Synthesis and recommendations

Grégoire Broquet (CEA), Dominik Brunner, Erik Koene, Gerrit Kuhlmann (Empa), Joffrey Dumont Le Brazidec (ENPC), Johanna Tamminen (FMI), Thomas Kaminski (iLab), Nicolas Bousserez (ECMWF), Julia Marshall (DLR), Diego Santaren, Antoine Berchet (CEA) and the WP4 team

coco2-project.eu

D4.8 SYNTHESIS AND RECOMMENDATIONS

Author(s):

Grégoire Broquet (CEA), Dominik Brunner, Erik Koene, Gerrit Kuhlmann (Empa), Joffrey Dumont Le Brazidec (ENPC), Johanna Tamminen (FMI), Thomas Kaminski (iLab), Nicolas Bousserez (ECMWF), Julia Marshall (DLR), Diego Santaren, Antoine Berchet (CEA) and the WP4 team

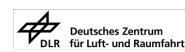# CoCO2: Prototype system for a Copernicus CO₂ service

**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

Table of Contents

Figures

1 Executive Summary

The deliverable provides a synthesis of the lessons learned from the activities in CoCO2 WP4. It provides a review of the conclusions from the previous deliverables D4.1-7 in this work package as well as new conclusions from the analysis in tasks T4.1-4. The main conclusions lead to recommendations and guidance regarding the spatial resolution of the atmospheric transport modelling and inversion and the observations to tackle the monitoring of the CO$_2$ anthropogenic emissions, the development of specific branch of local scale inversion based on computationally light data driven methods for the operational process of the XCO$_2$ spaceborne images of plumes downwind of emission hotspots, the coupling of such a branch with the CO2MVS multi-scale inversion prototype, the development of machine learning techniques for the local scale inversions, and regarding the benchmarking and inter-comparison of inverse modelling configuration at local and national scale, including the use of modular open source community codes. Further analysis and inter-comparisons with more complex synthetic tests cases or experiments using real data will be needed to refine the configuration assessment of the local scale inversion techniques based on lightweight techniques or machine learning, but the experiments in WP4 already provide strong insights on their strengths and accuracy, and on their level of readiness for operational applications. Large (national) scale inversions of anthropogenic CO$_2$ emissions keep on being more exploratory, and may finally have to connect to local scale inversions via the coupling of systems or scales within the multi-scale inversion prototype, or via the gradual increase in the spatial resolution of the national scale systems.

2 Introduction

2.1 Background

The CoCO2 project aims to develop a multi-scale inverse modelling prototype for the future CO$_2$ emission Monitoring and Verification Support Capacity (CO2MVS). Such a prototype will require proven and standard modelling and inversion techniques and configurations at the different spatial scales that are relevant for the global monitoring of anthropogenic CO$_2$ emissions. In this context, the CoCO2 work package 4 (WP4) developed and tested methods and systems to monitor the anthropogenic CO$_2$ emissions at national to facility/city scale. The first three tasks of WP4 (T4.1, T4.2 and T4.3) were dedicated to modelling and inversion strategies at the local scale. A wide range of models and methods were implemented and analysed with respect to their capabilities to simulate the plumes from large sources (cities, industrial sites), to detect such plumes in satellite XCO$_2$ images such as those from the future CO2M mission, and to quantify the sources. The fourth task (T4.4) was dedicated to national scale inversion systems, aiming to assess the optimal configurations (inversion parameters, type of input products, observation datasets etc.) and the potential of CO2M for the quantification of national to sub-national CO$_2$ and CH$_4$ anthropogenic emissions and natural fluxes. These different tasks were designed to demonstrate the capabilities of various implementations of local and national scale systems and to benchmark them separately at the local and national scales.

There was a subsequent need to synthesize the expertise and lessons learned from these tasks, to connect those from the local- and national-scale activities and to take advantage of this synthesized knowledge for the development of the global multi-scale inverse modelling prototype. This was covered by task 5 in WP4 (T4.5), corresponding to this deliverable D4.8.

2.2   Scope of this deliverable

2.2.1   Objectives of this deliverables

The aims of this final deliverable of WP4 are

- to synthesize the lessons learned from T4.1-T4.4 regarding the capabilities of the future CO2MVS operational system to monitor hot spot (urban, industrial) emissions and national to sub-national emission budgets
- to provide guidance for benchmarking of local and national scale inversion approaches and systems
- to synthesize the lessons learned from T4.1-T4.4 regarding the optimal strategies for local and national scale inversions
- to provide guidance for the integration of these strategies into the global multi-scale integrated prototype of the future operational CO2MVS developed in WP6
- to confront the results and conclusions from local and national scale inversions and to provide insights in how to build connections between local and national scale systems and between the multi-scale prototype and independently developed national systems

2.2.2   Work performed in this deliverable

This deliverable mainly consists of a synthesis of the different deliverables from tasks T4.1-T4.4, but also includes insights from WP5. This synthesis provides recommendations and guidance for the development of the multi-scale inversion prototype. Regarding these aspects, new specific discussions and conclusions are derived from the WP4 studies that have been documented in deliverables D4.1-7.

