore also reduce the tree

depth to 1. Plume size appears as the main criterion for IME and CS, with respectively 44 and 42 occurrences. As this criterion appears for less than half of the samples, we do not consider it as sufficiently relevant. For GP3, the error of the optimisation appears as the main criterion, without standing out here too (42 occurrences). We therefore choose not to retain these criteria.

B4   Study of the results by city

As the thresholds distributions are similar for all inversion methods, we choose to use the same threshold values than those found for the GP2 method (cf Section 5.2.1). We have the same 5 cities (Bogota, Lima, Los Angeles, Mexico City and Tehran) for which more than 90% of the synthetic images do not pass the selection criterion based on the spatial variability of the wind direction.
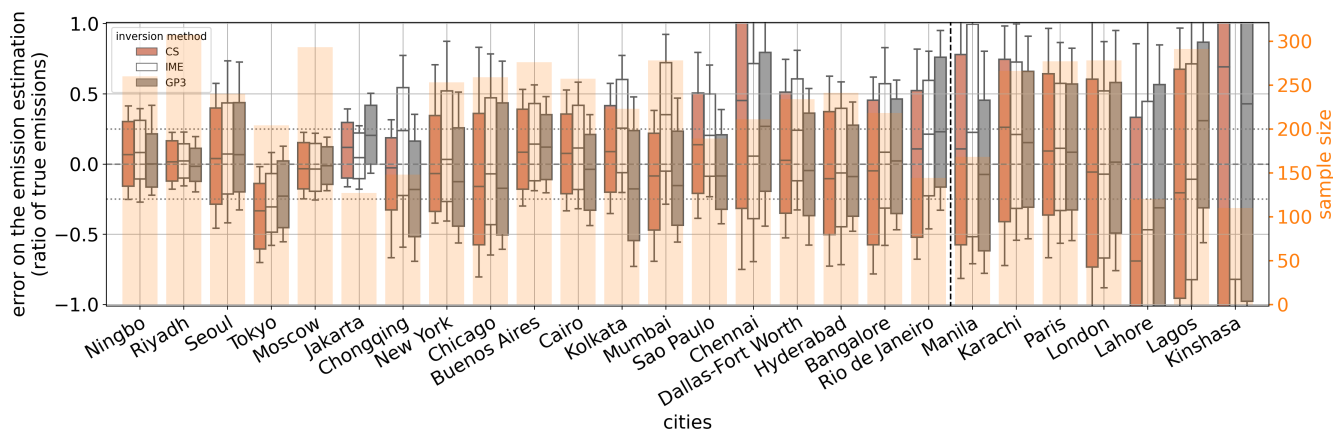
Figure B2. Distribution of the error on the emission estimates (boxplot) for the synthetic images passing the selection criterion on the spatial variability of the wind direction ($> 12°$). The orange bars show the number of synthetic images used. The dotted line separates the cities according to the discrimination criterion on the city's emissions budget with on the left the cities passing the criterion (emissions $> 2.1\text{ktCO}_2/\text{h}$) and on the right the cities not passing it. The cities are ranked in descending order according to their emissions budget.

As with the GP2 method, we can see (cf Fig. B2) that spread of the error generally decreases whith higher cities' emissions. However, here again, this parameter does not fully explain the disparity of the results between cities.

Author contributions. AD performed the data analysis and wrote most of the manuscript. AS ran the OLAM simulations, provided helpful explanations on the outputs, wrote part of Sections 2.1 and 2.2 and was at the origin of Section 5.3 with its pertinent comments. GB and TL closely supervised the redaction, took part in the design of the data analysis and improved the quality of both the scientific content and the clarity of the manuscript by their careful reviews. FMB supervised the work and gave useful remarks on the study and the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors want to thank Elena Fillola Mayoral from University of Bristol who corrected and rewrote the general
655 description of the decision tree algorithm (Section 4.2.1) during the revision process.This work was supported by the CNES (Centre National
d'Etudes Spatiales), in the frame of the TOSCA OCO-3-city project.

References

Broquet, G., Bréon, F.-M., Renault, E., Buchwitz, M., Reuter, M., Bovensmann, H., Chevallier, F., Wu, L., and Ciais, P.: The potential of satellite spectro-imagery for monitoring CO_2 emissions from large cities, Atmospheric Measurement Techniques, 11, 681–708, https://doi.org/10.5194/amt-11-681-2018, 2018.

