CoCO2

Prototype system for a Copernicus CO_2 service

D4.5 Local (city and power plant) scale inversion using transport models

Co-ordinated by ECMWF

CoCO2

Prototype system for a Copernicus CO_2 service

D4.5 Local (city and power plant) scale inversion using transport models

Dissemination Level: Confidential

Author(s): Joffrey Dumont Le Brazidec, Pierre Vanderbecken, Alban Farchi, Marc Bocquet (ENPC) Thomas Kaminski (iLab), Peter Rayner (iLab), Michael Voßbeck (iLab), Wolfgang Knorr (iLab), Sander Houweling (VUA) Gregoire Broquet (CEA), Michał Gałkowski (AGH), Tia Scarpelli (UEdin), Paul Palmer (UEdin)

Date: 22/02/2023

Version: 0.1

Contractual Delivery Date: 31/02/2023

Work Package/ Task: WP4/ T4.3

Document Owner: ENPC

Contributors: ENPC /iLab/UEDIN

Status: for Review

CoCO2: Prototype system for a Copernicus CO_2 service

Coordination and Support Action (CSA)
H2020-IBA-SPACE-CHE2-2019 Copernicus evolution – Research activities in support of a European operational monitoring support capacity for fossil CO2 emissions

Project Coordinator: Dr Richard Engelen (ECMWF)
Project Start Date: 01/01/2021
Project Duration: 36 months

Published by the CoCO2 Consortium

Contact:
ECMWF, Shinfield Park, Reading, RG2 9AX, richard.engelen@ecmwf.int

The CoCO2 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958927.

Table of Contents

1 Executive Summary

This deliverable documents the development of atmospheric inversion approaches at the local scale relying on high-resolution transport models for the quantification of city and industrial plant CO2 emissions based on spaceborne images of their XCO2 plumes, and potentially of images of pollutants like CO and NO2. The use of high-resolution transport models should help solving for the chemistry and providing estimates under atmospheric transport conditions that are too complex for the methods documented and analysed in deliverables D4.3 and D4.4. It could also provide insights into the spatial and sectoral distribution of the emission within the cities, or allow the co-assimilation of pollutants (CO, NO2) co-emitted by the fossil fuel combustion for the emission quantification. This ambitious objective requires limiting the impact of model errors. The deliverable documents three complementary studies on this topic, and synthesises their conclusions. The aim of the three studies is respectively to

a) develop advanced local scale inversion methods exploiting the modelling skills appropriately despite model errors, potentially using co-emitted species

b) assess the ability to solve for the spatial and/or sectoral resolution of the emissions within the cities and strengthen their emission budget via the use of optimal control vectors for city-scale inversion systems, and potentially via the co-assimilation of co-emitted species

c) assess the reliability of the co-assimilation of co-emitted species for the local scale inversions.

These three studies can be summarised as follows.

a) Segmentation and inversion of XCO2 images using Convolutional Neural Networks
Here, we use transport models to train convolutional neural networks (CNN) which analyse full XCO2 images (potentially with co-registered NO2 images) covering plumes from local sources -cities or power plants-, instead of relying solely on direct comparisons between model simulations and observations at a particular time. The analysis involves two steps: first, detecting the plume in the images, and second, quantifying the source emissions. The method is evaluated with a training and test of the CNN with pseudo CO2M XCO2/NO2 images from transport models covering the plumes from Paris, Berlin and large East German power plants.

When implementing and training the CNN appropriately, the segmentation of the XCO2 images to detect the plume with the CNN is found to be finer than that of the detection algorithms used in T4.2. This result applies when training the CNN with simulation sets that include or not simulations of the target source during the test phase. This demonstrates the ability of the CNN to extrapolate the training to new sources, which is critical in terms of computational cost for applications to large sets of sources. The CNN also demonstrates the ability to handle cases where plumes from different sources overlap. However, the analysis is currently limited to complete images, while gaps are expected in practice.The results of the inversion with the CNN are more exploratory at this stage but they demonstrate the relevance of the general approach. They offer clear perspectives for further developments.

b) Studying the ability to solve for the spatial and sectoral distribution of the emissions within a city with the Carbon Cycle Fossil Fuel Data Assimilation System (CCFFDAS) for Berlin

The CCFFDAS for Berlin is a city-scale inversion system, which includes process representations of anthropogenic emissions and natural fluxes. By contrast to pure transport inversion systems, rather than solving directly for the city emissions it solves for parameters underlying these process formulations. The analyses of synthetic CO2M errors within this CCFFDAS indicate that city-scale data assimilation systems can provide useful estimates of the total city emissions from non-energy generation when relying on XCO2 images only and that the co-assimilation of NO2 images achieves a sizeable reduction in the uncertainty in the emission of the individual districts of the city. The system also shows capabilities to solve for the emissions from individual power plants. However, in general, for non-energy generation emissions the above data streams are not sufficient to provide relevant

estimates at spatial resolutions finer than that of a district. Specific tests are conducted to investigate the impact of the potential variations of the NOx/CO2 emission ratios within the city when co-assimilating the NO2 images.

c) Analysis of the ratio between CO_2 and co-emitted species in actual observation for plumes from cities and large industrial plants

The third study targeted the analysis of the spatial and temporal variability of the CO-to-CO_2 concentration ratios in plumes from emission hotspots in Europe, in actual data from TROPOMI (for CO) and OCO-2 (for XCO_2). Current schemes for co-assimilating data from co-emitted species rely on emission ratios relatively constant in time or homogeneous in space. The analysis aims at evaluating the reliability of such an approach. Overlaps of OCO-2 with plumes identified in TROPOMI NO_2 images are searched for, testing different methods and varying the spatial and temporal averaging of OCO-2 observations in an effort to increase the number of identified overlaps. The outcome is a lack of overlapping between the set of plumes detected in the TROPOMI NO_2 data and in the OCO-2 data, which prevents from deriving robust statistics on the variability of the TROPOMI CO-to-OCO-2 CO_2 concentration ratios. As a result, the analysis highlights the challenges of the co-assimilation of CO and other trace gases and CO_2 data from the existing missions for local scale inversions, at least in Europe.

These different studies confirm the potential of using atmospheric transport models to improve the accuracy of the plume detection and inversion compared to the light plume techniques studied in deliverables D4.3 and D4.4, and to infer the emissions from specific plants or districts within a city. They also reveal the need for further developments and analysis, and the joint analysis of NO2, CO and CO2 satellite data raises warnings regarding the ability to co-assimilate data on co-emitted species from the current instruments. However, these studies provide promising perspectives regarding the potential to increase the accuracy of the local inversions, the potential of merging the different approaches analysed here, and the co-assimilation of the NO2 and CO2 images that will be co-registered by the future CO2M mission.

2 Introduction

2.1 Background

WP4 aims to develop and apply high-resolution atmospheric transport and inversion systems for limited geographical areas and various scales, such as mesoscale weather phenomena and individual plumes as seen by CO2M. The developed systems are to be improved for efficiency and robustness to be used operationally, and fed with anthropogenic emissions and ecosystem exchange data from other WPs. National data and inventories will also be considered for some inversions. One goal is to increase consistency within and between local and regional inversions, and to provide local and regional estimates with uncertainty information to be used in the global prototype.

In particular, WP4.3 is dedicated to the development of advanced local “plume inversion” approaches relying on atmospheric transport models to process the images, unlike WP4.2. The use of local scale atmospheric transport models should help overcome complex atmospheric transport conditions, which challenge the simple assumptions underlying the light plume inversion techniques that are tested in WP4.2. Atmospheric transport models also provide some potential to solve for the spatial and sectoral distribution of the emissions within urban areas, based on the capability to trace the signal from any component of the urban emissions. This potential is enhanced by the capability to co-assimilate in these models CO2 data with data on the pollutants co-emitted by the fossil-fuel combustion. Indeed, the co-assimilation of co-emitted species like NO2 (which will be co-registered with XCO2 by the CO2M mission, and which is monitored by various missions/instruments including Sentinel-5P/TROPOMI) and CO (monitored by various missions/instruments including Sentinel-5P/TROPOMI) should strengthen the ability to distinguish between biogenic fluxes and anthropogenic emissions from different sectors or areas within a city when quantifying and attributing the urban emissions. WP4.2 assessed the use of the co-emitted species to support the CO2 plume detection and inverted plumes of NO2, but it did not assess such a co-assimilation for the emission quantification.

However, traditional atmospheric inversion approaches use atmospheric transport models as a strong constraint in the data assimilation process, summarising the model errors on the simulation for a given time and location into a random and unbiased noise without spatial or temporal correlations, while the plume modelled by such models may easily deviate in average direction and shape from the actual ones. Such methods may thus be less flexible and may be prone to a poorer analysis of the satellite images than the light plume inversion techniques tested in WP4.2. Furthermore, the co-assimilation of co-emitted species relies on strong assumptions regarding the emission ratios between species. In particular, it often relies on average values from inventories while there can be large and uncertain variations of these ratios over wide ranges of spatial and temporal resolutions.

The overall aims of this deliverable are thus to assess the full potential of using local scale atmospheric transport models to tackle complex transport conditions and solve for the spatial and sectoral resolution of the urban emissions. This is done by developing methods to exploit the skills of the atmospheric transport model by overcoming the modelling uncertainties, and assessing the potential of the co-assimilation of co-emitted species.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverable

This document presents a perspective on local scale inversion for estimating CO2 emissions from cities and power plants, using atmospheric transport modelling and satellite images, in particular those from the future CO2M mission.

The overall aims detailed above are addressed in this document as 3 distinct objectives:

(i) the ultimate development of an advanced local scale inversion method overcoming the transport modelling errors and exploiting the modelling skills appropriately

(ii) assessing the ability to solve for the spatial and/or sectoral resolution of the emissions within the cities and strengthening their emission budget via the use of optimal control vectors for city-scale inversion systems.

(iii) assessing the (iii-a) potential and (iii-b) reliability of the co-assimilation of co-emitted species for the local scale inversions

For (i), the concept is to create a plume inversion system that incorporates machine/deep learning techniques, in addition to the full 4D meteo-transport models, to tackle complex cases. To achieve this goal, a comprehensive simulation database generated by a 4D meteo-transport model is essential. The goal is to assess the performance and accuracy of the advanced local inversion approach and determine its suitability for operational use. To estimate emissions from the plume, it is essential to detect it on satellite images. Thus, the detection of a plume, i.e. the identification of its contour, in a satellite image is a critical step in the evaluation of source emissions. An intermediate objective is therefore the development of a method to accurately detect or segment plumes in satellite images.

The objectives (ii) and (iii-a) are intrinsically linked since the co-assimilation of the co-emitted species is expected to support the spatial and sectoral attribution of the urban emissions in addition to the distinction of the emissions from the targeted source from surrounding fluxes. These objectives are thus targeted together with city scale inversions assimilating the NO2 images from CO2M in addition to the XCO2 images from this mission based on a so-called Carbon Cycle Fossil Fuel Data Assimilation System (CCFFDAS) developed for the city of Berlin. Both the ability to constrain the city scale emission budgets and that to solve for the spatial and sectoral resolution of the emissions within the cities are assessed with this system.

The objective (iii-b) mainly consists in assessing the knowledge on emission ratios and thus the reliability of the co-assimilation of satellite data on co-emitted species. To that end, actual satellite data are analysed for trace gas enhancement ratios ($\Delta XCO/\Delta XCO_2$) over urban regions in Europe. We use current satellite observations from OCO-2 and TROPOMI for 2018-2019. We use NO_2 plumes identified in the TROPOMI as a marker of polluted urban areas and then identify overlapping OCO-2 observations, testing different methods for finding these overlaps. We then analyse the enhancement ratios for each identified TROPOMI plume and OCO-2 overlap. The goal is to assess the temporal and spatial variability in enhancement ratios, and thus the ability of city-scale inversions co-assimilating these trace gases to accurately represent the observed relationship between CO2 and other trace gases in their inversion systems.

2.2.2 Work performed in this deliverable

The three groups of objectives detailed above (i, ii and ii-a, iii-b) lead to three sets of analysis.

Convolutional Neural Networks trained with atmospheric transport models to process the images

Conventional threshold-based methods for plume detection rely on the signal-to-noise ratio of the plume, where the signal is the enhancement of CO2 inside the plume above the background field and the noise is the variability in measurements due to the precision of the instrument and interference from other anthropogenic and biospheric fluxes. However, as shown by (Kuhlmann, Broquet, et al. 2019), the expected precision of the CO2M CO2 product (<0.7 ppm, MRDv3) results in a signal-to-noise ratio that is too low for reliable detection of CO2 plumes or parts of CO2 plumes using threshold-based methods for many cities and power plants. To overcome this limitation, CO2M will utilise an additional nitrogen dioxide (NO2) instrument on the same platform, which serves as a proxy for CO2 and significantly improves plume detection capabilities.

In this deliverable, we rely on deep learning methods (Chollet 2017; Zhang et al. 2022) to cope with the signal-to-noise ratio (SNR) problem in plume detection and inversion problems.The method should allow for a better assessment of parts of the image with a low signal-to-noise ratio. In other words, the method should improve the assessment (detection or inversion) of low emissions plumes (by assisting in the overall assessment of these plumes). But the method should also improve the assessment of high emissions plumes (by assisting in the assessment of low SNR parts of these plumes). In general, we expect the performance of the deep learning method to be better at distinguishing plumes from other noise streams. In particular, we rely on convolutional neural networks (CNNs) to segment plumes more accurately than thresholding techniques and perform inversion by learning and capturing plume-specific spatial patterns. Plumes may indeed have certain spatial properties or shapes that can be exploited by an algorithm capable of extracting and learning these features. The image dataset used to train and test the CNN model is based on fields of column-averaged dry air mole fractions of CO2 (XCO), simulated in the vicinity of the targeted sources (Grand Paris, Île-de-France (IdF), Berlin, and various power plants). Each image consists of (at least) a targeted source plume and the other nearby biogenic and anthropogenic fluxes, plus the instrumental noise typical of the sensor onboard COM. Clouds are not included in the CO2 images for simplicity.

The CNN segmentation method is compared for reference to the thresholding plume segmentation method described by (Kuhlmann, Broquet, et al. 2019; Kuhlmann et al. 2021), which is available as part of a Python package for data-driven emission quantification
(ddeq; https://gitlab.com/empa503/remote-sensing/ddeq).

Analysis with a city-scale CCFFDAS over Berlin

We employ a Carbon Cycle Fossil Fuel Assimilation System (CCFFDAS) set up for a domain around Berlin at the resolution of CO2M (Kaminski et al., 2022a) to investigate the benefit of co-assimilation of complementary NO2 in general, the impact of the correlations in the uncertainties of NO2:CO2 emission factor ratios, the representation of the uncertainty in the spatial structure of the inflow from the boundary in the inverse problem, and the dependence of the uncertainty in the city's emission estimates on the spatial scale.

Analysis of concentration ratios in current satellite XCO2, CO and NO2 observations

CO2 city-scale inversions can benefit from the co-assimilation of species co-emitted from combustion like CO and NO2. In order to take advantage of these trace species, we need to accurately represent the variability in their enhancement ratios compared to CO2. We can use existing satellites to investigate the observed variability across Europe.

In this deliverable, we use TROPOMI and OCO-2 observations to investigate the temporal and spatial variability in enhancement ratios. We first test out methods to identify overlaps of the two observing systems. Our focus is on using TROPOMI NO2 plumes to identify high pollution regions and then looking for OCO-2 observations that cross the plume in space and time. We then look at the variability in enhancements over Europe and specifically in Madrid, Spain, where a number of plume and OCO-2 overlaps are identified.

2.2.3 Deviations and countermeasures

Even though this deliverable is of type 'public' we would like to delay the public release, in order not to hamper publication in the peer reviewed literature.

This deliverable was highly impacted by staffing problems with the departure of one of the main participants at SRON much earlier than expected, and issues for the expected recruitments (at AGH, and then at SRON to compensate for the early departure, see below). As a result, this deliverable does not document analysis from city scale inversions for Krakow (at AGH) and the Randstad area (at

SRON) while it was initially envisaged. However, the work plan in WP4.3 has been adapted to overcome these issues:

- the topics that had to be investigated in D4.5 have been entirely covered with the analysis by CEREA, iLab and UEdin
- the plan for the analysis by AGH and SRON has been entirely redirected towards the objectives of the final deliverable D4.8

Details regarding the deviation and counter measures for AGH

Implementation of the local inversion system for Kraków (AGH University of Science and Technology) has faced unexpected and significant delays related to the recruitment of personnel. The issue has finally been alleviated in early 2023. Another delay was caused by unexpected necessity to transfer the modelling framework software to the new supercomputer, which was fully completed only in Nov. 2022. At the moment of writing this report the inversion system is actively being implemented, focusing on providing input to the Tasks active within the Project (T4.5, with deliverable D4.8).

Details regarding the deviation and counter measures for VUA

At VUA, we experienced difficulties in replacing personnel who left early in the project. Meanwhile we managed to recover from this situation with the first person starting in October 2022, followed by the second in January 2023. This affected the work on WRF-CTDAS in WP4.3, and a dataset from the DALES model for the Randstad planned in WP4.1. In the meanwhile the Dutch supercomputing system was renewed, and therefore the first task was to get the WRF-CTDAS system, planned to be used for D4.5, up and running on the new system. Since the appointment of new personnel, we have been progressing well but not in time to deliver at the schedule of the D4.5 report. As for partner AGH, we plan to make use of the configuration developed in WP4.3 to deliver results for the D4.8 deliverable.

Details regarding the deviation and counter measures for UEDIN

There was an early departure of the main staff for this deliverable from UEDIN. This has led them to only focus on XCO enhancement ratios rather than on XCO and XNO2 ratios. They find that the existing observing systems prevent in-depth analysis of enhancements due to a lack of satellite overlaps, so their main conclusions are not significantly impacted.

3 Segmentation and Inversion methods

3.1 Simulation of the synthetic datasets

3.1.1 Simulation of the CO2 fields

Two different atmospheric transport models are used to simulate the CO2 fields which provide the XCO2 images.
Simulations in the Paris region by WRF-Chem V3.9.1 are based on the configuration of (Lian et al. 2021) (see Dumont Le Brazidec et al. 2022 for more details) while simulations in the Berlin region including neighbouring power plants are taken from the SMARTCARB project (Kuhlmann, Clément, et al. 2019).

Paris data consist of 3-month meteorology and CO2 transport simulations on a nesting of three domains with different spatial resolutions (1, 5, and 25 km). Initial and boundary conditions (ICBC) are forced with ERA-5 re-analysis fields at a resolution of 0.75 degrees for the meteorological simulations and CAMS 3-hourly update interval global CO2 atmospheric inversion products for the CO2 simulations. High-resolution inventories, the TNO GHGco v3.0 TNO-MACC II and the VERIFY D2.1 v1.0 (Denier van der Gon et al. 2021), are used to simulate CO2 concentrations over the entire domain. Finally, biogenic fluxes are computed with the Vegetation Photosynthesis and Respiration Model (VPRM) (Mahadevan et al. 2008) model coupled online with the WRF-Chem V3.9.1 model.

The SMARTCARB simulations were run with the COSMO-GHG model for a domain centred on Berlin and covering several neighbouring power plants. The simulations were used to generate synthetic CO2M observations and to assess different plume detection and inversion methods (Kuhlmann et al. 2020). The model fields consist of hourly data over one year with a spatial resolution of 0.01 degrees and sixty vertical layers from 0 to 24 km. MeteoSwiss COSMO-7 analyses are used as the meteorological initial and boundary conditions, while the CO2 boundary conditions correspond to the fields of the ECMWF (European Centre for Medium-Range Weather Forecasts) free-running global CO2 simulations with 137 levels . Biogenic CO2 fluxes are modelled offline with the VPRM diagnostic biosphere model. Finally, the TNO-MACC III inventory is used for modelling anthropogenic emissions in most of the regions. Berlin emissions, however, are modelled with the help of a detailed inventory. The main configuration parameters are summarised in (Dumont Le Brazidec et al. 2022).