Furthermore, some of the developments and tests in tasks T4.1-4 have been extended and improved in T4.5 to strengthen the conclusions in this final deliverable.

2.2.3   Deviations and counter measures

This deliverable inherits deviations from other deliverables due to staffing problems and due to the lack of suitable input data or results. Some of the delayed activities of Tasks T4.1 to T4.4 were planned to be resumed in T4.5 to provide additional inputs for this synthesis. However, as detailed in the following, time was ultimately too short to complete these activities despite significant efforts.

One of the objectives of T4.4 was to assess the potential of CO2M for national-scale inversions of the CO$_2$ anthropogenic emissions based on Observing System Simulation Experiments (OSSEs). This objective could not be fulfilled because the required synthetic CO2M data, generated in T5.4 of WP5, were provided too late to be incorporated by the national-scale inverse modelling teams. Therefore, the assessment of the potential of CO2M at national scale provided in this synthesis is only based on results obtained in T4.4 with real observations from the current observation network.

A second delayed activity was to conduct specific experiments with two nested systems coupling national (from T4.4) and city scale (from T4.3) inversions to provide insights into the optimal coupling between these scales. However, the full configuration of these systems was not ready early enough to enable a proper analysis in the frame of T4.5. Alternatively, we expected some insights from the assimilation of the local and national scale Monte Carlo inversion estimates (generated in WP4 and delivered to WP6 in the frame of D4.7) in the multi-scale inversion prototype of WP6 to feed the topic of the reconciliation between local and national scale analysis. However, the assimilation of local and national scale inversion products in the multi-scale inversion prototype is still in the testing phase.

The question of how to optimally couple local and national scales can thus not be answered in this report, which instead focusses on discussing the complementarity between local and national-scale inversions. This is not only because of a lack of results from WP4, but also due to limitations in the current observing system. Indeed, there is (1) a lack of transects of plumes

from CO_2 emissions hotspots in existing satellite observations of total column CO_2 (XCO_2) from OCO-2/3 over Europe (see D6.5 in WP6), (2) a lack of CO_2 stations suitable for the estimation of urban CO_2 emissions, and (3) a lack of control of anthropogenic CO_2 emissions at the national scale even in Europe despite its comparatively dense network of in-situ stations (as demonstrated by the inversions in T4.4 which were focused on this continent). This limitation made it impossible to compare the results from the local and national scale inversions using real observations, and to identify optimal configurations or coupling strategies to increase the consistency between the local and regional estimates.

3   Synthesis from the local scale activities in T4.1-T4.3

The activities dedicated to local scale modelling and inversions were split into three tasks: T4.1 focused on the local scale modelling of CO_2 and NO_x plumes from cities and power plants, T4.2 focused on the use of light-weight data driven techniques for the detection and inversion of XCO_2 and NO_2 plumes in satellite observations, and T4.3 was dedicated to the development of approaches exploiting the information from high-resolution atmospheric transport models for the detection and inversion of XCO_2 and NO_2 plumes in satellite images.

3.1   Conclusions from plume modelling activities (T4.1)

Task 4.1 consisted of a CO_2 and NO_x/NO_2 simulation intercomparison for seven cases of strong point sources. The aim of this task was to **assess the performance** of current high-resolution transport models, **identify critical elements** affecting the simulations, and **develop recommendations** for the optimal simulation of the plumes. The task also aimed to build **a library of plumes** for method development and benchmarking. A modelling protocol was established in the frame of D4.1 and refined in the frame of D4.2 (see the appendix of D4.2). The protocol ensured that differences between the simulation outputs were only due to (1) the model used to simulate the atmospheric transport, and (2) the horizontal resolution of the model. Five models were considered (COSMO-GHG at 1.1 km horizontal resolution, ICON-ART at 2 & 6 km, LOTOS-EUROS at 1 km, MicroHH at 100 m, WRF-CHEM at 1 km).