Center For International Earth Science Information Network-CIESIN-Columbia University and International Food Policy Research Institute-IFPRI and The World Bank and Centro Internacional De Agricultura Tropical-CIAT: Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extents Grid, https://doi.org/10.7927/H4GH9FVG, 2011.

Chevallier, F., Broquet, G., Zheng, B., Ciais, P., and Eldering, A.: Large CO 2 emitters as seen from satellite: Comparison to a gridded global emission inventory, Geophysical Research Letters, https://doi.org/10.1029/2021gl097540, 2022.

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.

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., Avis, C., Braverman, A., Castano, R., Chang, A., Chapsky, L., Cheng, C., Connor, B., Dang, L., Doran, G., Fisher, B., Frankenberg, C., Fu, D., Granat, R., Hobbs, J., Lee, R. A., Mandrake, L., McDuffie, J., Miller, C. E., Myers, V., Natraj, V., O'Brien, D., Osterman, G. B., Oyafuso, F., Payne, V. H., Pollock, H. R., Polonsky, I., Roehl, C. M., Rosenberg, R., Schwandner, F., Smyth, M., Tang, V., Taylor, T. E., To, C., Wunch, D., and Yoshimizu, J.: The Orbiting Carbon Observatory-2: First 18 months of science data products, Atmospheric Measurement Techniques, 10, 549–563, https://doi.org/10.5194/amt-10-549-2017, 2017.

Eldering, A., Taylor, T. E., O'Dell, C. W., and Pavlick, R.: The OCO-3 mission: Measurement objectives and expected performance based on 1 year of simulated data, Atmospheric Measurement Techniques, 12, 2341–2370, https://doi.org/10.5194/amt-12-2341-2019, 2019.

Feng, S., Lauvaux, T., Newman, S., Rao, P., Ahmadov, R., Deng, A., Díaz-Isaac, L. I., Duren, R. M., Fischer, M. L., Gerbig, C., Gurney, K. R., Huang, J., Jeong, S., Li, Z., Miller, C. E., O'Keeffe, D., Patarasuk, R., Sander, S. P., Song, Y., Wong, K. W., and Yung, Y. L.: Los Angeles megacity: A high-resolution land-atmosphere modelling system for urban CO_2 emissions, Atmospheric Chemistry and Physics, 16, 9019–9045, https://doi.org/10.5194/acp-16-9019-2016, 2016.

Frankenberg, C., Thorpe, A. K., Thompson, D. R., Hulley, G., Kort, E. A., Vance, N., Borchardt, J., Krings, T., Gerilowski, K., Sweeney, C., Conley, S., Bue, B. D., Aubrey, A. D., Hook, S., and Green, R. O.: Airborne methane remote measurements reveal heavytail flux distribution in Four Corners region, Proceedings of the National Academy of Sciences of the United States of America, 113, 9734–9739, https://doi.org/10.1073/pnas.1605617113, 2016.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Quéré, C. L., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., Arneth, A., Arora, V. K., Bates, N. R., Becker, M., Bellouin, N., Bittig, H. C., Bopp, L., Chevallier, F., Chini, L. P., Cronin, M., Evans, W., Falk, S., Feely, R. A., Gasser, T., Gehlen, M., Gkritzalis, T., Gloege, L., Grassi, G., Gruber, N., Özgür Gürses, Harris, I., Hefner, M., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jain, A. K., Jersild, A., Kadono, K., Kato, E., Kennedy, D., Goldewijk, K. K., Knauer, J., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lindsay, K., Liu, J., Liu, Z., Marland, G., Mayot, N., McGrath, M. J., Metzl, N., Monacci, N. M., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K., Ono, T., Palmer, P. I., Pan, N., Pierrot, D., Pocock, K., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Rodriguez, C., Rosan, T. M., Schwinger, J., Séférian, R., Shutler, J. D., Skjelvan, I., Steinhoff, T., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tanhua, T., Tans, P. P., Tian, X., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R.,

Walker, A. P., Wanninkhof, R., Whitehead, C., Wranne, A. W., Wright, R., Yuan, W., Yue, C., Yue, X., Zaehle, S., Zeng, J., and Zheng, B.: Global Carbon Budget 2022, Earth System Science Data, 14, 4811–4900, https://doi.org/10.5194/essd-14-4811-2022, 2022.

Grell, G. A. and Dévényi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophysical Research Letters, 29, 10–13, https://doi.org/10.1029/2002GL015311, 2002.