3.1.2 Parametrisation of the CO2 field simulations

CNN segmentation models are trained and tested on images of fixed size: XCO2 images of 160×160 pixels are extracted from the Paris and SMARTCARB datasets. The images are extracted such that the hotspot is located in the centre of the image and the chosen size ensures that most of the hotspot plume is present in the image. The issue of image integrity (presence of gaps due to satellite swathing or the presence of clouds) is not addressed in this work and should be the subject of a future publication or technical report. The native resolution of km of the SMARTCARB data is maintained during this extraction phase, while Paris data are mapped from the original 200 pixels in longitude and 165 pixels in latitude to 160×160 pixels: the new image concentrations are calculated by cubic spline interpolation which gives images with a resolution of ~1 km in IdF. A wide variety of fields and plumes are needed to train an efficient plume segmentation model. The dataset diversity and size is achieved through:

- seasonal variability (January, March, and August for the Paris data, and a whole year for SMARTCARB);
- geographical variability (Paris, and various locations in Germany);
- plume type variability: single power plant plumes (a single major anthropogenic plume on the image) with Lippendorf, multiple plumes (several major anthropogenic plumes on the image) with Jänschwalde or Boxberg, cities (Grand Paris, Berlin) or cities with an extended suburb (the

Î le-de-France including Paris region, IdF). The Paris data is split into two parts to assess the ability of the CNN to retrieve plumes from the Paris conurbation alone (Grand Paris) or from the entire Paris region (IdF).

To fully account for the detectability factors affecting the signal-to-noise ratio, the satellite instrumental noise must be taken into account. A Gaussian random noise typical of CO2M is used and added to the simulated XCO2 fields. Considering these various factors, the generation of a XCO2 image can be summarised in three steps: simulation of the hotspot anthropogenic plume, addition of the simulated background (biogenic and other anthropogenic fluxes), and addition of the instrument noise. This is illustrated on Fig. 1.

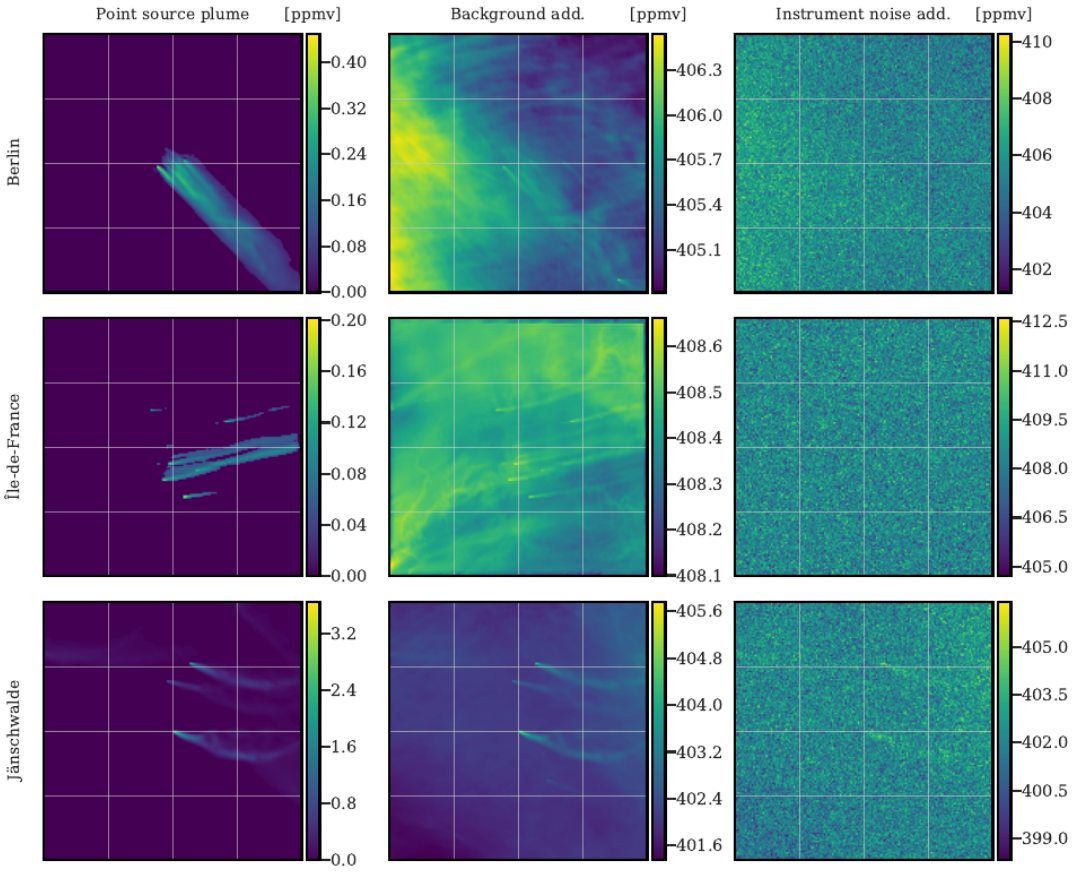

Figure 1: Examples for the construction of three simulated XCO2 satellite images. Each row shows the generation of a sample XCO2 image; three hotspots are considered at random times and days: Berlin (top), Paris (middle), and Jänschwalde (bottom) which is located near other power plants, which explains the presence of multiple plumes. The left column displays the anthropogenic hotspot plumes with the concentrations in ppmv indicated on the colour bars. In the middle column is shown the addition of background (biogenic and anthropogenic fluxes). Finally, on the right column is revealed the full simulated image used as input to the CNN model with the addition of satellite instrument noise.

We provide the CNN model with full noisy images (right panels of Fig. 1) and we design it to return the plume masks of the hotspot plumes (left panels of Fig. 1).

Data augmentation techniques are applied to the training data. The training images are randomly shifted, zoomed, sheared, flipped and rotated variants of the original images. Specifically, each image used for training the CNN has been subject to:

• a random horizontal and vertical shift up to 0.2 (the border values are then used to fill the missing values of the new image, as shown in Fig. 2);

• a random zoom up to 0.2;

• a potential horizontal or vertical flip with a probability of 0.5;

• a random rotation up to 90 degrees;

• a random shear, i.e., a distortion along an axis (while the other axis is fixed), up to 45 degrees.

This data augmentation is meant to (i) raise the performance of the CNN model: data augmentation artificially and substantially increases the number of training data, thus reducing the risk of overfitting, (ii) raise the representativeness of our plume database through the enforcement of geometrical invariance.

Figure 2 shows two examples of data augmented fields and associated plumes.

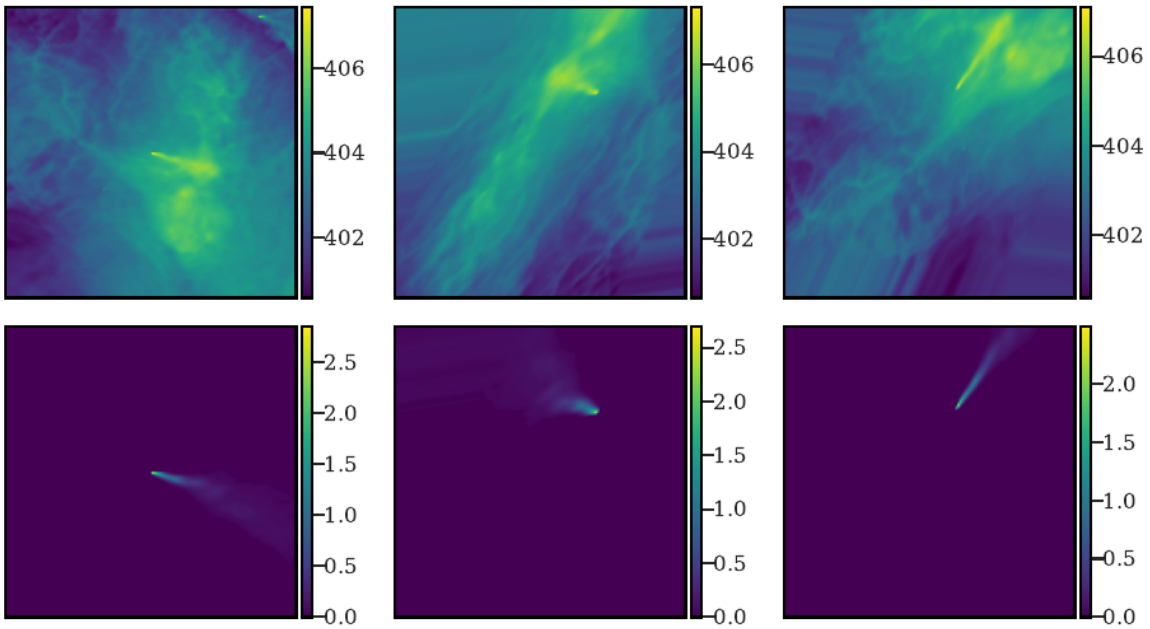

Figure 2: Examples of augmented XCO2 fields (top row, without instrumental noise) and corresponding plumes (bottom row). The left column corresponds to the original XCO2 field and plume. The middle and right columns correspond to the same XCO2 fields, after shearing, flipping, rotating and translating operations. These are typical examples of what is used as input to the CNN model.

Extrapolation or distortion of plumes due to data augmentation can lead to non-physical plumes. Yet, we empirically found that the use of such plumes improves the ability of the CNN model to segment real plumes.

3.2 Segmentation: methodology

3.2.1 Problem description

The plume segmentation problem can be defined, for a given image, as the detection of all pixels composing the plume.

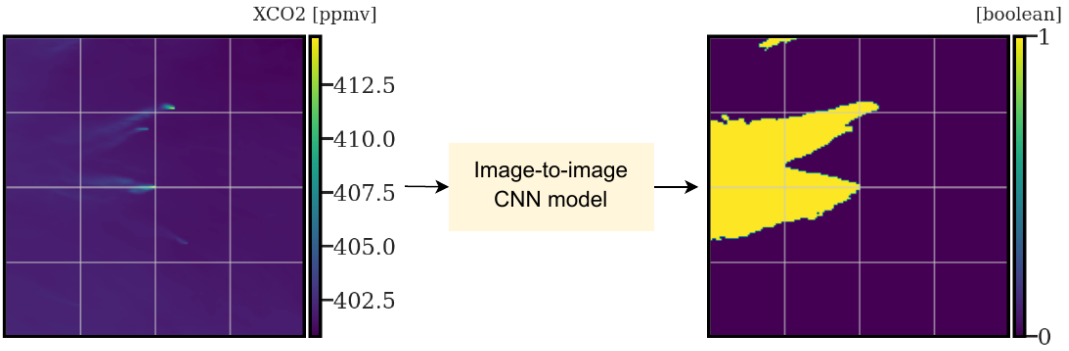

Figure 3: Illustrating the principle of plume segmentation. An image-to-image model (here, a CNN) is used to extract from a XCO2 field the anthropogenic plumes from one or several hotspots.

This problem can be seen as an image-to-image problem, where the goal is to translate the original image into a Boolean map where pixels are assigned to True (part of a hotspot plume) or False (not part of a hotspot plume), as shown in Fig. 3. Segmentation of a plume is different from the simpler problem of detecting the "existence" of a plume, as performed by Finch et al (2022). Many algorithms can be designed to perform such a translation. However, in this deliverable, we dispose of a labelled dataset as both the input XCO2 field and the corresponding targeted plume are available. In this context of supervised learning, for image processing, CNNs are particularly effective .

These algorithms are based on learning specific patterns of increasing complexity using smaller and simpler patterns (the filters). The larger and more complex patterns are specific to the learned targets (here, the plumes from the targeted sources). The filters are optimised to allow the learning of these complex target-specific features. This optimisation is done automatically, unlike most algorithms where the filters would have to be chosen manually (feature engineering).

The CNNs decompose as a training step (which includes validation) and a test step. In the training step, the selected CNN model, described in section 3.3, is trained with XCO2 field/Boolean map pairs. The Boolean map is composed of pixels equal to 1 if the pixel has a positive XCO2 concentration corresponding to the simulated anthropogenic plume, or 0 if the pixel does not. For a given XCO2 field, the CNN model knows the target Boolean plume and learns to output a probability map that best matches it (supervised learning). The shapes of the input and output are equal and each pixel in the output represents the probability that the pixel in the input belongs to the anthropogenic plume. In the testing step, the CNN model is applied to new input images, none of them seen during the learning phase, to assess its ability to generalise to new data.

3.2.2 Loss function

The loss function is a measure of the discrepancy between the truth (the Boolean map representing the real plume) and the prediction (a probability map). Many loss functions can be used, each of them defining what the CNN model should learn from the data, what the priorities are and which differences can be overlooked. The definition of a plume, according to the CNN model, is embedded in the characterisation of the loss function.

A classical loss function used for segmentation problems is the binary cross entropy (bce). This definition uses the plume Boolean map as target (truth, or label), and gives an equal weight to pixels with a high plume concentration and to pixels with a low plume concentration, which is questionable. Two Boolean plumes are shown in Fig. 4: the middle row images represent the transformation of the top row plumes into Boolean targets: images of 0 and 1 depending on whether the plume concentration of the pixel is greater than the threshold ppmv.

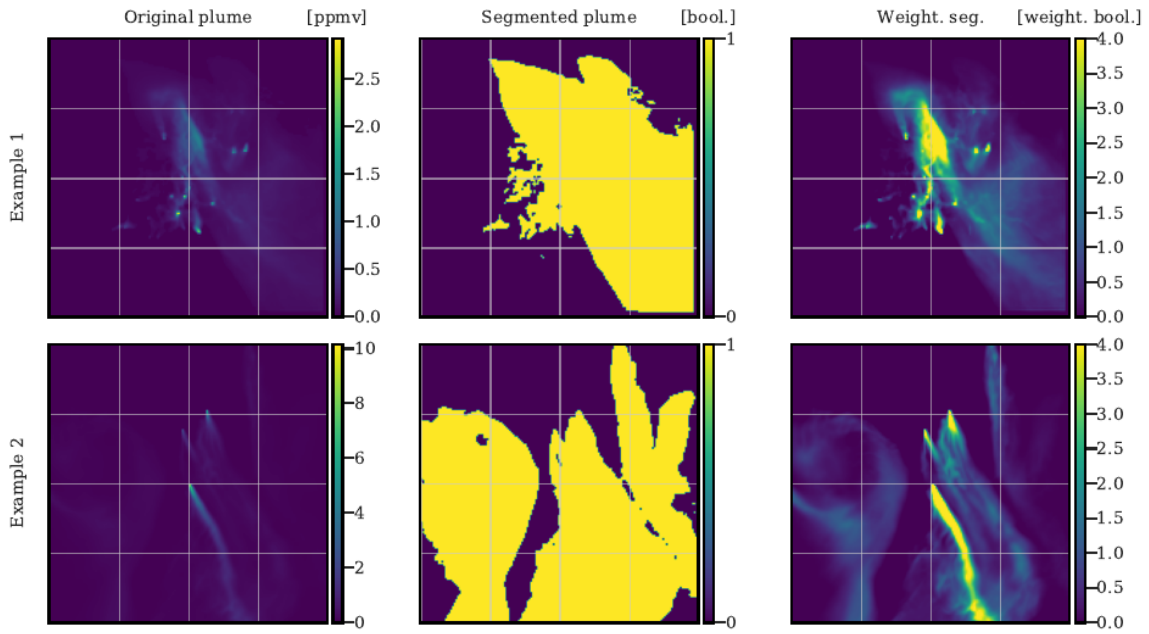

Figure 4: Examples of XCO2 plumes (left), corresponding Boolean maps representing plume positions (middle) and weighted Boolean maps representing the plume positions (right).

These Boolean targets are visually far from representative of the plumes: the bulk of the signal, the mass of CO2, is contained in a much narrower area. In practice, this choice hinders convergence and deeply degrades the performance of the CNN, since many pixels with low plume concentration are difficult to detect. A threshold could be used to generate more representative Boolean targets, but due to the diversity of plume types, no universal threshold exists.

To overcome this problem, the pixel loss is weighted by a function proportional to the plume concentration of the pixel. The weight function, depending on the plume concentration in the pixel, is linear. More details can be found in (Dumont Le Brazidec et al. 2022).With this weighting function, the model:

- is heavily penalised if it makes an error on a pixel with a high plume concentration;
- is penalised very little (even insignificantly) if it makes an error on a pixel associated with a low plume concentration;
- is moderately penalised if it makes an error on a non-plume pixel.

The result of this weighting can be observed on the right column of Fig. 4: each pixel is still a Boolean but weighted depending on the plume concentration of the pixel.

The new loss function is differentiable which is a necessary condition for the application of the gradient descent backpropagation algorithm. Moreover, the weighting is carried out independently for each field/plume pair, and not uniformly for the whole dataset. This latter choice would have penalised low-emission hotspots and favoured high-emission ones. This loss function is referred to as weighted bce (weighted binary cross entropy) or wbce in the following.

3.2.3 U-net model

The deep learning model chosen to address this image-to-image problem follows the U-net architecture, a CNN encoder-decoder originally developed for biomedical image segmentation , but later successively applied in many domains. This architecture is composed of (i) a downsampling or encoder phase where the resolution of the input image decreases and the number of feature channels increases, and (ii) of an upsampling or decoder phase where the resolution is increased to its original

shape while the number of feature channels decreases symmetrically to the downsampling phase. The encoder captures and learns aggregated information locally, progressing until it captures close to the entire image. The encoder works in the same way as, for example, a classification model and can be built with any conventional CNN classifier. The decoder uses the encoded information to build the output. The particularity of the U-net architecture is the use of skip connections where encoded layers are directly carried to the decoder part. In other words, the decoder part collects the high-resolution features from the encoder part through concatenation, to prevent any information loss.

Many encoder and decoder architectures can be used: an example is illustrated in Fig. 5.

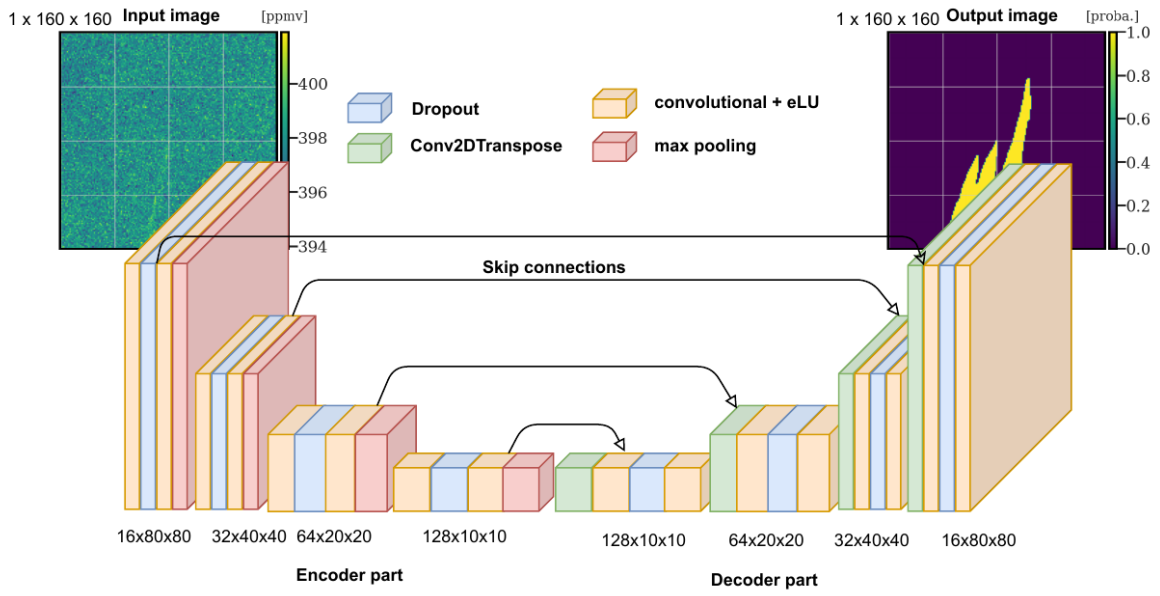

Figure 5: The XCO2 field/plume pairs are fed into a U-Net that learns to distinguish the spatial features of the plume from the background.