To **assess the performance** of the models, we compared them to in-situ data. The best fits between the simulated and observational data were obtained with MicroHH at 100 m horizontal resolution, reproducing both the "width" and "amplitude" of observed plumes very closely. With degrading resolution, the fit becomes worse as both the "width" (spreading out more) and "amplitude" (decreasing) of the plume is captured less well. We also compared the models to remotely sensed data (including TROPOMI NO_2 data) but we only made a qualitative assessment based on this comparison, namely, that all models did a reasonable simulation of the plume properties as visible in the remotely sensed data.

The **critical elements** affecting the simulations depend on the modelling targets. The MicroHH set-up was the only one which was sufficiently close to matching the in-situ data. However, the resolution required to generate realistic synthetic CO2M observations (e.g. for inversion OSSEs) can be slightly coarser than that of MicroHH. All models at 2 km or finer resolution appeared to generate similar types of plumes in CO2M images. The use of high-resolution models did not seem to be required for such simulations owing to the 2 km spatial resolution of CO2M.

The **recommendations** from this study emphasized that *model spatial resolution* (the finer resolution, the better the fit to the observations), *atmospheric boundary layer development* (the better the model simulates the true atmospheric boundary layer development or thus numerically reproduces meteorology, the better the fit to the observations) and *chemistry scheme* were the major contributors to the simulation (in)accuracy. As mentioned in the previous paragraph, the recommendations depend on the modelling target. A very high-

resolution large eddy simulation (LES) model with a full chemistry scheme (like MicroHH) is required to simulate in-situ data of reactive trace gases. Our *qualitative* approach suggests that the 2 km atmospheric models driven by an accurate meteorological scheme (like the 'online' COSMO-GHG, ICON-ART, WRF-CHEM models; as well as the LOTOS-EUROS 'offline' model) can simulate 2-km column-averaged XCO_2 data sufficiently well.

In the frame of D4.8, we want to bring some nuance to the above conclusions from D4.2:

- Our approach here was *qualitative* rather than *quantitative.* A remaining question was why the MicroHH plumes remained more elevated over the background (i.e., more concentrated and less dispersive) than the plumes from the other models, even when sampled to synthetic CO2M pixels. One possible reason for this is that plume rise was accounted for in a simplified way by simulating power plant stacks as point sources in horizontal space with altitude-varying profiles. For coarse models, this led to (too) fast initial mixing, since emissions from a point source are immediately spread over the size of a grid box. For high-resolution models like MicroHH, however, this led to too little mixing and thus the plume remained more concentrated compared to observations close to the point source. Resuming the activities from T4.1, the MicroHH developers (Krol et al., 2023) have made changes to their code since the delivery of D4.2 to allow plume rise to be simulated explicitly within their model, which led to a more realistic dispersion of the trace gases close to the point source. This approach can also account for a possible lifting of the top of the atmospheric boundary layer by the rising plume. This presents an interesting avenue for further research.

- The models used for Numerical weather prediction (NWP, i.e., COSMO, ICON-ART and WRF at 1 or 2 km horizontal resolution and with the corresponding tracer transport modules), and the LES model (at 100 m horizontal resolution) showed *similar, but not identical* performance to simulate 2 km x 2 km CO2M pixel data. The current suggestion is that the mesoscale NWP models likely perform well enough, and thus that high resolution LES models are likely not needed, but a more quantitative approach to confirm this could be valuable. A new study is suggested, assessing whether the simulated MicroHH plumes fit well with remotely sensed total-column data at long distances, in addition to fitting well with the in-situ data recorded relatively close to power plant stacks.

- The **library of plumes** developed in T4.1 (https://zenodo.org/records/7448144) gathers relatively simplified total column images with no systematic errors and other issues that may impact the accuracy of actual CO2M images. In other words, the training of emission estimation methods on this library of plumes (e.g., the fine-tuning of direct estimation methods, or the training of neural networks based on the corresponding pseudo-images) may not be sufficient to tackle the CO2M data that will eventually be recorded. It is currently beyond the state-of-the-art to compute realistic synthetic CO2M data for such a purpose, mainly due to missing knowledge on the systematic error patterns that one may expect in reality.

- The downwind profiles of $NO_2:CO_2$ (or $NO_x:CO_2$) ratios was highly variable across the models. These profiles are not just a function of how the chemistry module is implemented in the models (e.g., full chemistry versus exponentially decaying functions for the simulation of NO_x). They are also a function of the model spatial resolution. As the reaction rates vary throughout the plume (e.g., they are different at the edges of turbulent eddies than within them) a higher resolution model should better capture the true chemistry. However, we did not get sufficient comparisons to in-situ data to assess such a tendency. As CO2M will co-register XCO_2 and NO_2 images, more research should be done to refine the modelling of the plume chemistry effects to support the NO_x and CO_2 emission estimates.