Grell, G. A. and Freitas, S. R.: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, Atmospheric Chemistry and Physics, 14, 5233–5250, https://doi.org/10.5194/acp-14-5233-2014, 2014.

Gurney, K. R., Liang, J., Patarasuk, R., Song, Y., Huang, J., and Roest, G.: The Vulcan Version 3.0 High-Resolution Fossil Fuel CO_2 Emissions for the United States, Journal of Geophysical Research: Atmospheres, 125, https://doi.org/10.1029/2020JD032974, 2020.

Gurney, K. R., Liang, J., Roest, G., Song, Y., Mueller, K., and Lauvaux, T.: Under-reporting of greenhouse gas emissions in U.S. cities, Nature Communications, 12, 1–7, https://doi.org/10.1038/s41467-020-20871-0, 2021.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Sabater, J. M., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on pressure levels from 1959 to present, https://doi.org/10.24381/cds.bd0915c6, 2018.

Kiel, M., Eldering, A., Roten, D. D., Lin, J. C., Feng, S., Lei, R., Lauvaux, T., Oda, T., Roehl, C. M., Blavier, J. F., and Iraci, L. T.: Urban-focused satellite CO_2 observations from the Orbiting Carbon Observatory-3: A first look at the Los Angeles megacity, Remote Sensing of Environment, 258, https://doi.org/10.1016/j.rse.2021.112314, 2021.

Krings, T., Gerilowski, K., Buchwitz, M., Reuter, M., Tretner, A., Erzinger, J., Heinze, D., Pflüger, U., Burrows, J. P., and Bovensmann, H.: MAMAP - A new spectrometer system for column-averaged methane and carbon dioxide observations from aircraft: Retrieval algorithm and first inversions for point source emission rates, Atmospheric Measurement Techniques, 4, 1735–1758, https://doi.org/10.5194/amt-4-1735-2011, 2011.

Kuhlmann, G., Brunner, D., Broquet, G., and Meijer, Y.: Quantifying CO_2 emissions of a city with the Copernicus Anthropogenic CO_2 Monitoring satellite mission, Atmospheric Measurement Techniques Discussions, 1, 1–33, https://doi.org/10.5194/amt-2020-162, 2020.

Lei, R., Feng, S., Danjou, A., Broquet, G., Wu, D., Lin, J. C., O'Dell, C. W., and Lauvaux, T.: Fossil fuel CO_2 emissions over metropolitan areas from space: A multi-model analysis of OCO-2 data over Lahore, Pakistan, Remote Sensing of Environment, 264, 0–11, https://doi.org/10.1016/j.rse.2021.112625, 2021.

Lei, R., Feng, S., Xu, Y., Tran, S., Ramonet, M., Grutter, M., Garcia, A., Campos-Pineda, M., and Lauvaux, T.: Reconciliation of asynchronous satellite-based NO2 and XCO_2 enhancements with mesoscale modeling over two urban landscapes, Remote Sensing of Environment, 281, 113 241, https://doi.org/10.1016/j.rse.2022.113241, 2022.

Lespinas, F., Wang, Y., Broquet, G., Bréon, F. M., Buchwitz, M., Reuter, M., Meijer, Y., Loescher, A., Janssens-Maenhout, G., Zheng, B., and Ciais, P.: The potential of a constellation of low earth orbit satellite imagers to monitor worldwide fossil fuel CO_2emissions from large cities and point sources, Carbon Balance and Management, 15, 1–12, https://doi.org/10.1186/s13021-020-00153-4, 2020.

Lian, J., Wu, L., Bréon, F.-M., Broquet, G., Vautard, R., Zaccheo, T. S., Dobler, J., and Ciais, P.: Evaluation of the WRF-UCM mesoscale model and ECMWF global operational forecasts over the Paris region in the prospect of tracer atmospheric transport modeling, Elementa, 6, https://doi.org/10.1525/elementa.319, 2018.

Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D. B., and Crisp, D.: Quantifying CO_2 Emissions From Individual Power Plants From Space, Geophysical Research Letters, 44, 10,045–10,053, https://doi.org/10.1002/2017GL074702, 2017.

730   Nassar, R., Moeini, O., paul Mastrogiacomo, J., Dell, C. W. O., Nelson, R. R., Kiel, M., Chatterjee, A., Eldering, A., and Crisp, D.: Tracking CO 2 emission reductions from space : A case study at Europe's largest fossil fuel power plant, pp. 1–15, https://doi.org/10.3389/frsen.2022.1028240, 2022.