Such a U-net is built on top of convolutional layers that locally aggregate the information. Furthermore, the encoder uses maxpooling layers which decrease the resolution of the image, while the decoder resorts to upsampling layers. Finally, dropout layers are used to reduce overfitting. However, in this paper, we use a generalised architecture, not shown for the sake of readability since more than 270 layers and 5 million parameters are used. The encoder used is the EfficientNetB0 CNN architecture (Tan et Le 2020) which is built with specific convolution layers (based on depth-wise convolutions) and a squeeze-and-excitation optimisation. Several encoders have been considered and tested: ResNet , DenseNet and self-made alternatives. The decoder phase is a repetition of convolution and upsampling layers.

A dropout rate of 0.2 is used in the encoder part. The activation layers in the encoder part are swish functions , whereas relu functions are chosen for the decoder part. The normal kernels are chosen to initialise convolutional layers to avoid vanishing or exploding gradients during the first epochs. To get a probability map, the final output is activated by a sigmoid function. We use an initial learning rate of 1E-3 with Adam optimiser and a reduce on plateau strategy after considering different configurations. The batch size is set to 32 samples, and the number of epochs is set to 500 which ensures the convergence of learning. The final model weights are the best performing weights on the validation dataset.

3.2.4 Train, validation and test datasets

The complete dataset is divided into training, validation and test subsets. Since a plume at a certain time strongly resembles the plume of the next hour, the validation and test sets consist of subsets of plumes on two consecutive days. For a given month, the test dataset always consists of the plumes of

the 4th, 5th, 15th, 16th days of the month. The training, validation and test datasets are used to train the model, to tune its hyperparameter and to test the optimal model, respectively.

The input XCO2 fields are standardised using the mean and variance over all pixels of all images of the training data set. All the results of the following section 4 were obtained on the test dataset, unobserved until the final evaluation. Furthermore, the results are obtained on a non augmented test dataset since wbce metrics of the model on an augmented or non-augmented test datasets are similar and we are primarily interested in segmenting non-augmented images.

3.3 Segmentation: application

To evaluate the performances of the CNN plume segmentation, two alternative segmentations to compare to (hereafter called references) are described. Then, the U-net algorithm is trained and tested in two configurations.

In section 4.2, the first configuration, we investigate the ability of the U-net to generalise to new data from the same region. The U-net is trained and tested on pairs of XCO2 and plume images in Grand Paris, IdF, Berlin, Lippendorf and in plume clusters centred at Jänschwalde or Boxberg. Several training setups are considered: the CNN is trained either on all available data or only on data from one location.

In section 4.3, the second configuration, we investigate the ability of the U-net to extrapolate on unseen data from another area. The U-net is trained on Grand Paris, IdF, Jänschwalde, Lippendorf, Boxberg images and tested on Berlin images.

3.3.1 Alternative segmentations to compare to

3.3.1.1 Neutral reference

Two references are considered to assess the quality of the CNN segmentation through their wbce scores. First, we use a constant probability map as a first reference, which is in practice equivalent to a prediction of non-segmentation of the plume. Since the wbce metric only deals with probabilities (rather than Boolean values) this constant value is a probability and must be chosen. For each hotspot dataset, this probability is found as the one minimising the wbce over all images of that hotspot with a differential evolution algorithm. In practice, for each hotspot, the calculated probability is close to 0.15-0.2. This first segmentation reference output is called neutral reference in the following. Figure 6 shows the histograms of the wbce computed on the plume cluster centred at Boxberg and the Berlin plume with respect to the neutral reference.

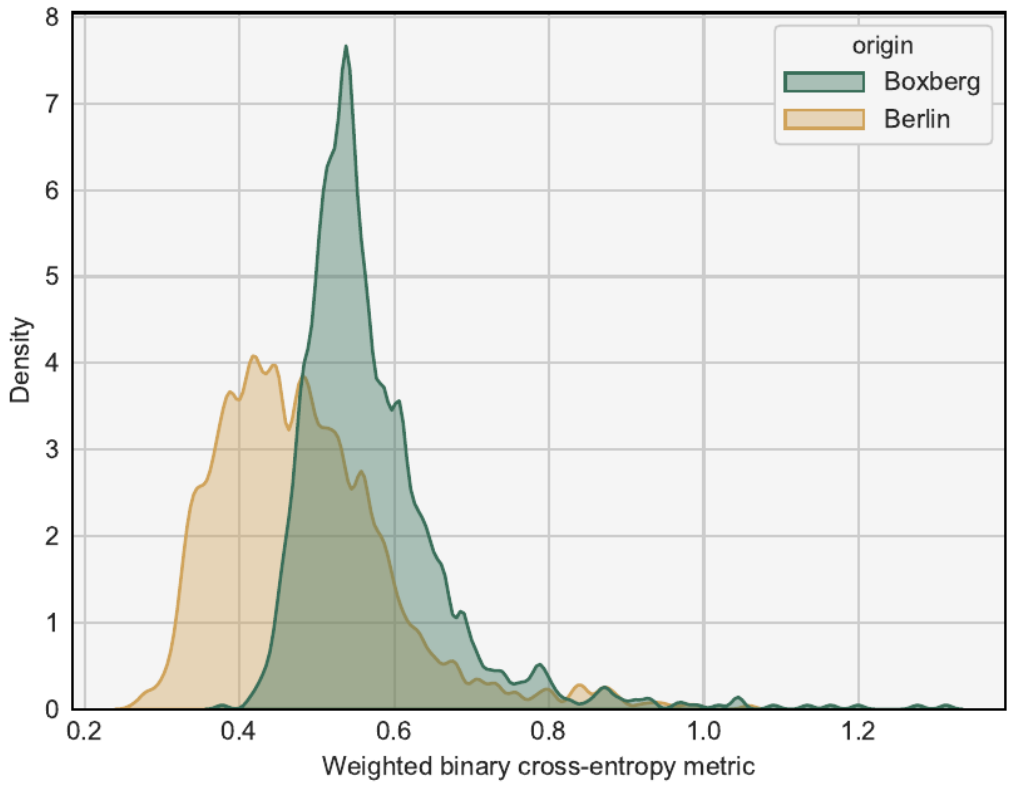

Figure 6: Histograms of the weighted-bce scores over all images in Berlin and Boxberg obtained with the neutral reference.

Large variations can be observed: the neutral wbce score for the Berlin images varies between 0.25 and 1. This means that the score associated with a segmentation is very dependent on the image considered: for one image, a score of 0.5 corresponds to a good segmentation and for another, such a score is equal to the neutral score and therefore equivalent to the absence of plume detection. Therefore, in the following, to make the segmentation scores more consistent over the samples, the wbce of an image is systematically divided by its wbce obtained with the neutral reference segmentation. This new metric is called nwbce (normalised weighted binary cross entropy metric) and a score of 1 means that the resulting segmentation is no better than "no detection".

3.3.1.2 A segmentation technique based on thresholding: ddeq

The second reference to be compared with our segmentation method is the detection algorithm implemented in the Python package for data-driven emission quantification (ddeq). This algorithm can be described as a thresholding method: it first detects signal enhancements that are significant in relation to instrument noise and background variability, and then identifies plumes as coherent structures. Since the algorithm returns a Boolean map, the identified non-plume and plume pixels (0 and 1) are mapped to two values, defined independently for each hotspot. These two values are chosen so as to minimise the wbce over all images from the hotspot. Figure 7 shows four applications of the ddeq algorithm to the CO2 images (two plume clusters centred in Boxberg, two in Berlin).

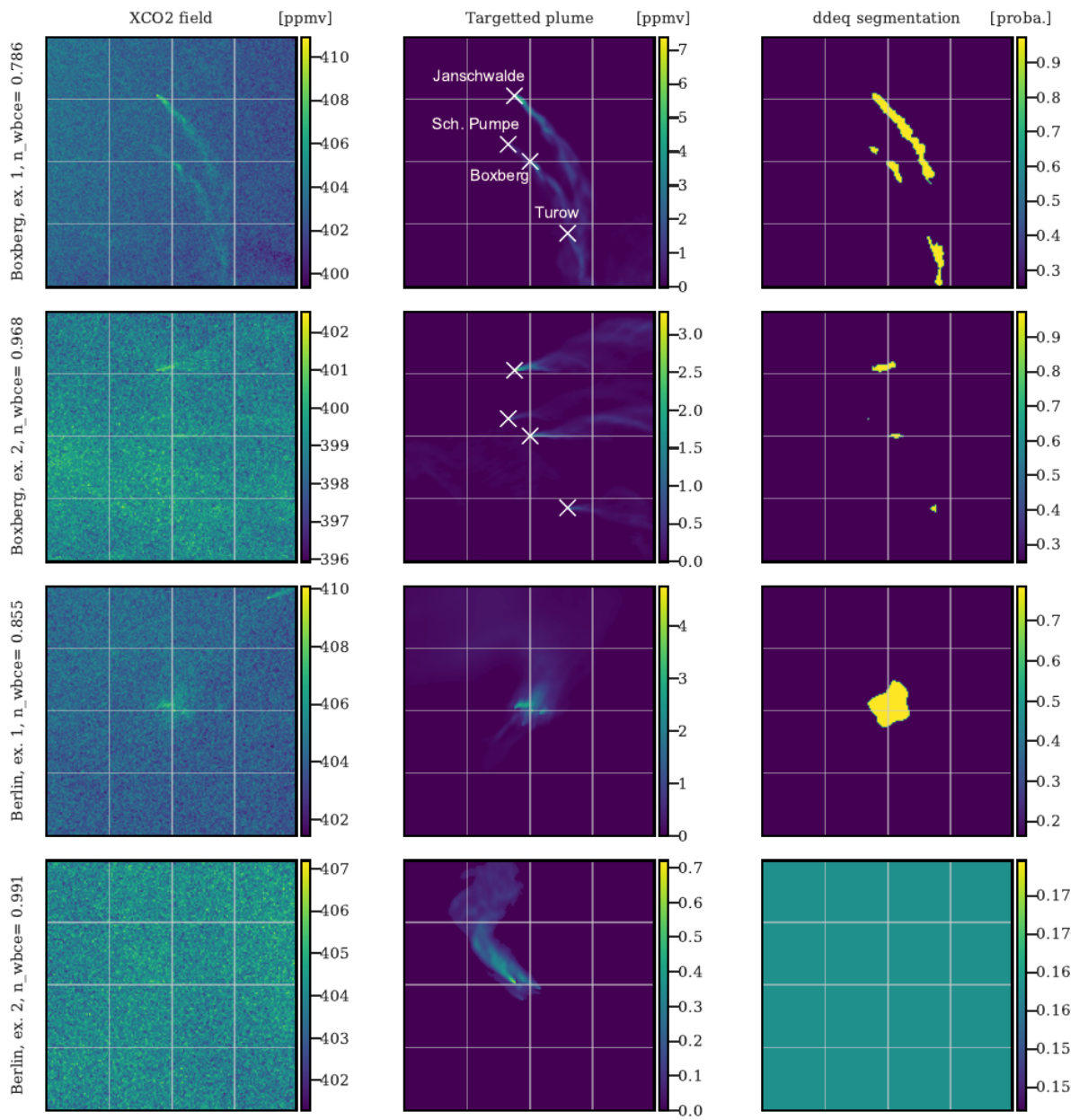

Figure 7: Four examples of ddeq plume segmentation algorithm applications on simulated satellite images centred at Boxberg (first two rows) or Berlin (last two rows). The first column corresponds to XCO2 simulated satellite images in ppmv, the second column to the targeted plumes, and the third column to the predictions of the ddeq segmentation algorithm mapped to probability images. All times given on the left of the figures are in UTC.

The first plume cluster (centred in Boxberg) image (1st row) obtains a much better nwbce (0.79) than the second example (0.97) because the plume signal to background ratio is much higher. The same is true for the Berlin plume segmentations. Furthermore, due to the low signal to noise ratio, no plume is detected on the fourth example and a constant probability map is returned (which gives a score close to the neutral and not 1, because the mapped probability is different). The thresholding method allows the segmentation of plumes, or portions of plumes, associated with signals above the background. But if no visible signal above the background is detected, the plume is not identified.

3.3.2 Generalisation on new data from the same region

3.3.2.1 Score histograms

Figure 8 presents the kernel density estimates of the nwbce scores of the U-net and ddeq segmentation methods according to the origin of the XCO2 field/plume pair.

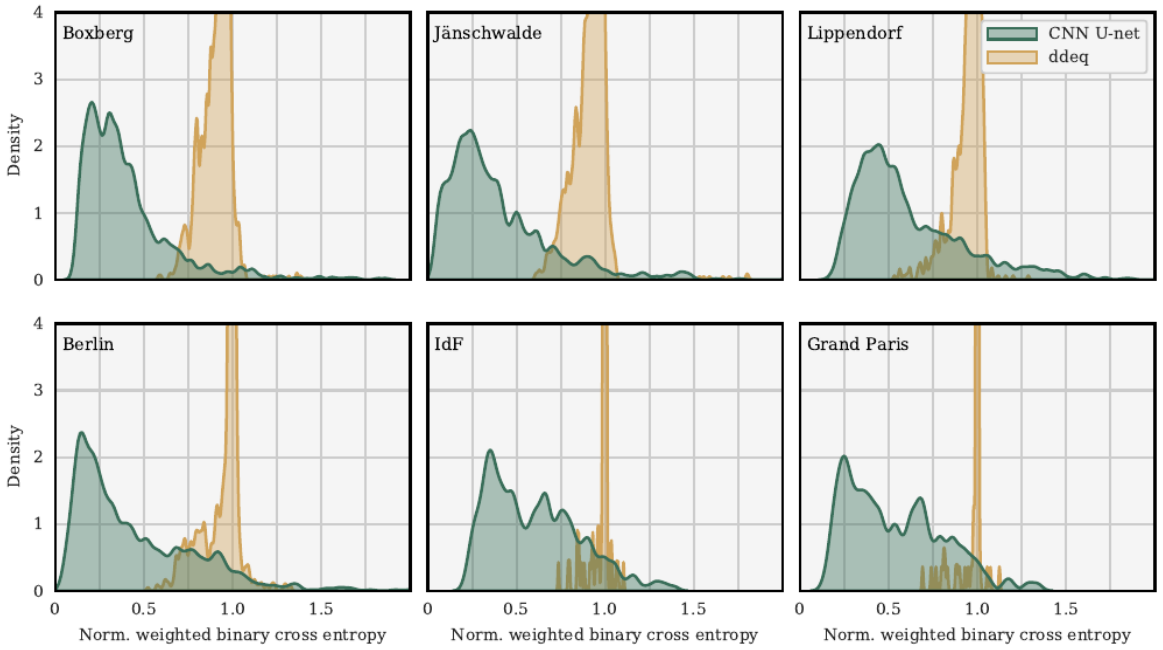

Figure 8: Kernel densities of the histograms of the nwbce scores of the U-net and ddeq segmentation techniques over all images of the various geographical domains.

As a general rule, the lower the score, the better the segmentation. As shown in the following examples, scores between 0 and 0.5 correspond usually to very good to good segmentation, and scores between 0.5 and 0.75 non-perfect but usable segmentation. A score of 1 is neutral (neither worse nor better than predicting no plume) and a score above 1 corresponds to a worse segmentation than the neutral , i.e. a segmentation of the wrong part of the image. A number of applications with scores are presented in the following.

On all hotspots, our deep learning model consistently outperforms the ddeq segmentation method on the nwbce metric. For example, the average nwbce over all Berlin images is 1.0 for the neutral (by definition), 0.95 for the ddeq method, and 0.44 for the CNN segmentation. The average nwbce is over all Janschwälde images 0.9 for the ddeq method and 0.4 for the CNN method. Note, however, that the CNN is optimised on the nwbce metric whereas the ddeq segmentation method is not. The choice of a metric is to some extent arbitrary and the difference between the two methods would change if another metric, and/or another definition of the plume, were chosen.

The best segmentation scores are obtained on Jänschwalde and Boxberg, which is consistent with the fact that these images contain several plumes of high intensity. The histogram of the Lippendorf nwbce metric shows overall very good results but with a large variance, and a significant part of the scores above 1. The distribution of the Berlin fields has a wider tail than that of the power plant fields: this can be explained by the shape of the city plumes, which are generally more complex and therefore more difficult to segment than the straight power plant plumes. The poorer results over Grand Paris and IdF on average are due to the smaller amount of available images and the low SNR of Grand Paris plumes. The small plumes specific to IdF (outside Grand Paris) are almost never recovered as shown in the IdF histogram, similar to the Grand Paris histogram but slightly shifted to the right.

3.3.2.2 Berlin predictions

In Fig. 9, we present four typical Berlin plume segmentations with the U-net algorithm. The XCO2 images (left column) are fed into the CNN which yields the segmentation probability maps (right column) of the XCO2 plumes. For the classical binary cross-entropy metric, a pixel with a probability equal to 0.5 means a "no-information" on the class of the pixel (plume or non-plume). We assume that this can be extended to the wbce metric: all pixels with a probability greater than 0.5 can be considered as part of a plume, while pixels below 0.5 can be considered as pixels that are not part of a plume. Consequently, a divergent colour map is used to represent the CNN model predictions.

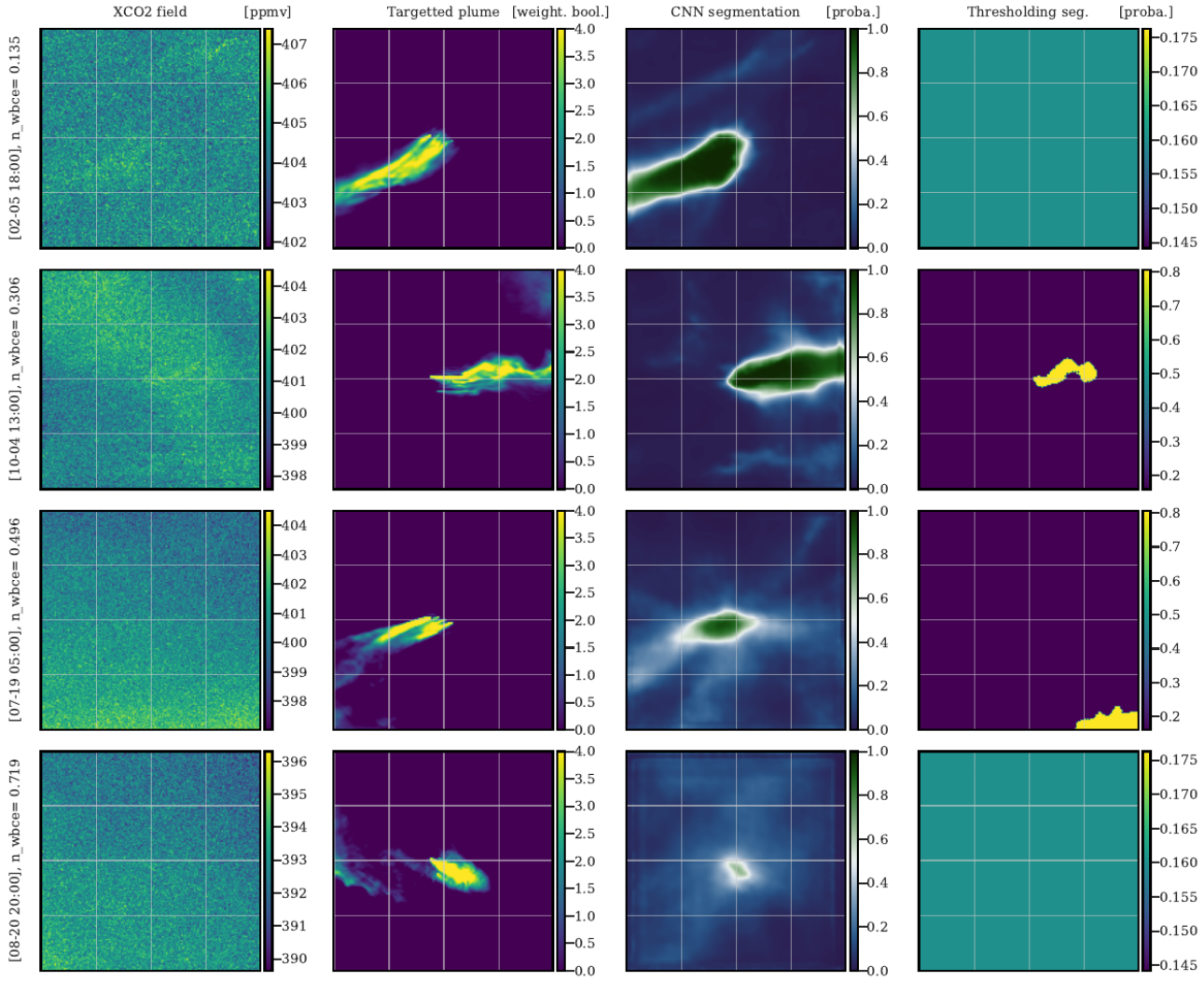

Figure 9: Examples of the application of U-net on images in the Berlin region. The first, second and third columns correspond to XCO2 images of Berlin, weighted Boolean plumes and CNN predictions as probability maps, respectively. The fourth column shows the application of the detection algorithm implemented in the Python package ddeq. The first, second, third and fourth rows are representative of the first, second, third and fourth quartiles of nwbce scores, respectively. All times given on the left of the figures are in UTC.

The first and second rows show a very accurate segmentation: the model predicts the correct direction, shape and thickness of the plume. The third plume is rather well recovered with some inaccuracies: in particular, the tail of the plume is reconstructed with less accuracy, which was expected since the concentrations on the tail reach very low values. Moreover, the core of the plume is segmented with less confidence: the probabilities of the plume pixels are close to 0.75. In general, prediction confidence is positively correlated with the nwbce score. Similarly, uncertainty, represented by the number of pixels close to 0.5, and the nwbce score are inversely correlated. To a certain extent, this is true for all

hotspots and is a measure of model uncertainty. It can also be used in evaluations without access to the truth to quantify how certain the predictions are. Confirming this correlation, the fourth row shows a very uncertain prediction that still correctly finds the direction and core of the plume. In all images, the position of the plume origin is always accurate. This is not trivial because in the training set horizontal and vertical shifts are used, which means that the plume origin is not known in advance by the model. We note that the plume is often masked by background variability and instrument noise, which does not prevent its detection by the CNN.

3.3.2.3 Plume cluster centred in Boxberg predictions

Figure 10 shows the segmentation of four images centred at Boxberg power plant. Sources are from bottom to top of the images: Turow, Boxberg, Schwarze Pumpe and Jänschwalde. The four images in order are representative of the four quartiles of their nwbce score, respectively.

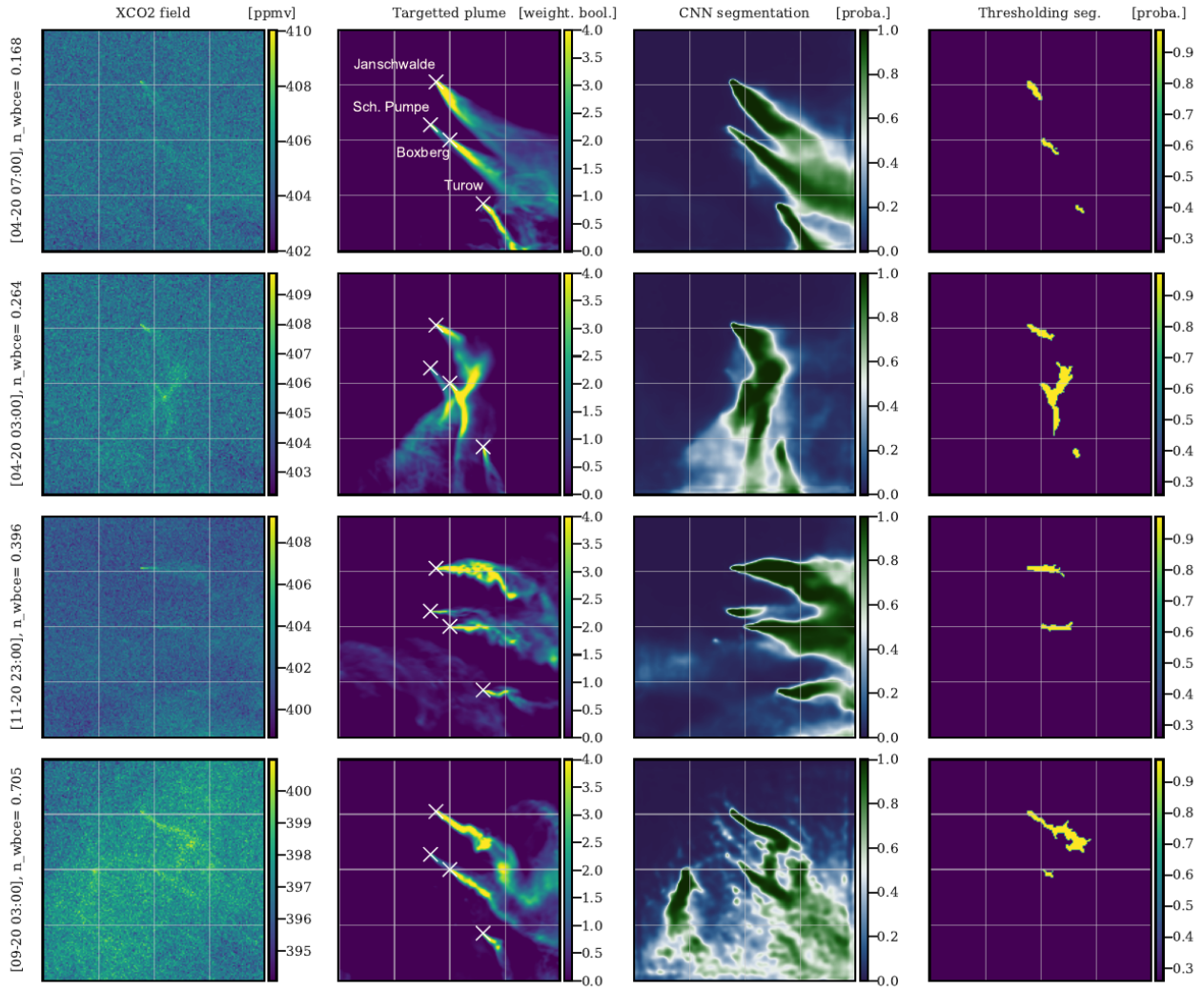

Figure 10: Examples of the application of U-net on images centred at Boxberg. Sources are from bottom to top in each image: Turow, Boxberg, Schwarze Pumpe and Jänschwalde. The first, second and third columns correspond to XCO2 images, weighted Boolean plumes and CNN predictions as probability maps, respectively. The fourth column shows the application of the detection algorithm implemented in the Python package ddeq. The first, second, third and fourth rows are representative of the first, second, third and fourth quartiles of nwbce scores, respectively. All times given on the left of the figures are in UTC.

All first three segmentations are very accurate: the origins, thicknesses and directions of the plumes are accurately reconstructed. Some failures are the mixing of the two plumes in the centre of the first image, the no detection of the Schwarze Pumpe plume in the second image, or the wrong evaluation of

the direction of the Turow plume in the third image. The fourth segmentation high nwbce is mainly due to the addition by the model with a high probability of a ghost plume on the left of the image: a clear enhancement on the XCO2 field at the same location explains the U-net error. The absence of power plants or major cities in the area raises questions on the origin of this enhancement.

3.3.3 Extrapolation on unseen data from another region

In this section, we investigate the performance of the U-net when trained and tested on data from different regions. Such a task is more difficult than generalising on plumes from the same region, where the training and test sets have more similarities from the local meteorological and pollution climatology. To study the potential for extrapolation, the U-net model is trained on the Paris, Jänschwalde, Boxberg and Lippendorf fields and tested on the Berlin fields. Berlin is chosen because cities are a particularly complicated case as their signal is lower and because, in this way, we can rely on a large set of images to validate and test the CNN model.

Figure 11 shows the histograms of the nwbce scores for all Berlin test images depending on the method used (left), and, with the CNN method in the case of geographical extrapolation, of all nwbce scores for several ranges of the Berlin emission rate at the time of the image (right).

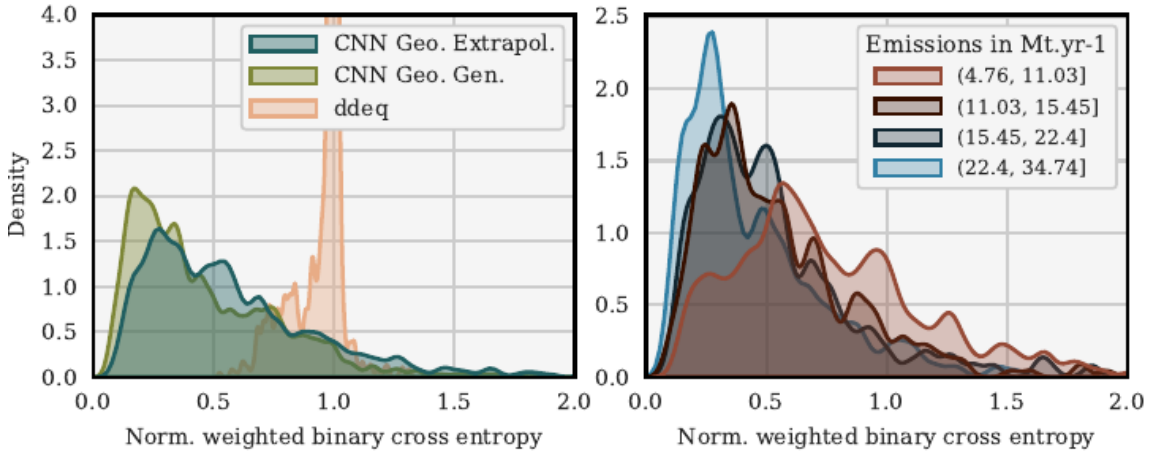

Figure 11: Histograms of the nwbce image scores over all test images of Berlin. Plumes/segmentations are classified depending on the method used (left). In the geographical extrapolation case, the CNN segmentation outputs are classified in four equivalent clusters according to the corresponding emissions (right).

The mean nwbce score of all prediction-truth pairs is 0.57 and the median is 0.49 – higher in both cases than when the model is trained on Berlin images only (see Section 4.2), but still very satisfying: the model extrapolates well and outperforms the ddeq segmentations according to the nwbce metric. In addition, the main divergence between the generalisation and extrapolation histograms is a shift to the right of the part of the histogram between the scores of 0 and 0.75. It is explained later in this section that segmentations with an nwbce below 0.75 are generally good enough for inversion. In other words, the switch from generalisation to extrapolation mainly degrades highly accurate segmentations to "only" accurate segmentations.

On the right histogram, it can be observed that the results, quite naturally, deteriorate in the case of low-emission plumes: for high-emission plumes, the density peaks at 0.25 whereas it peaks at 0.5 in the case of low-emission plumes. The variance of the low-emission plumes nwbce metric density is also significantly higher.

In Fig. 12, we present four typical Berlin plume segmentations with the U-net algorithm. The four images from top to bottom are illustrative of the four quartiles, respectively (according to their nwbce score).

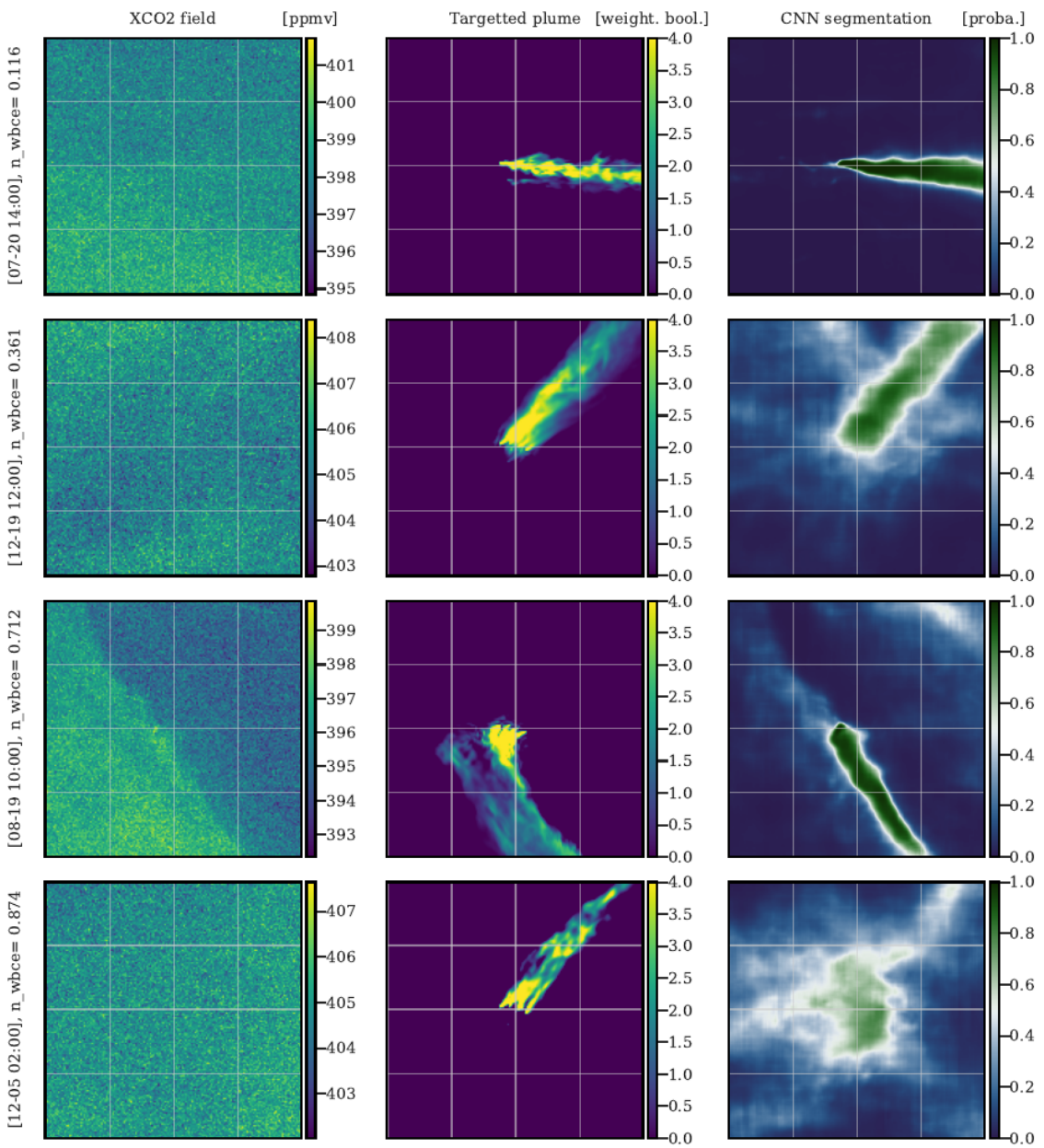

Figure 12: Examples of the application of the U-net on images of Berlin. The first, second and third columns correspond to XCO2 images of Berlin, weighted Boolean plumes and CNN predictions as probability maps, respectively. The first, second, third and fourth rows are representative of the first, second, third and fourth quartiles, respectively. All times given on the left of the figures are in UTC.

The first three images show segmentations that recover the direction and origin of the plume. The thickness of the plume is also well reconstructed in the first two examples but a part of the plume is missed in the third example, which gives the largest fraction of the error; this miss is probably due to a gradient in the background field. The second and third examples also show a significant number of pixels at values around 0.5, expressing the uncertainty of the model. In these examples, the plume is masked by background variability and instrument noise, yet is detected by the CNN. In the fourth example, the deep learning model fails to detect the plume and yet diagnoses higher uncertainty. Further analysis shows that most of the last quartile results are difficult to use for inversion because they partially or completely miss the plume, or express too much uncertainty.

3.4 Inversion

The results for the inversion with the CNN are exploratory at this stage. The experiments yield positive results and demonstrate the relevance of the general approach for inversions. Furthermore, they provide clear perspectives for further developments. In particular, a specific approach is needed to improve the training set to ensure good results when the training set does not include simulations from the source targeted during the testing phase.
This section mainly demonstrates that we are able to use deep neural networks to perform inversion tasks. The analysis suggests that significantly better results could be obtained in the future.

3.4.1 Inversion theory

The purpose of this section is to describe a methodology for inversing XCO2 satellite images, based on the use of the segmentation model.

The inversion of XCO2 images is the estimation of CO2 emissions corresponding to a plume detected on a satellite XCO2 image. It is a regression task which may be tackled by a statistical or machine learning statistical model taking as inputs the XCO2 image with a XCO2 plume and yielding as output one or several scalars representing the CO2 emission rates.
Some ancillary data (predictors) can be added to the XCO2 image to help the regression achieve better precision. In particular, NO2 images, results of the segmentation algorithm or winds can be used to help the model to estimate the emissions associated with a XCO2 plume on an image.

An inversion model is defined here as a statistical model that uses a certain amount of information as input and outputs an emission rate. Inputs can be images at a time t of NO2, XCO2, outputs of the segmentation model, wind direction and velocity. The images are centred around a point source whose emissions are to be reconstructed. The point source is always positioned in the centre of the image so that the model knows its position. The output is a vector representing the evolution of the emission rate over the hours preceding the input images. If this vector is of size 1, we seek to reconstruct a single emission rate.

3.4.2 Inversion results with a linear convolutional neural network

The inversion model is trained and evaluated with:
- as inputs, the NO2 and XCO2 fields, the U and V wind components as column averaged fields, and the output of the segmentation model. The plume field should be sufficient to predict the plume emissions. Therefore, we expect the model to predict emissions from a combination of at least the XCO2 field and the segmentation model output;
- as output, a single emission rate.
The segmentation model was fed with NO2, XCO2, U and V wind images.
We built a dataset consisting of 64*64 pixel images of size 2 km centred in Berlin, Jänschwalde, Dolna Odra, Turow, Patnow, and Lippendorf. We initially considered a model trained and evaluated on data centred at the same point source, and such an inversion model produced excellent results. But a key challenge is to train a model capable of extrapolating to new points source (not used during training).