Nielsen-Gammon, J. W., Powell, C. L., Mahoney, M. J., Angevine, W. M., Senff, C., White, A., Berkowitz, C., Doran, C., and Knupp, K.: Multisensor estimation of mixing heights over a coastal city, Journal of Applied Meteorology and Climatology, 47, 27–43,
735   https://doi.org/10.1175/2007JAMC1503.1, 2008.

Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic CO_2, version 2016 (ODIAC2016): A global monthly fossil fuel CO_2 gridded emissions data product for tracer transport simulations and surface flux inversions, Earth System Science Data, 10, 87–107, https://doi.org/10.5194/essd-10-87-2018, 2018.

Pasquill, F.: The estimation of the dispersion of windborne material, Meteorological Magazine, 90(1063), 33–49, 1961.

740   Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M., Pétron, G., Hirsch, A. I., Worthy, D. E., Werf, G. R. V. D., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North
745   American carbon dioxide exchange: CarbonTracker, Proceedings of the National Academy of Sciences of the United States of America, 104, 18 925–18 930, https://doi.org/10.1073/pnas.0708986104, 2007.

Pillai, D., Buchwitz, M., Gerbig, C., Koch, T., Reuter, M., Bovensmann, H., Marshall, J., and Burrows, J. P.: Tracking city CO_2 emissions from space using a high-resolution inverse modelling approach: A case study for Berlin, Germany, Atmospheric Chemistry and Physics, 16, 9591–9610, https://doi.org/10.5194/acp-16-9591-2016, 2016.

750   Reuter, M., Buchwitz, M., Schneising, O., Krautwurst, S., O'Dell, C. W., Richter, A., Bovensmann, H., and Burrows, J. P.: Towards monitoring localized CO_2; emissions from space: co-located regional CO_2; and NO_2; enhancements observed by the OCO-2 and S5P satellites, Atmospheric Chemistry and Physics Discussions, pp. 1–19, https://doi.org/10.5194/acp-2019-15, 2019.

Schuh, A. E., Otte, M., Lauvaux, T., and Oda, T.: Far-field biogenic and anthropogenic emissions as a dominant source of variability in local urban carbon budgets: A global high-resolution model study with implications for satellite remote sensing, Remote Sensing of
755   Environment, 262, 112 473, https://doi.org/10.1016/j.rse.2021.112473, 2021.

Sierk, B., Fernandez, V., Bézy, J.-L., Meijer, Y., Durand, Y., Courrèges-Lacoste, G. B., Pachot, C., Löscher, A., Nett, H., Minoglou, K., Boucher, L., Windpassinger, R., Pasquet, A., Serre, D., and te Hennepe, F.: The Copernicus CO2M mission for monitoring anthropogenic carbon dioxide emissions from space, p. 128, https://doi.org/10.1117/12.2599613, 2021.

Taylor, T. E., Eldering, A., Merrelli, A., Kiel, M., Somkuti, P., Cheng, C., Rosenberg, R., Fisher, B., Crisp, D., Basilio, R., Bennett, M.,
760   Cervantes, D., Chang, A., Dang, L., Frankenberg, C., Haemmerle, V. R., Keller, G. R., Kurosu, T., Laughner, J. L., Lee, R., Marchetti, Y., Nelson, R. R., O'Dell, C. W., Osterman, G., Pavlick, R., Roehl, C., Schneider, R., Spiers, G., To, C., Wells, C., Wennberg, P. O., Yelamanchili, A., and Yu, S.: OCO-3 early mission operations and initial (vEarly) XCO_2 and SIF retrievals, Remote Sensing of Environment, 251, 112 032, https://doi.org/10.1016/j.rse.2020.112032, 2020.

Ullrich, P. A., Jablonowski, C., Kent, J., Lauritzen, P. H., Nair, R., Reed, K. A., Zarzycki, C. M., Hall, D. M., Dazlich, D., Heikes, R., Konor,
765   C., Randall, D., Dubos, T., Meurdesoif, Y., Chen, X., Harris, L., Kühnlein, C., Lee, V., Qaddouri, A., Girard, C., Giorgetta, M., Reinert, D., Klemp, J., Park, S. H., Skamarock, W., Miura, H., Ohno, T., Yoshida, R., Walko, R., Reinecke, A., and Viner, K.: DCMIP2016: A

review of non-hydrostatic dynamical core design and intercomparison of participating models, Geoscientific Model Development, 10, 4477–4509, https://doi.org/10.5194/gmd-10-4477-2017, 2017.