We therefore consider in a second step a case of geographical extrapolation: we train a model on data from a certain sample of point sources and evaluate the model on a point source not included in this sample. For example, we train a model on data centred on Berlin, Dolna Odra, Patnow, and Opole and evaluate the model on Lippendorf. Tests with the convolutional neural network EfficientNetB0 and a linear convolutional neural network have been performed.

The linear convolutional neural network consists of a simple 42,000-parameter network with no

activation layers, making it by construction linear. Here we present the performance of this model trained on images centred at Opole, Berlin, Dolna Odra, and Patnow and evaluated on images centred at Lippendorf. The model is trained on NO2, XCO2 and segmentation images over 2000 epochs. The model does not overfit and the last iteration of the model is used. The average absolute error obtained on all pairs of evaluation images/emission rates is 3.9 Mt/year and the average relative error obtained is 24%.
On Fig. 13, we provide probability densities of the mean absolute error, mean relative error, true and predicted emission rates, and predicted emission rates divided by the true emission rates.

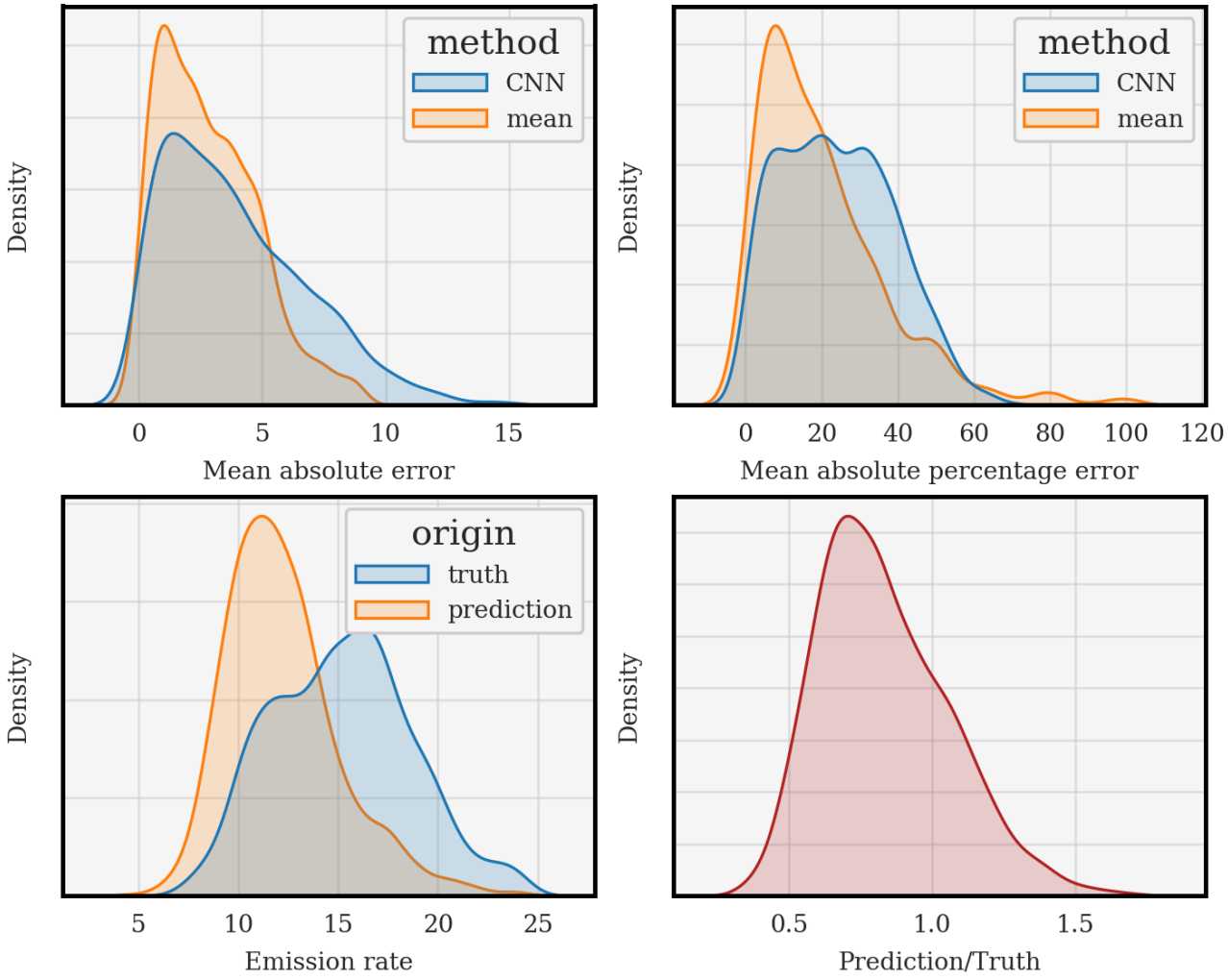

Figure 12a: Probability densities assessing the performance of the linear convolutional neural network trained on images of Opole, Berlin, Dolna Odra, Patnow and tested on Lippendorf. The model that predicts the mean of the emission rates of Lippendorf is provided as a strong baseline. Top-left: kernel densities (histogram fitting) of the mean absolute error; top-right: kernel densities of the mean relative error; bottom-left: kernel densities of the emission rates; bottom-right: kernel density of the prediction divided by the truth.

The probability densities of the absolute and relative errors show that the CNN predictions are on average close to the actual emission rates. The underestimation of the predictions compared to the truth is probably due to the insufficient number of input power plants to learn from.

3.5 Conclusions

Emissions can be assessed from CO2 plumes of hot spots in the satellite images. This data-driven assessment needs to detect plumes from satellite images, which is difficult for a thresholding method due to the low signal-to-noise ratio of the plume. Deep learning and convolutional neural network (CNN) techniques could provide more accurate plume detection, because of their ability to learn and capture plume-specific spatial patterns, which do not necessarily depend on a significant concentration enhancement.

In this deliverable, we first evaluate the ability of CNNs to accurately detect the mask of a plume in a gap-free XCO2 satellite image using simulated CO2 fields. Each synthetic XCO2 image is the sum of the anthropogenic plume of a major hotspot (a city or a power plant), a background from other biogenic and anthropogenic fluxes, and a random Gaussian noise to simulate the satellite instrumental errors.

Our plume detection model is based on a CNN encoder-decoder, the U-net algorithm with an EfficientNetB0 backbone. It is an image-to-image model, which transforms the full XCO2 field into a map showing the positions of the anthropogenic emission plumes. For training, we develop a novel loss function that penalises more the errors made on pixels associated with high plume concentrations and thus yields a more accurate definition of a plume than a simple threshold value. This CNN is trained and tested in two contexts. First, the capacity of the model to generalise on unseen data from the same region is evaluated. The U-net shows very good performance: most plumes are precisely segmented, the origin, thickness and shape of the plume being often accurately retrieved. Second, we evaluate the ability of the model to extrapolate to unobserved data from another region. Specifically, the model is trained with simulated fields of Paris and power plants and tested with fields in the Berlin area. The segmentations are slightly less accurate than in the first context, but are nevertheless very satisfactory: about half of the Berlin plumes are accurately segmented, with plume shape, thickness, direction and origin recovered, and 75% of the segmentations are accurate enough to be used for inversion.

The observed good performance of the U-net architecture is due to the ability of the convolution layers to capture detailed spatial patterns corresponding to plumes even when the concentrations of these plumes are partially covered by high satellite noise or background variability. It allows the model to outperform segmentations by the thresholding technique according to the concentration-weighted metric, whether the model trained on some data is tested on data from the same region or not. The U-net is effective over a wide variety of plumes (cities, power plants, diverse regions, several levels of hotspot emissions). Its training time is less than one day, while once the model is trained, the evaluation of a new image is achieved in less than a second. However, although the model performs better when trained and tested on data from the same region, it would be too expensive to generate simulations on all the cities and power plants whose plumes we wish to segment. Therefore, we believe that the goal is the development of a "universal" CNN, trained only on a limited sample of cities and power plants and highly efficient on all of them. The model in this paper, trained on Paris and power plants data and tested on Berlin, already shows accurate and very satisfactory segmentations of the Berlin plumes, but the results need to be confirmed on multiple cases, in particular with changing orography within the scene, and to be extended to incomplete images.

It is very likely that many other techniques could be applied to improve these segmentations, which could be based on:

- more advanced and powerful NN architectures such as transformers or on CNN networks with more parameters;
- improving the distribution of the data, by increasing the amount of images used, or by using more carefully chosen augmentation techniques.

For all these reasons, CNN methods appear to be very suitable for CO2 plume segmentation problems on satellite data. However, the model was evaluated on simulated data which does not take into account all the problems of plume detectability presented by real satellite images in particular clouds and patterns of systematic errors due to surface reflectance and aerosol dependency of the retrievals. Consequently, the method needs to be extended and validated on full OSSEs: fields with clouds and satellite swaths taken into account, and afterwards on real satellite data.
The CNN can be adapted to handle images with gaps (e.g. clouds) as inputs. A first idea is to reconstruct the images (fill the gaps) using inpainting techniques. A second, and intuitively more compelling, idea is to train a gap-ignoring CNN (using partial convolutions or simply filling the gaps with negative values).

As the CO2 emission rate is proportional to the mass and shape of the corresponding plume, an accurate segmentation of the plume should lead to an accurate emission estimate.
The linear convolution model is trained and tested in the cases of generalisation or geographical extrapolation. In both cases, good performance is obtained when evaluated on the test data set. This demonstrates the possibility of creating an efficient universal hotspot estimator from a CNN, fed by these segmented plumes and a sufficient data set.

The development of a better dataset is essential to improve the current results. This includes making better use of the segmentations obtained from the CNN plume detection model. Another perspective would be to identify the individual contributions of the segmentation and inversion models to the inversion performance (recall that the inversion model is based on the segmentation model). These could be identified by comparison with the methods of T4.2.

4 Analyses of the spatial and sectoral resolution and of the added value of co-emitted species with a Carbon Cycle Fossil Fuel Data Assimilation System

This section describes a series of experiments using a regional version of the Carbon Cycle Fossil Fuel Data Assimilation System (CCFFDAS, Kaminski et al., 2022) that apply the Quantitative Network Design (QND) approach. As the system is described in the published literature we focus on the aspects that are particularly relevant for this study and refer for more detail to Kaminski et al. (2022a).

4.1 Method

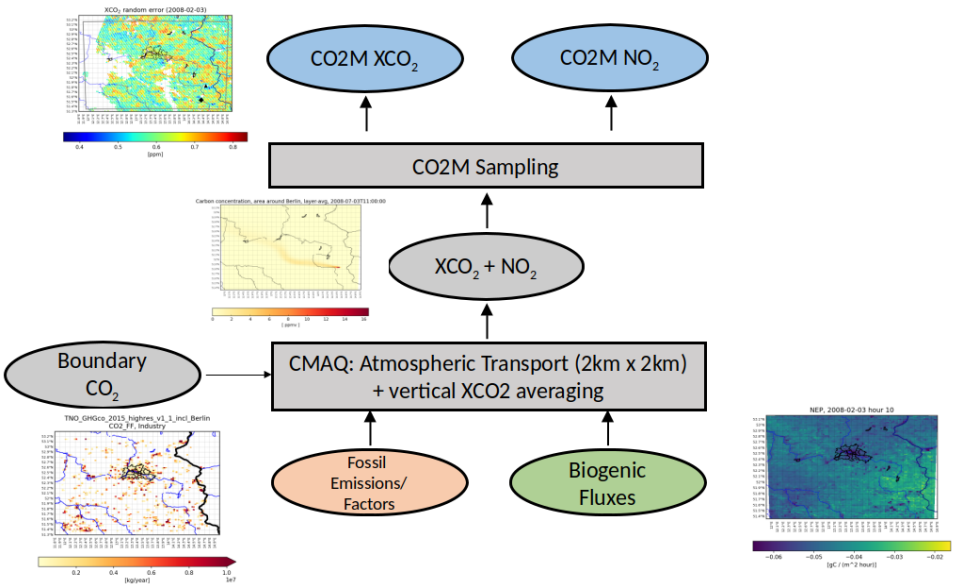

Figure 13: The flow of information in the forward sense through the modelling chain.

The CCFFDAS is built around a modelling chain to simulate two CO2M XCO2 images over the Berlin area, one on 3 February 2008 and one on 3 July 2008, starting 24 h before the respective acquisitions. The flow of information in the forward sense is shown in Figure 13. The CO2M observation impact is assessed through the quantitative network design (QND) approach that is described below. This approach is based on a representation of the modelling chain through a Jacobian matrix that quantifies the sensitivity of the measurements as a function of the control vector. Our control vector consists of the surface emissions into each grid cell and the lateral inflow of CO2 as well as scaling factors of the NO2/CO2 emission ratio. Our 24 h simulation period is sufficiently long to ensure that the initial concentration has left our 200 × 200 km2 domain under typical wind conditions, i.e., we can safely ignore it in the control vector.

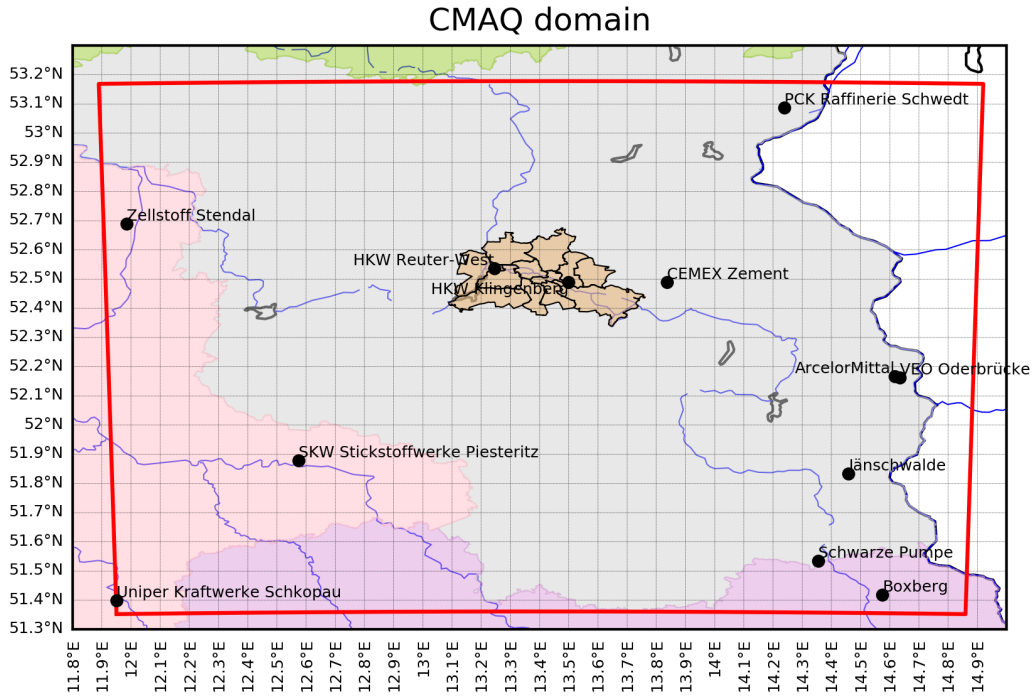

Figure 14: Model domain including the locations of the largest emitting power plants..

To link the emission models to XCO2 and NO2 observations by CO2M we use version 5.2.1 of the Community Multiscale Air Quality model (Zenodo, 2018), which can be run as an offline tracer transport model with meteorological input fields derived from the Weather and Research Forecast model (WRF, version 3.9.1.1, (Skamarock et al. 2008). Our model domain of 200 km by 200 km around Berlin is shown in Figure 14. Over this domain we operate the model at a horizontal resolution of 2 km by 2 km with 32 vertical layers.

For the time scales addressed in the present study, CO2 is considered chemically inert. By contrast, the loss of NO2 due to chemical processes needs to be taken into account. The model achieves this through an atmospheric lifetime. For the summer case we use a lifetime of 4 h and for the winter case we regard also NO2 as inert.

Name	fuel type	CO2 emission [MtC/yr]	NO2/CO2 emission ratio
LEAG, Kraftwerk Jänschwalde	solid	6.48	0.0029
Kraftwerk Boxberg	solid	5.33	0.0025
LEAG, Kraftwerk Schwarze Pumpe	solid	3.36	0.0017
Uniper Kraftwerke GmbH (Schkopau)	solid	1.46	0.0024
PCK Raffinerie GmbH Schwedt	liquid	0.87	0.0025
ArcelorMittal Eisenhüttenstadt GmbH	gaseous	0.72	0.0018
Vattenfall Europe Wärme HKW Reuter-West	solid	0.59	0.0030
SKW Stickstoffwerke Piesteritz GmbH	solid	0.59	0.0008
VEO Vulkan-Energiewirtschaft - Oderbrücke GmbH	gaseous	0.46	0.0001
CEMEX Zement GmbH	solid	0.39	0.0036

Table 1: Fuel type, annual CO2 emissions, and NO2/XCO2 emission ratio for the 10 power plants in the domain with highest CO2 emissions.

For the electricity generation sector we take the locations of power plants, temporal emission profiles, as well as annual emissions of CO2 and NOx from a data set (Super et al. 2020) compiled as part of the European commission's CHE project (Balsamo et al. 2021). Table 1 shows the relevant characteristics of the 10 largest power plants in the domain with highest CO2 emissions. The control vector for the electricity generation sector consists of the CO2 emission of each power plant and of a scaling factor for the NO2 to CO2 emission ratio, which also absorbs uncertainties in the NO2 to NOx ratio. Prior uncertainties of the CO2 emission are assumed to be a constant fraction of 20% of the emission. With regard to the degree of differentiation of the scaling factor of the NO2/XCO2 emission ratio we explore three cases:

- Uniform: All plants in our domain share the same scaling factor.
- Per type: All plants in our domain of the same fuel type share the same scaling factor.
- Per plant: Each plant in our domain has an individual scaling factor.

fuel type	σ
solid	0.098
liquid	0.245
gaseous	0.924
average	0.422

Table 2: 1 sigma uncertainty range in the NO2 to CO2 emission ratio for each fuel type.

The prior value for the scaling factor is 1. The (relative) prior uncertainty of the emission factor ratio is calculated from reported emission factor uncertainties averaged for several countries, following the approach used by (Super et al. 2020), see Table 2.

Our “other” sector accounts for fossil fuel emissions from all sectors except electricity generation. Other sector emissions of CO2 and NOx are also taken from the data set by (Super et al. 2020). As the "other" sector includes several processes, the resulting emission ratio varies spatially depending on the contributions of different processes. The control vector for the other sector consists of the CO2 emission into each model grid cell and of a scaling factor for the NO2 to CO2 emission ratio. In our experiments, we use a (spatially uncorrelated) prior uncertainty of 52.8% of the emissions into a grid cell for each grid cell, which translates to a 20% prior uncertainty when aggregated over Berlin. For the scaling factor of the NO2/CO2 emission ratio we use a prior of 1 and the prior uncertainty for the average over fuel types.

The terrestrial biosphere model we used to calculate the natural terrestrial CO2 exchange fluxes is based on the Simple Diagnostic Biosphere Model (SDBM, (Knorr et Heimann 1995), which was used by (Thomas Kaminski, Scholze, Rayner, Voßbeck, et al. 2022) for assimilation of CO2 and by (T. Kaminski et al. 2017) for assimilation of XCO2. Here we use a new implementation on the 2 km by (Thomas Kaminski, Scholze, Rayner, Houweling, et al. 2022) 2 km grid of the transport model with a time step of 1 hour. It calculates the uptake of CO2 by photosynthesis (expressed as Gross Primary Productivity, GPP) using a light-use efficiency approach and ecosystem respiration using a temperature dependency. The control vector consists of 5 parameters that were calibrated against a set of eddy-covariance sites. We use a prior uncertainty range of 20% for each of the parameters.

The Quantitative Network Design (QND) formalism is presented in detail by (Thomas Kaminski et Rayner 2017), see also (Tarantola 2005) and (Rayner, Michalak, et Chevallier 2019). In brief, it performs a rigorous uncertainty propagation from the observations to a target quantity of interest relying on the indirect link from the observations to the target variables established by a numerical model. The link has to be indirect, because, in general there is no direct link from the observations to the target quantity. There are, however, direct links from the control vector, which includes the uncertain inputs to the modelling chain, to target quantities and observations. For our experiments, the control vector is composed of the fossil fuel emissions from power plants, the fossil fuel emissions from the other sector, scaling factors for the NO2/CO2 emission ratio, and the parameters of the terrestrial biosphere model.

The target quantities are fossil fuel emissions for each power plant and from the other sector on the 2 × 2 km2 pixel scale and aggregated to larger scales, including the scale of Berlin districts and of the entire city. The observational impact on the target quantities is quantified by the following two-step procedure: The first step uses the observational information to reduce the uncertainty in the control vector, i.e., from a prior to a posterior state of information. The second step propagates the posterior uncertainty forward to the simulated target quantity.

Each CO2M satellite has a repeat cycle of 11 days. We apply our QND framework around the CCFFDAS to the simulated uncertainty in single measurement retrievals of CO2M for two specific overpasses with low cloud cover over our study domain around Berlin, one in winter and one in summer. Random and systematic errors were estimated from an error parameterisation formula that exploits the availability of aerosol information from a multiangular polarimeter onboard each CO2M satellite. We computed the variance of the XCO2 observational uncertainty as the sum of the variances of random and systematic errors. This reflects the assumption that systematic and random errors are independent of each other and also in space.

For the retrieval uncertainty in tropospheric column NO2 we follow the estimates of (Lorente et al. 2019) for the TROPOMI NO2 retrieval. For the random uncertainty we use their estimate of the slant column error for the TROPOMI retrieval, which is 0.5–0.6 · 1015 molec.cm-2.. We use a systematic error of 0.3· 1015 molec.cm-2. We regard random and systematic error as independent and compute the variance of the NO2 observational uncertainty as the sum of the variances of random and systematic errors. We note that one of the activities in CoCO2 WP5 is to investigate the sensitivity of the posterior uncertainties with respect to the uncertainty in NO2.

The inflow from the lateral boundary is represented in the control vector by one component for each group of 5 grid cells in the horizontal at each of the vertical levels with prior uncertainty of 0.53 g(C)/day.

4.2 Results

To investigate the impact of the degree of differentiation of the scaling factor of the NO2/XCO2 emission ratio we explore three cases:

- Uniform: All plants in our domain share the same scaling factor.
- Per type: All plants in our domain of the same fuel type share the same scaling factor.
- Per plant: Each plant in our domain has an individual scaling factor.

and contrast these to a case with XCO2 only. All experiments were conducted for the summer and winter days.

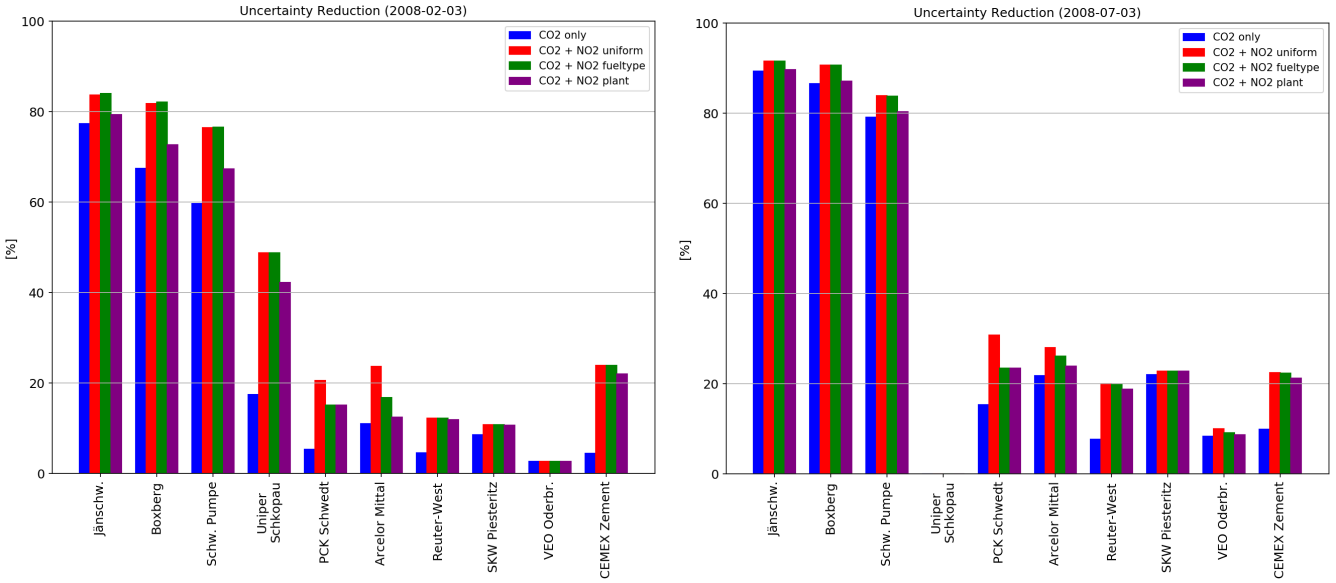

Figure 15: Uncertainty reduction for ten largest power plants and all experiments in winter (left) and summer (right).

Figure 15 displays the uncertainty reductions for the largest power plants in the domain for the winter day (left) and the summer day (right). We note that these results are derived by an approach completely different from that used in Section 3. Here we are exploiting a combined transport and process model rather than using a statistical relation (Neural Network). This renders a quantitative comparison difficult. We first compare the cases without NO2 (blue bars) and with NO2 and uniform scaling factor (red bars), which generally shows the positive impact of complementary NO2 observations. They increase the uncertainty reduction for the large power plants in the summer and winter cases. In relative terms the increase in uncertainty reduction is particularly high for the plants which were not well observed with XCO2 alone and those with high NO2/CO2 emission ratio (compare Table 2). The absolute reduction in posterior uncertainty through the addition of the NO2 measurements is, however, highest for the larger power plants. In relative terms the increase in uncertainty reduction is higher in the winter case, when the constraint by XCO2 alone is weaker. The best overall performance of XCO2 and NO2 is, however, achieved in the summer case, when both random and systematic errors are lower.

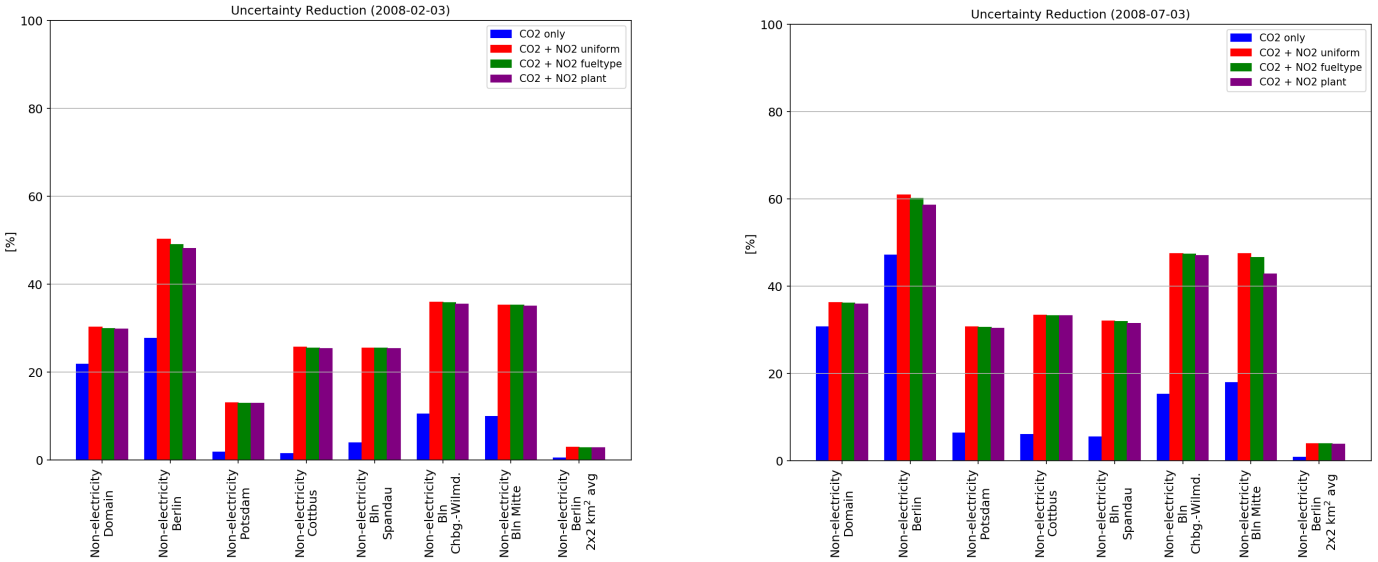

Figure 16: Uncertainty reduction for the other sector (non electricity) at spatial scales from entire domain to city (Berlin, Potsdam, and Cottbus) to Berlin district to grid cell in winter (left) and summer (right).

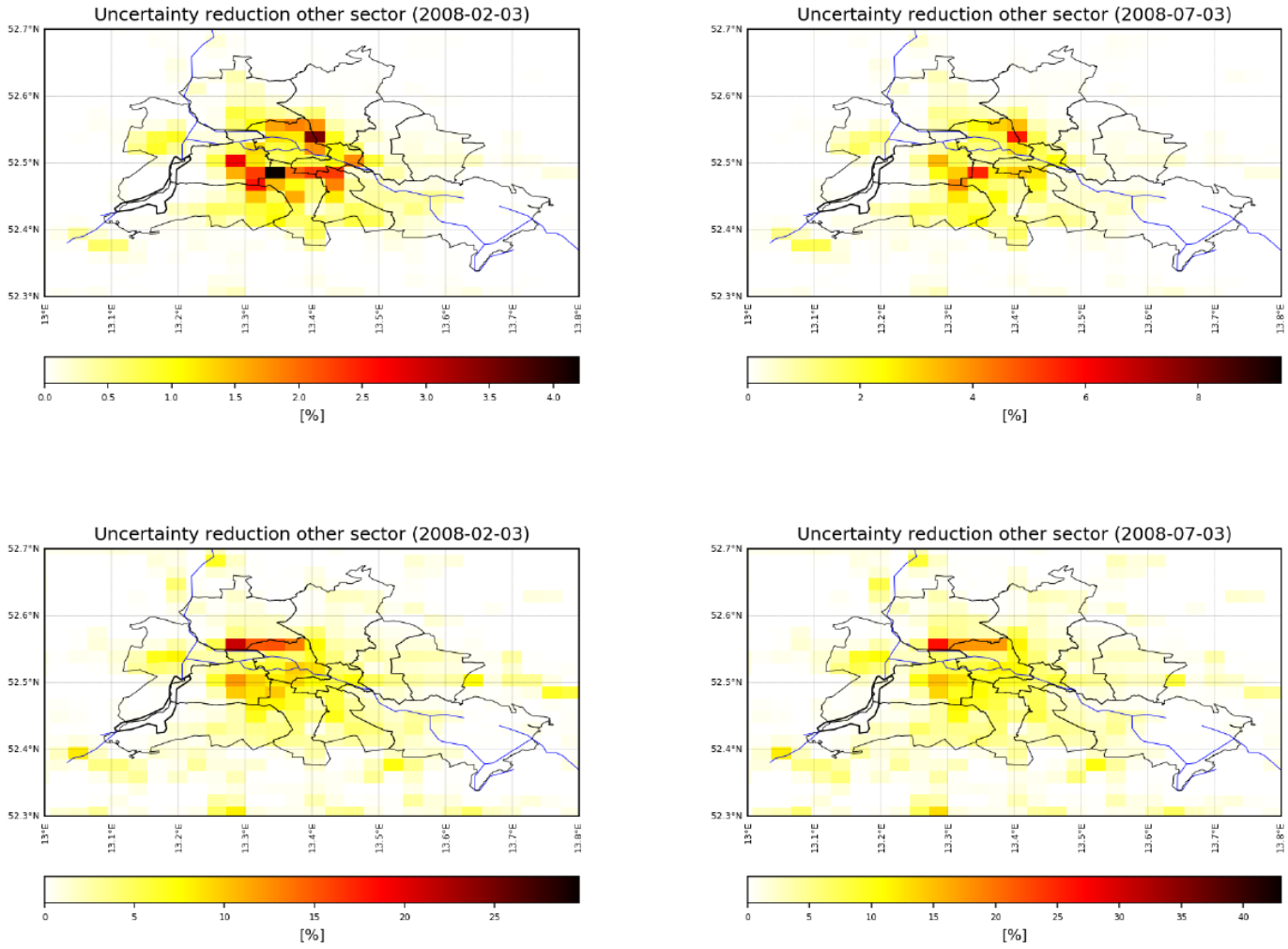

Figure 17: Uncertainty reduction for other sector emissions at 2 km by 2 km scale over Berlin in winter(left) and summer (right) with (top) without NO2 (bottom)

For the other sector, the extra NO2 measurements significantly increase the uncertainty reduction on all scales on the summer and winter days (Figure 16). On the grid cell scale uncertainty reductions reach now 25% in the winter case and 40% in the summer case (bottom panels of Figure 17). High values reflect the combination of high emission ratio with high CO2 emissions. On the Berlin district scale the relative increase in uncertainty reduction is particularly high for districts that were not well constrained by XCO2 alone such as Spandau. In absolute terms the posterior uncertainty decreases most for districts with larger emissions such as Charlottenburg-Wilmersdorf. With the extra NO2 measurements the uncertainty reduction in the other sector emissions aggregated over Berlin increases to about 50% in the winter case and 60% in the summer case. The best overall performance of XCO2 and NO2 for the other sector is achieved in the summer case.

The differentiation of the scaling factors in the NO2/CO2 emission ratio has an impact on the uncertainty reduction. As expected, a uniform scaling factor yields higher uncertainty reduction than a scaling factor per plant. This is because the uniform scaling factor is constrained by the atmospheric observations of all plants (transfer of information between plants), while an independent plant-specific scaling factor is only constrained by the atmospheric observations of the plant in question. In other words, the case of the uniform scaling factor imposes more prior knowledge as it removes the independence of the scaling factors. In between these two cases lies the case with a scaling factor per fuel type, with the exception of the largest of the ten power plants where it outperforms the case with uniform scaling factor. The four largest plants belong to the type burning solid fuel, for which there are two competing effects when changing from the case “uniform” to the case “fuel type”. First, the prior uncertainty in the emission ratio for solid fuel is considerably lower than the average we use in the uniform case (see Table 2), which increases the performance of the NO2 measurements. Second, the transfer of information from one power plant to the next through the use of the same scaling factor for the emission rate is obviously weaker in the case “fuel type” than in the case “uniform”, because the information is shared between fewer power plants. For the larger plants the first effect dominates the second. In the winter case this concerns the first four power plants and in the summer case the first two.

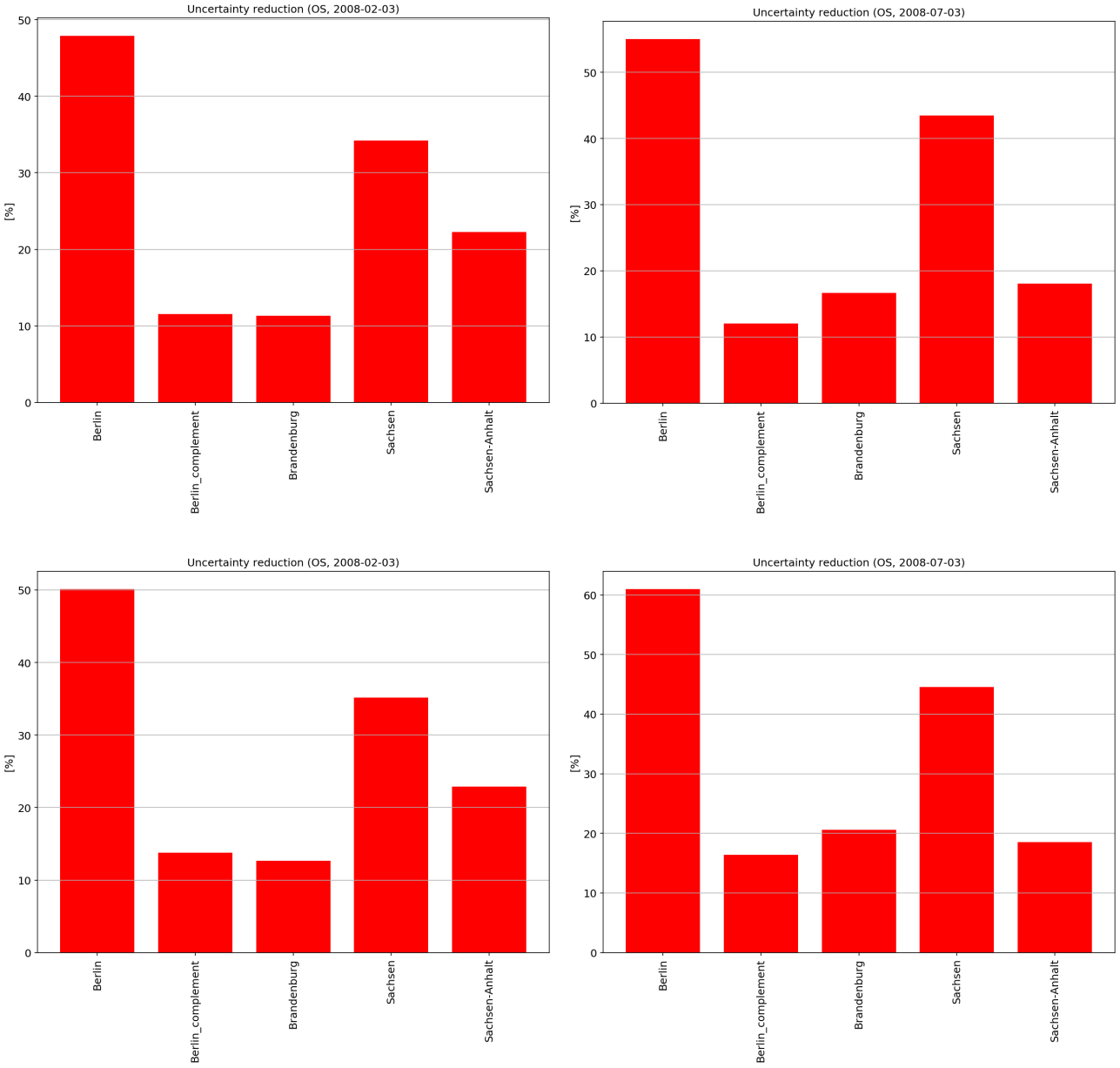

Figure 18: Uncertainty reduction for Berlin, its complement, Brandenburg and the intersection of two Bundesländer with the domain for winter (left) and summer (right) cases when small scale structure of lateral inflow is (top) or is not (bottom) assumed unknown

To explore the impact of the representation of the lateral boundary we investigated two cases:

- The first case treats the small scale structure of the boundary as unknown in that it includes a component of the control vector for each group of 5 grid cells in the horizontal at each of the vertical levels with a prior uncertainty of 12gC/s.