UNDESA: World Urbanization Prospects, vol. 12, https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf, 2018.

770 UNFCCC: Report of the Conference of the Parties on its nineteenth session (FCCC/CP/2013/10/Add.3), UNFCCC Conference of the Parties, pp. 1–54, http://unfccc.int/resource/docs/2013/cop19/eng/10a03.pdf, 2013.

Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O., Xia, Y., and Huang, Y.: Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes, Atmospheric Measurement Techniques, 11, 5673–5686, https://doi.org/10.5194/amt-11-5673-2018, 2018.

775 Varon, D. J., McKeever, J., Jervis, D., Maasakkers, J. D., Pandey, S., Houweling, S., Aben, I., Scarpelli, T., and Jacob, D. J.: Satellite Discovery of Anomalously Large Methane Point Sources From Oil/Gas Production, Geophysical Research Letters, https://doi.org/10.1029/2019GL083798, 2019.

Varon, D. J., Jacob, D. J., Jervis, D., and McKeever, J.: Quantifying Time-Averaged Methane Emissions from Individual Coal Mine Vents with GHGSat-D Satellite Observations, Environmental science & technology, 54, 10 246–10 253, https://doi.org/10.1021/acs.est.0c01213,
780 2020.

Walko, R. L. and Avissar, R.: The Ocean-Land-Atmosphere Model (OLAM). Part I: Shallow-water tests, Monthly Weather Review, 136, 4033–4044, https://doi.org/10.1175/2008MWR2522.1, 2008a.

Walko, R. L. and Avissar, R.: The Ocean-Land-Atmosphere Model (OLAM). Part II: Formulation and tests of the nonhydrostatic dynamic core, Monthly Weather Review, 136, 4045–4062, https://doi.org/10.1175/2008MWR2523.1, 2008b.

785 Wang, Y., Broquet, G., Ciais, P., Chevallier, F., Vogel, F. R., Wu, L., Yin, Y., Wang, R., and Tao, S.: Potential of European 14CO$_2$ observation network to estimate the fossil fuel CO$_2$ emissions via atmospheric inversions, Atmospheric Chemistry and Physics, 18, 4229–4250, https://doi.org/10.5194/acp-18-4229-2018, 2018.

Wang, Y., Ciais, P., Broquet, G., Bréon, F. M., Oda, T., Lespinas, F., Meijer, Y., Loescher, A., Janssens-Maenhout, G., Zheng, B., Xu, H., Tao, S., Gurney, K. R., Roest, G., Santaren, D., and Su, Y.: A global map of emission clumps for future monitoring of fossil fuel CO$_2$
790 emissions from space, Earth System Science Data, 11, 687–703, https://doi.org/10.5194/essd-11-687-2019, 2019.

Wilson, A. M. and Jetz, W.: Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions, PLoS Biology, 14, 1–20, https://doi.org/10.1371/journal.pbio.1002415, 2016.

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.

795 Wu, D., Lin, J. C., Fasoli, B., Oda, T., Ye, X., Lauvaux, T., Yang, E. G., and Kort, E. A.: A Lagrangian approach towards extracting signals of urban CO$_2$ emissions from satellite observations of atmospheric column CO$_2$ (XCO$_2$): X-Stochastic Time-Inverted Lagrangian Transport model ("X-STILT v1"), Geoscientific Model Development, 11, 4843–4871, https://doi.org/10.5194/gmd-11-4843-2018, 2018.

Wu, D., Liu, J., Wennberg, P. O., Palmer, P. I., Nelson, R. R., Kiel, M., and Eldering, A.: Towards sector-based attribution using intra-city variations in satellite-based emission ratios between CO 2 and CO, Atmospheric Chemistry and Physics, pp. 1–32, https://doi.org/10.5194/acp-
800 2021-1029, 2022.

Ye, X., Lauvaux, T., Kort, E. A., Oda, T., Feng, S., Lin, J. C., Yang, E. G., and Wu, D.: Constraining Fossil Fuel CO$_2$ Emissions From Urban Area Using OCO-2 Observations of Total Column CO$_2$, Journal of Geophysical Research: Atmospheres, 125, 1–29, https://doi.org/10.1029/2019JD030528, 2020.

Zheng, T., Nassar, R., and Baxter, M.: Estimating power plant CO_2 emission using OCO-2 XCO_2 and high resolution WRF-Chem simulations, Environmental Research Letters, 14, https://doi.org/10.1088/1748-9326/ab25ae, 2019.

805