- The second case treats the small scale structure of the boundary as known and only solves for the large scale structure in that it includes a single scalar multiplier for the inflow from the lateral boundary into the control vector with a prior uncertainty set to the sum of the prior uncertainties into the individual grid cell groups defined above.

Again, we explored the winter (wind from southwest) and summer cases (wind from east). In both seasons and for all regions the uncertainty reduction for the other sector is higher when small scale structure of the lateral inflow is treated as known (Figure 18). For example, for Berlin and summer it increases from ~55% to ~61% when small scale structure in lateral boundaries is assumed to be known. For the complement of Berlin, which is closer to the boundaries, it increases by about 1/3 in summer (from 11.97% to 16.31%).

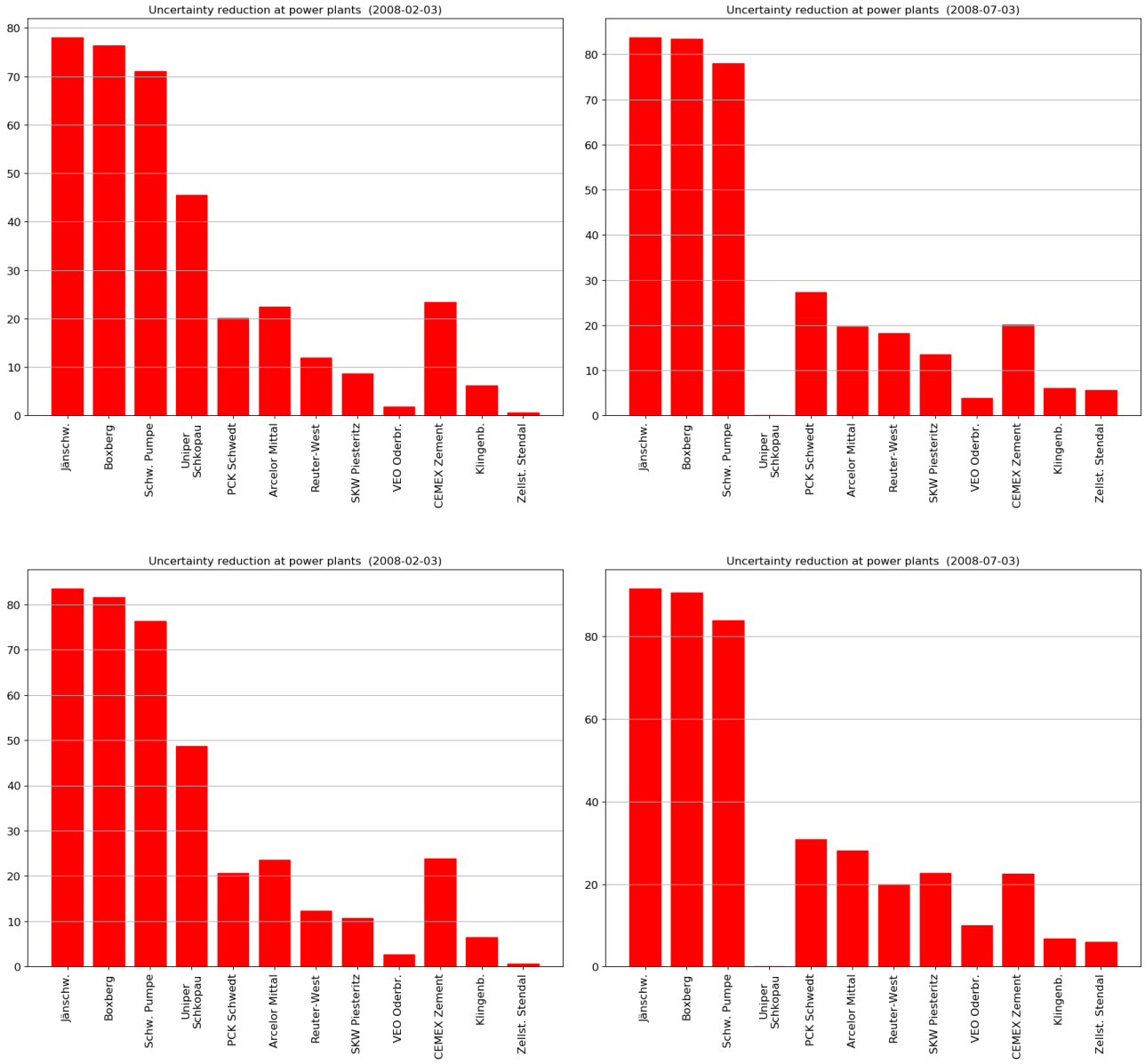

Figure 19: Uncertainty reduction for the 12 largest emitting power plants in the domain for winter (left) and summer (right) cases when small scale structure of lateral inflow is (top) or is not (bottom) assumed unknown.

In both seasons and for all power plants the uncertainty reduction is higher when small scale structure of the lateral inflow is treated as known (Figure 19). The extent of this increase depends on the location of the plant in combination with the wind direction. For example, for the plant "Arcelor Mittal", which is located close to the eastern boundary of the domain, the increase on the summer day with easterly wind direction is almost as high as 1/2 (from 19.6% to 28.1%).

4.3 Summary and Conclusions

The study conducted a number of assessments of observation impact in terms of the posterior uncertainty in fossil fuel emissions over 24 h on scales ranging from 2 to 200 km. This means the assessments include temporal and spatial scales typically not covered by inventories. Our typical metric is the relative reduction of the uncertainty in emissions through the information in the CO2M measurements compared to the prior uncertainty. The assessments differentiate the fossil fuel CO2 emission into two sectors, an energy generation sector (power plants) and the complement ("other sector").

We find that XCO2 measurements alone provide a powerful constraint on emissions from larger power plants and a constraint on emissions from the other sector that increases when aggregated to larger spatial scales. NO2 measurements provide a powerful additional constraint on the emissions from power plants and from the other sector. Through the atmospheric constraint, more prior information on the CO2 emissions from power plants or on the differentiation of the NO2/CO2 emission factor reduces the uncertainty in CO2 emissions from the other sector.

If we ignore uncertainty in the small-scale structure of the lateral inflow this leads to a sizable overestimation of the uncertainty reduction. In other words we are too confident of our posterior emission estimate. In the example calculation the uncertainty reduction for one of the large power plants increases by 1/2 (from 19.6% to 28.1%) when uncertainty in small scale structure of the lateral inflow is ignored. For the non-energy-generation, for Berlin and summer it increases from ~55% to ~61% when small scale structure in lateral boundaries is assumed to be known. For the complement of Berlin, which is closer to the boundaries, it increases by about 1/3 in summer (from 11.97% to 16.31%). The other effect of ignoring uncertainty in small scale structure of lateral inflow will be a bias in the emission estimate that compensates for errors in the small-scale structure.

Our results suggest that the capability of CO2M measurements to constrain fossil fuel emissions varies between summer and winter cases. The main factor behind the larger constraint in the summer case are lower random and systematic errors in XCO2 measurements. There are, however, exceptions related to factors such as cloud cover and atmospheric transport.

The setup of the CCFFDAS and of the experiments focus on specific uncertain elements in the processing chain and its inputs. Further uncertain factors such as structural model errors are not covered, so that the quantitative assessments with the current prototype might be interpreted as a lower limit for posterior uncertainty. For example, we'd expect structural errors in the terrestrial biosphere model to come into play in particular in summer. We can, however, expect that, with sufficient research effort, future CCFFDAS's will be able to benefit from more accurate models of fossil fuel emissions (finer sectoral resolution, further observational constraints), of natural fluxes (more terrestrial observations including CO2M measurements of solar induced fluorescence, better meteorological driving data), and of atmospheric transport (constrained by observations of local meteorological conditions) and improved prior information. In this sense our performance assessments may provide a realistic indication of what can be achieved. In summary we find that the combination of CO2M with a suitable inversion system can provide useful estimates for urban scale emission reporting/verification. As for the global CCFFDAS (Thomas Kaminski, Scholze, Rayner, Houweling, et al. 2022), possible application modes are either a verification mode, in which the system is operated largely independently from inventory information or a synergy mode that derives a best emissions estimate by integrating bottom up information.

5 Analyses of the added value of co-emitted species using real satellite data

We analyse trace gas ratios across Europe, looking at the observed variability in trace gas to CO_2 ratios. Trace gas ratios can be used as a tool in multi-tracer local-scale inversions, but this requires knowledge and accurate representation of the relationship between CO_2 and co-emitted trace gases. Trace gases, like CO, are co-emitted with CO_2 during combustion and can act as a tracer for (incomplete-)combustion activities. Using these tracers potentially enhances our ability to separately estimate CO_2 anthropogenic and biogenic emissions. Because these species are co-emitted during combustion, we expect that enhancement ratios over populated areas like cities will have higher enhancements of shorter lived tracer concentrations, like CO, compared to CO_2 enhancements. We investigate enhancement ratios of CO compared to CO_2 in regions of observed pollution plumes (NO_2), investigating the temporal and spatial variability in ratios.

5.1 OCO-2 and TROPOMI plume overlap methods

For the identification of satellite CO_2 atmospheric column (XCO_2) and CO atmospheric column (XCO) ratios, we use the existing observing network which includes OCO-2 for XCO_2 observations and TROPOMI for XCO observations. We focus on TROPOMI observations flagged as NO_2 plumes using machine learning, as described in (Finch, Palmer, et Zhang 2021)Finch et al. (2022). Figure 20 shows the distribution of these plumes across Europe (-15° to 35° E longitude and 34° to 66° N latitude) for 2019. We use the latitude and longitude location and corresponding column concentrations for the plume centre as defined by Finch et al. (2022).

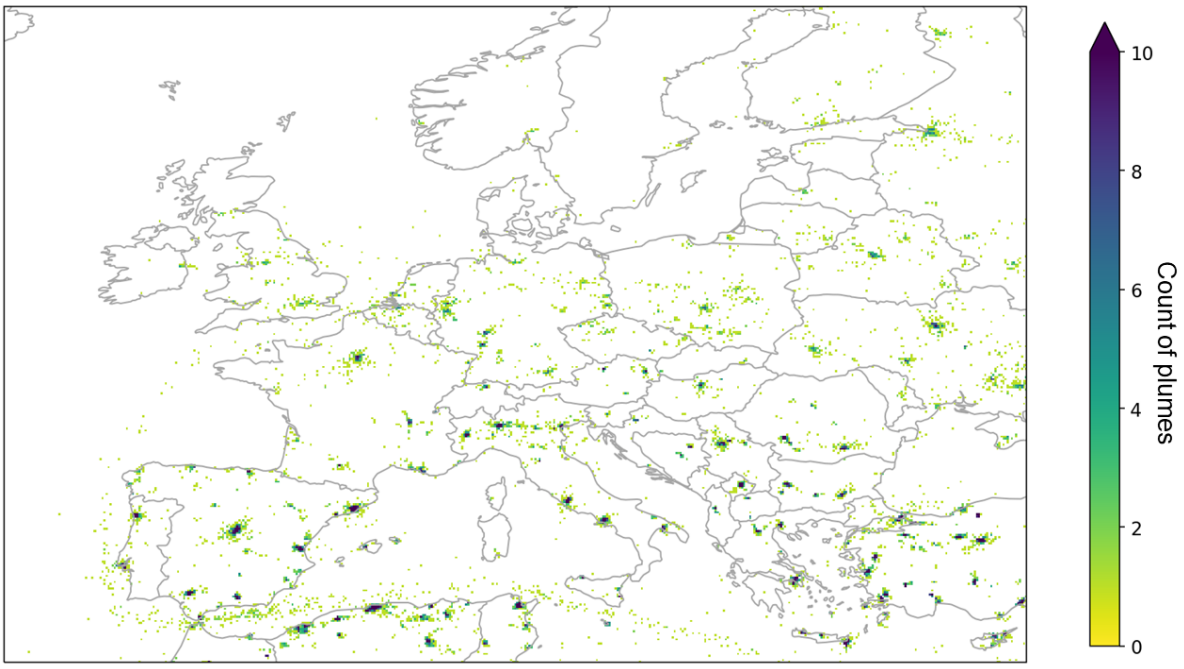

Figure 20: Map of 2019 TROPOMI NO_2 plumes used for identifying TROPOMI and OCO-2 overlaps in 2019. Plume identification methods are described in Finch et al. (2022).

Before calculating our species ratios, we remove background CO_2 and CO so we can better isolate the elevated concentrations due to local emissions. We estimate background contributions monthly using the lowest 5% of observations over the Atlantic Ocean averaged over -35° to -25° W longitude. We first tried grouping by 10° latitude bands, but found difficulties with background groups missing, so all latitude bands are treated as one region and we find the lowest 5% of observations in this region (30°-60° N). The temporal averaging by month is done to remove noise. The background is removed from observations to get the enhancement by species, ΔXCO_2 and ΔXCO, which we then use to calculate the enhancement ratio $\Delta XCO/\Delta XCO_2$.

We test different methods for identifying successful overlaps of OCO-2 observations and our identified TROPOMI plumes. We consider a successful overlap to be where at least one OCO-2 observation occurs within an allowable distance of a plume. Plumes are kept as individual observations with no temporal averaging with the exception of the 'Consistent plumes' which we describe next. We test allowable overlap distances of 1km and 10km. For the OCO-2 observations, we either require them to occur on the same day as the plume or we average the OCO-2 observations by month and only require them to occur within the same month as the plume.

We test a third method of time averaging where we first identify 'Consistent plumes' as defined by at least 5 plumes occurring within a 1km^2 or 10km^2 region. We find annual average plume concentrations for these 'Consistent plumes' and identify OCO-2 observations, also averaged annually, that fall within these 1km^2 or 10km^2 regions. This method is meant to reduce the variability introduced by taking annual averages by only considering plumes that consistently occur throughout the year.

5.2 Discussion of identified overlaps and enhancement ratios

Figure 21 shows the number of identified overlaps for 2018-2019 in our European domain. We find that very few overlaps are identified when we require the plume and OCO-2 observation to occur on the

same day. When we ease the time requirement to monthly overlaps, the number of overlaps increases by over an order of magnitude. For the 'Consistent plume' method we also find few overlaps and this is due to the lack of occurrence of 'Consistent plumes'. Given the limitations of the 'Consistent plumes' method, we do not consider it in further analysis.

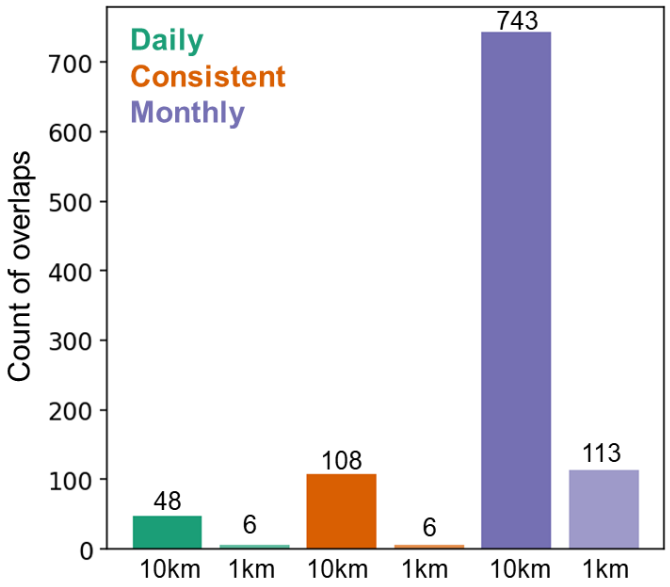

Figure 21: Number of OCO2 and TROPOMI plume overlaps identified over Europe for 2018-2019. For each overlap, we found at least one OCO-2 observation that was made within 1km or 10km of an identified TROPOMI plume. To identify overlaps we temporally average the OCO-2 observations by day (green) or month (purple). We also show the overlap counts for our 'Consistent plume' method (orange).

Figures 22 and 23 show the variability in XCO_2, XCO, and the enhancement ratio $\Delta XCO/\Delta XCO_2$ for the overlap methods using the same day (Figure 23) and the same month (Figure 22) for identifying overlaps. In general, the variability in the enhancement ratio follows the variability in XCO, with a few exceptions including in May in Figure 22 when the enhancement ratio changes less from April to May than XCO because of an increase in XCO_2. Figure 22 shows that within a given month there is large variability in XCO and the enhancement ratio across the European domain. We find that when overlaps are restricted to the same day (Figure 23), the range of enhancement ratios is lower than the monthly overlaps, but given the low number of overlaps found it is difficult to provide further analysis.

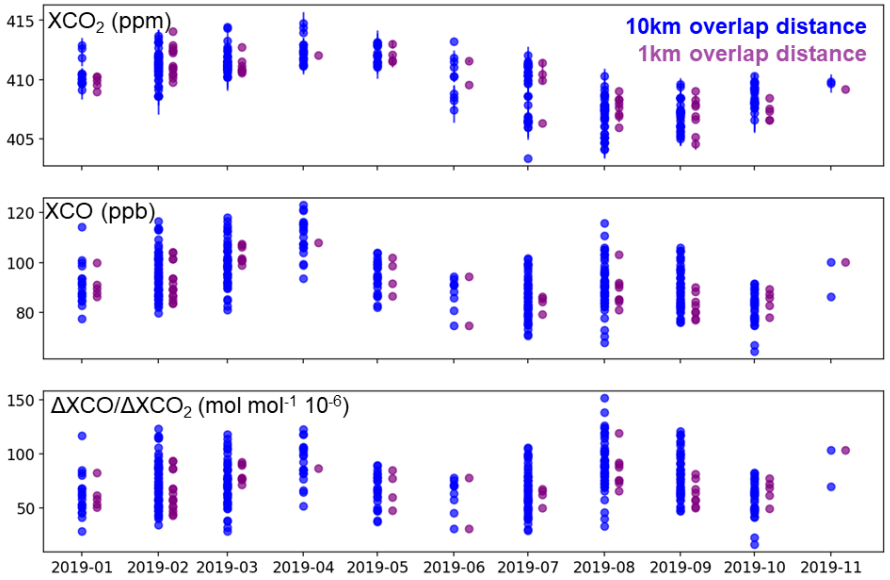

Figure 22: European XCO$_2$, XCO, and ΔXCO/ΔXCO$_2$ ratios in 2019 for the same month overlap method. The ΔXCO for each plume is used without averaging, while the ΔXCO$_2$ observations overlapping each plume are averaged over the whole month and averaged in space (10km^2 or 1km^2). The standard deviations for monthly averaged observations are shown as vertical lines.

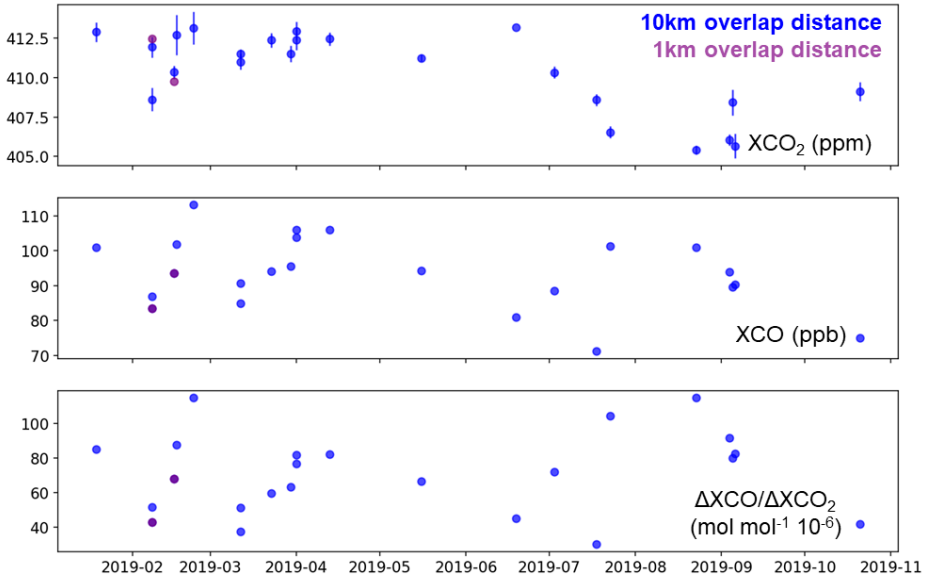

Figure 23: European XCO2, XCO, and ΔXCO/ΔXCO$_2$ ratios in 2019 for the same day overlap method. The ΔXCO for each plume is used without averaging, while the ΔXCO$_2$ observations overlapping each plume are averaged in space (10km^2 or 1km^2) for each day. The standard deviations for daily averaged observations are shown as vertical lines.

Figure 24 shows the distribution of ΔXCO/ΔXCO$_2$ ratios across Europe for 2019 using the various overlap methods. The enhancement ratio is generally lower in southern Europe, following lower XCO, with the highest ratios in regions like the southern U.K. and northern Italy. There are fewer counts of plume overlaps in northern Europe due to cloud cover and seasonal snow, making it difficult to comment on spatial variability of enhancements in the north. We do have a significant number of

overlaps in some southern cities like Madrid. Figure 25 shows enhancement ratios for Madrid in 2019 where we find varying ratios in space and time, though ratios never reach the values seen in northern European cities shown in Figure 24.

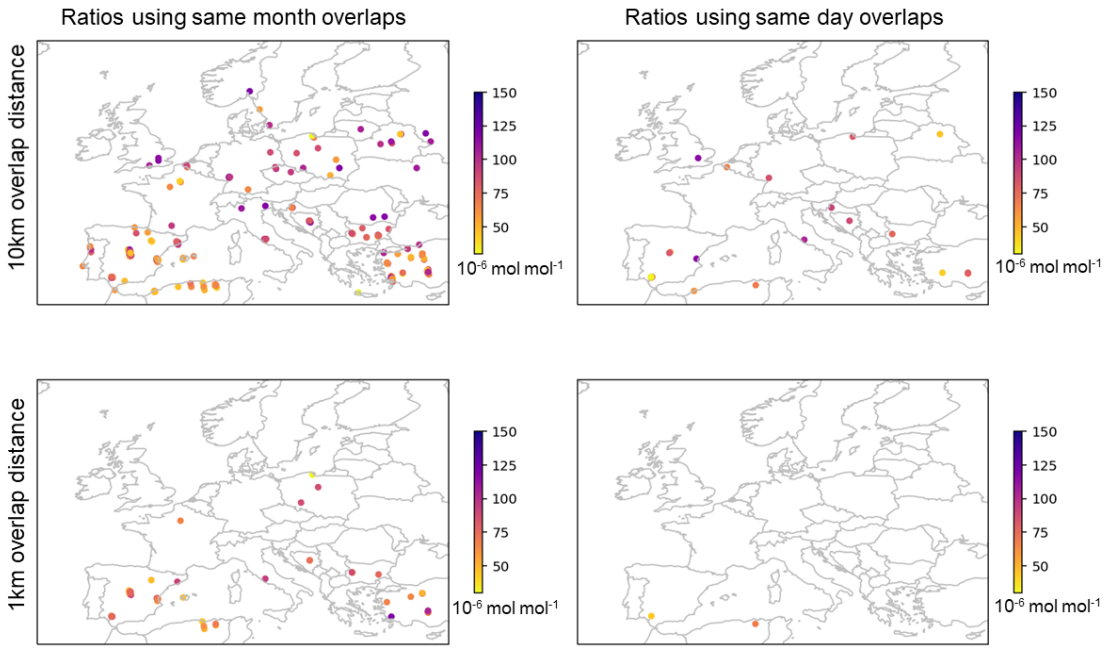

Figure 24: ΔXCO/ΔXCO$_2$ ratios over Europe in 2019. The XCO$_2$ observations overlapping each plume are shown averaged in space (10km^2 or 1km^2) and time (1 month or 1 day).

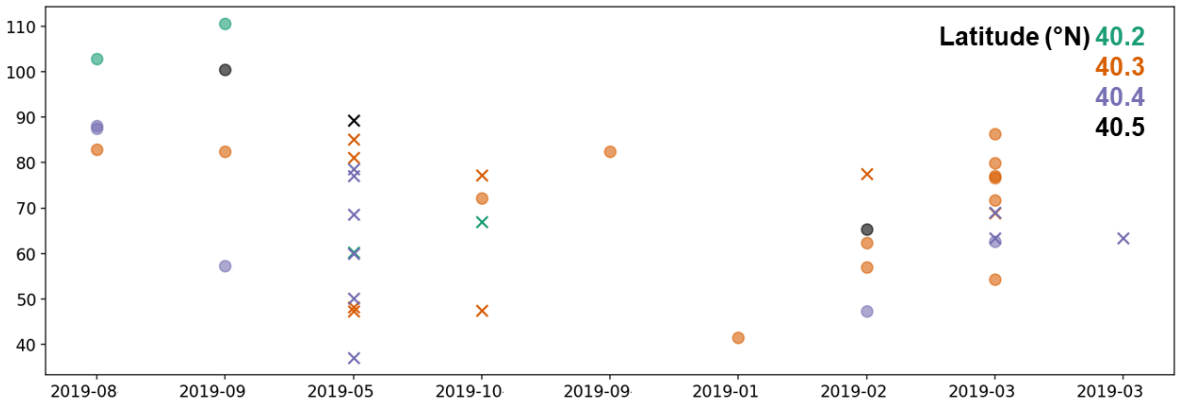

Figure 25: ΔXCO/ΔXCO$_2$ ratios over Madrid in 2019. Ratios are shown for the monthly averaging with 10km overlap method. Colours designate latitude (as shown) and shape designates longitude with spheres as -3.8°W and crosses as -3.7°W.

5.3 Conclusions

We find limitations in analysing ΔXCO/ΔXCO$_2$ enhancement ratios over Europe with the existing satellite observation system. There are few overlaps of OCO-2 XCO$_2$ observations within a close distance and on the same day as TROPOMI XCO over Europe, especially in northern regions (where

there are also high emission regions). Future satellites like CO2M that provide same day and location observations of XCO$_2$ and co-emitted trace gases, will overcome the limitations we find for the existing system. We find the most enhancement ratios when using a 10km distance between OCO-2 observations, averaged by month, and TROPOMI plumes. When using this method, we find that the variability in enhancement ratios is driven by XCO variability in our TROPOMI plumes and can vary at the city-scale within a year. A more in-depth analysis of temporal variability in enhancement ratios requires either modelling or satellites observing both XCO$_2$ and trace gases.

6 General conclusion

This deliverable documents three complementary studies regarding the potential and requirements of local scale inversion systems processing satellite images of XCO2 and of co-emitted species based on atmospheric transport models. These studies provide an ensemble of results on different aspects of the problem, but from which several general conclusions can be hired.

First, they confirm the potential of transport models to strengthen local scale inversions of emissions from cities and industrial plants in comparison to the light plume detection and inversion techniques studied in deliverables D4.3 and D4.4. Such a use of the model can improve the accuracy of the plume detection and thus, the emission estimates when applying the light plume inversion techniques to the corresponding plume identification (see section 3). Even though they are more preliminary, the tests of transport model and CNN based approaches for the plume inversion itself are promising and also demonstrate potential to improve the accuracy of the emission estimates (see section 3). Finally, the use of transport models can provide the potential to infer the emissions from specific plants or districts within a city (see section 4).

However, further developments and tests in more realistic conditions (e.g. testing the CNN based approaches to images with loss of pixels) are required to better characterise the potential of these approaches when processing real observations. Furthermore, these studies also show that exploiting this potential of atmospheric transport models for local scale inversions could require complex approaches to overcome transport modelling errors and that this complexity can raise challenges. On one hand, the computational cost of the CNN based approach studied in section 4 may not be critical, since some results presented in this section demonstrate that the approach could be trained with simulations for a limited number of sources over a limited period of time even if targeting a large amount of sources over a long period of time. Initially, the generation of these simulations would be relatively expensive compared to the use of the light plume detection inversion techniques presented in deliverables D4.3 and D4.4. However, in the long run, the cost of the CNN trained on these simulations would not be very large. On the other hand, the complex use of the model, which is not directly compared to the observations, in the approach proposed in section 3 may more critically challenge the ability to solve for the sectoral and spatial distribution of the emissions within cities. This ability is assessed here in section 4 based on city scale inversions with pseudo-data in which the transport model uncertainties are simulated in a relatively simple way that does not fully challenge the local inversions. This explains that these city scale inversions can safely rely on direct comparisons between the model and the satellite images using traditional metrics from larger scale inversion frameworks. A combination between the approaches developed in section 3 and 4 is probably needed to fully solve for the problem of spatial and sectoral resolution of the emissions. The topic needs to be addressed once the ability to quantify the total emissions using CNN-based approach will be strengthened.

Regarding the use of co-emitted species, the analysis of real satellite data in section 5 raises warnings. The lack of overlapping between the CO or NO2 plumes detected in TROPOMI data and of XCO2 observations from OCO-2 data highlights the challenges of the co-assimilation of existing data. However, the co-assimilation of the NO2 and XCO2 images from CO2M could be critical when solving for the spatial and sectoral resolution of the emissions within cities (section 4) and their co-registration by CO2M would overcome the problem of the lack of overlapping identified by the analysis of current data. Indeed, the analyses of synthetic CO2M errors within the CCFFDAS around Berlin indicates that city-scale data assimilation systems can provide useful estimates of the total city emissions from non-energy generation when relying on XCO2 images only but that the co-assimilation of NO2 images achieves a sizeable reduction in the uncertainty in the emission of the individual districts of the city. However, in general, for non-energy generation emissions the above data streams are not sufficient to provide relevant estimates at spatial resolutions finer than that of a district. Furthermore, the impact of

the variations of the emission ratios between species at fine spatial and temporal scales on the ability to cross information from different species in city scale inversion systems remains to be fully assessed (section 5).

Finally, sections 3 and 4 highlight, as well as deliverables D4.3 and D4.4, the impact of the uncertainties in fine scale patterns of the XCO2 background field underlying the plume from the targeted sources. The accurate modelling of such fields at the local scale is challenging so that the information from the images on this field is critical and the segmentation of the image between the plume and the background may still be a prerequisite for the local scale inversion approaches based on models as well as for the light plume inversion techniques analysed in deliverables D4.3 and D4.4.

Further developments and analysis are thus needed to fully exploit the potential of atmospheric transport models for local scale inversions. In the near-term, the light plume detection and inversion techniques analysed in deliverables D4.3 and D4.4 may keep on being the most suitable approaches to process images for a large number of target sources. However this deliverable raises promising insights on the use of local scale atmospheric transport models, even for the process of large datasets.

7 References

Balsamo, Gianpaolo, Richard Engelen, Daniel Thiemert, Anna Agusti-Panareda, Nicolas Bousserez, Grégoire Broquet, Dominik Brunner, et al. 2021. « The CO2 Human Emissions (CHE) Project: First Steps Towards a European Operational Capacity to Monitor Anthropogenic CO2 Emissions ». *Frontiers in Remote Sensing* 2. https://www.frontiersin.org/articles/10.3389/frsen.2021.707247.

Chollet, Francois. 2017. *Deep Learning with Python*. 1st éd. USA: Manning Publications Co.

Denier van der Gon, H.A.C, Stijn Dellaert, Ingrid Super, Jeroen Kuenen, et Antoon Visschedijk. 2021. « VERIFY : Observation-based system for monitoring and verification of greenhouse gases ». Horizon 2020 Societal challenge 5: Climate action, environment, resource efficiency and raw materials. https://cordis.europa.eu/project/id/776810/reporting.

Dumont Le Brazidec, Joffrey, Pierre Vanderbecken, Alban Farchi, Marc Bocquet, Jinghui Lian, Grégoire Broquet, Gerrit Kuhlmann, Alexandre Danjou, et Thomas Lauvaux. 2022. « Segmentation of XCO_2 Images with Deep Learning: Application to Synthetic Plumes from Cities and Power Plants ». *Geoscientific Model Development Discussions*, décembre, 1-29. https://doi.org/10.5194/gmd-2022-288.

Finch, Douglas, Paul Palmer, et Tianran Zhang. 2021. « Automated Detection of Atmospheric NO_2 Plumes from Satellite Data: A Tool to Help Infer Anthropogenic Combustion Emissions ». *Atmos. Meas. Tech.*, août, 1-21. https://doi.org/10.5194/amt-2021-177.

Kaminski, T., M. Scholze, M. Vossbeck, W. Knorr, M. Buchwitz, et M. Reuter. 2017. « Constraining a Terrestrial Biosphere Model with Remotely Sensed Atmospheric Carbon Dioxide ». *Remote Sensing of Environment*, Earth Observation of Essential Climate Variables, 203 (décembre): 109-24. https://doi.org/10.1016/j.rse.2017.08.017.

Kaminski, Thomas, et Peter Julian Rayner. 2017. « Reviews and Syntheses: Guiding the Evolution of the Observing System for the Carbon Cycle through Quantitative Network Design ». *Biogeosciences* 14 (20): 4755-66. https://doi.org/10.5194/bg-14-4755-2017.

Kaminski, Thomas, Marko Scholze, Peter Rayner, Sander Houweling, Michael Voßbeck, Jeremy Silver, Srijana Lama, et al. 2022. « Assessing the Impact of Atmospheric CO2 and NO2 Measurements From Space on Estimating City-Scale Fossil Fuel CO2 Emissions in a Data Assimilation System ». *Frontiers in Remote Sensing* 3. https://www.frontiersin.org/articles/10.3389/frsen.2022.887456.

Kaminski, Thomas, Marko Scholze, Peter Rayner, Michael Voßbeck, Michael Buchwitz, Maximilian Reuter, Wolfgang Knorr, et al. 2022. « Assimilation of Atmospheric CO2 Observations from Space Can Support National CO2 Emission Inventories ». *Environmental Research Letters* 17 (1): 014015. https://doi.org/10.1088/1748-9326/ac3cea.

Knorr, Wolfgang, et Martin Heimann. 1995. « Impact of drought stress and other factors on seasonal land biosphere CO2 exchange studied through an atmospheric tracer transport model ». *Tellus B: Chemical and Physical Meteorology* 47 (4): 471-89. https://doi.org/10.3402/tellusb.v47i4.16062.

Kuhlmann, Gerrit, Grégoire Broquet, Julia Marshall, Valentin Clément, Armin Löscher, Yasjka Meijer, et Dominik Brunner. 2019. « Detectability of CO_2 Emission Plumes of Cities and Power Plants with the Copernicus Anthropogenic CO_2 Monitoring (CO2M) Mission ». *Atmospheric Measurement Techniques* 12 (12): 6695-6719. https://doi.org/10.5194/amt-12-6695-2019.

Kuhlmann, Gerrit, Dominik Brunner, Grégoire Broquet, et Yasjka Meijer. 2020. « Quantifying CO_2 Emissions of a City with the Copernicus Anthropogenic CO_2 Monitoring Satellite Mission ». *Atmospheric Measurement Techniques* 13 (12): 6733-54. https://doi.org/10.5194/amt-13-6733-2020.

Kuhlmann, Gerrit, Valentin Clément, Julia Marshall, Oliver Fuhrer, Grégoire Broquet, Christina Schnadt-Poberaj, Armin Löscher, Yasjka Meijer, et Dominik Brunner. 2019. « SMARTCARB – Use of Satellite Measurements of Auxiliary Reactive Trace Gases for Fossil Fuel Carbon Dioxide Emission Estimation ». Zenodo. https://zenodo.org/record/4034266.

Kuhlmann, Gerrit, Stephan Henne, Yasjka Meijer, et Dominik Brunner. 2021. « Quantifying CO2 Emissions of Power Plants With CO2 and NO2 Imaging Satellites ». *Frontiers in Remote Sensing* 2: 14. https://doi.org/10.3389/frsen.2021.689838.

Lian, Jinghui, François-Marie Bréon, Grégoire Broquet, Thomas Lauvaux, Bo Zheng, Michel Ramonet, Irène Xueref-Remy, Simone Kotthaus, Martial Haeffelin, et Philippe Ciais. 2021. « Sensitivity to the Sources of Uncertainties in the Modeling of Atmospheric CO_2 Concentration within and in the Vicinity of Paris ». *Atmos Chem Phys* 21 (13): 10707‑26. https://doi.org/10.5194/acp-21-10707-2021.

Lorente, A., K. F. Boersma, H. J. Eskes, J. P. Veefkind, J. H. G. M. van Geffen, M. B. de Zeeuw, H. a. C. Denier van der Gon, S. Beirle, et M. C. Krol. 2019. « Quantification of Nitrogen Oxides Emissions from Build-up of Pollution over Paris with TROPOMI ». *Scientific Reports* 9 (1): 20033. https://doi.org/10.1038/s41598-019-56428-5.

Mahadevan, Pathmathevan, Steven Wofsy, Daniel Matross, Xiangming Xiao, Allison Dunn, John Lin, Christoph Gerbig, J. Munger, Victoria Chow, et Elaine Gottlieb. 2008. « A Satellite-based Biosphere Parameterization for Net Ecosystem CO2 Exchange: Vegetation Photosynthesis and Respiration Model (VPRM) ». *Glob. Biogeochem. Cycles.* 22 (juin). https://doi.org/10.1029/2006GB002735.

Rayner, Peter J., Anna M. Michalak, et Frédéric Chevallier. 2019. « Fundamentals of Data Assimilation Applied to Biogeochemistry ». *Atmospheric Chemistry and Physics* 19 (22): 13911‑32. https://doi.org/10.5194/acp-19-13911-2019.

Skamarock, C., B. Klemp, Jimy Dudhia, O. Gill, Dale Barker, G. Duda, Xiang-yu Huang, Wei Wang, et G. Powers. 2008. « A Description of the Advanced Research WRF Version 3 ». https://doi.org/10.5065/D68S4MVH.

Super, Ingrid, Stijn N. C. Dellaert, Antoon J. H. Visschedijk, et Hugo A. C. Denier van der Gon. 2020. « Uncertainty Analysis of a European High-Resolution Emission Inventory of CO_2 and CO to Support Inverse Modelling and Network Design ». *Atmospheric Chemistry and Physics* 20 (3): 1795‑1816. https://doi.org/10.5194/acp-20-1795-2020.

Tan, Mingxing, et Quoc V. Le. 2020. « EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks ». *arXiv:1905.11946 [cs, stat]*, septembre. http://arxiv.org/abs/1905.11946.

Tarantola, Albert. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898717921.

Zhang, Aston, Zachary C. Lipton, Mu Li, et Alexander J. Smola. 2022. « Dive into Deep Learning ». arXiv. https://doi.org/10.48550/arXiv.2106.11342.

Document History

Version	Author(s)	Date	Changes
	Joffrey Dumont Le Brazidec, Pierre Vanderbecken, Alban Farchi, Marc Bocquet (ENPC), Grégoire Broquet (CEA), Thomas Kaminski, Peter Rayner, Michael Voßbeck, Wolfgang Knorr (iLab), Tia	22/02/2023	Initial version

	Scarpelli (UEdin), Paul Palmer (UEdin)		
	//	05/05/2023	After review

Internal Review History

Internal Reviewers	Date	Comments
Name (Organisation)	dd/mm/yyyy	

Estimated Effort Contribution per Partner

Partner	Effort
ENPC	6
CEA	1
iLab	5
UEdin	
Total	