3.2 Conclusions from the light plume inversion activities (T4.2)

In task T4.2, we studied different methods to derive local scale CO_2 (and NO_x) emissions from cities and power plants based on XCO_2 (and NO_2) images from individual or ensembles of overpasses of a satellite spectro-imager such as CO2M. The focus has been on computationally light methods, i.e., approaches which are heavily data-driven and which do not rely on numerical atmospheric transport models. **The methods have been benchmarked** using 1) synthetic CO2M data extracted from the SMARTCARB data set simulated with the COSMO-GHG model, that covers 16 emission sources in central Europe (Kuhlmann et al 2019) and 2) real TROPOMI NO_2 observations over the Matimba/Medupi large emission source in South Africa.

We have studied four methods for analysing instantaneous plume images: the inversion of a Gaussian Plume model (GP), the Integrated Mass Enhancement method (IME) and two versions of the Cross-Sectional Flux method (CSF and Light CSF, LCSF). In addition, the Divergence method (Div) was tailored for CO_2 emission quantification at the annual scale by averaging satellites images (Hakkarainen et al., 2022). Deliverable D4.3 provides a theoretical background and an initial general description for these methods. Deliverable D4.4 and then the publications from T4.2 (Hakkarainen et al. 2022, 2023a and 2023b, Kuhlmann et al. 2023 and Santaren et al. 2023) provide details and updates regarding the specific configurations that have been tested.

The **computation times** of the different implementations were analysed using one month of cloud-free XCO_2 and NO_2 data over the SMARTCARB domain. It should be noted that the computation time is less dependent on the specific method than on the choice of the pre-processing algorithm, the optimization of the implementation, and the number of fitting parameters selected for each method. For example, in the tests documented in D4.4, CSF, IME and GP used a complex pre-processing algorithm for detecting the plume locations and for computing a curved coordinate system following the plume's ridge, taking about 20 min computation time. On the other hand, the LCSF defined a non-curved coordinate system based on the wind direction near the source, which is significantly faster (<3 min). The computation time to determine the emissions from the pre-processed images are <30s for IME, 3-9 min for LCSF, and 1-5 min for CSF. With 90 min the GP method was the most expensive method. This can be explained by the specific implementation used here, including an optimization of the centre curve of the Gaussian plume model, which required recomputing the curved coordinate system in each iteration of the inversion process. The Div method takes about 20 minutes. The methods were not implemented with the goal of optimizing computation times, thus a reduction of the computation time is possible and might be desired when applying the methods for near-real-time operational monitoring of the sources worldwide. Overall, computation time should not be a major roadblock when implementing one (or several) of these methods in an operational system.

All methods were implemented in the open source *ddeq* Python library for data-driven emission quantification (Kuhlmann et al. 2023). The library provides shared data formats (using the xarray library) and functions for input, pre-processing, post-processing and output, with the flexibility for the different methods to use different sequences of processing steps. The only requirement to apply the methods to new datasets (i.e., satellite data, wind products or source locations), is to implement new functions to read the data to the data format shared by the library. The overall processing of the data is thus highly automatized.

The initial conclusions from the SMARTCARB benchmarking inter-comparison exercise have been detailed in deliverable D4.4, and then updated in Santaren et al. (2023), with **scores of RMS errors in the annual emission estimates** of 20% (GP), 27% (CSF), 31% (LCSF), 55% (IME) and 79% (Div). These scores correspond to the most realistic benchmarking scenario, which includes the loss of observations in the satellite swath due to clouds, realistic observational noise and the use of ERA-5 winds for the analysis (instead of the wind field from

the COSMO-GHG simulation behind the benchmarking dataset) which have a coarser horizontal resolution (~0.25° vs. ~0.01°) than the satellite images.

The tests showed that GP, CSF and LCSF performed best in retrieving instantaneous emissions, with interquartile ranges in deviations from the true emissions between 20% - 60% for most cases. The evaluation of the results included the **confrontation of the (generally partial) diagnostics of uncertainty by the different methods against the true error in the emission estimates**. We introduced thresholds for the acceptable uncertainty below which estimates are selected. The **overall accuracy and number of estimates** for a given method are impacted by the values assigned to these thresholds (Fig. 1): when this threshold is low, the number of estimates strongly decreases and the overall accuracy increases more or less depending on the inversion method. The uncertainty estimates of the CSF and IME methods were sufficiently reliable and accurate to be applied for outlier detection and data screening while the uncertainty estimation of the GP and LCSF methods should be improved. Ensemble approaches which gather results of different combinations of inversion methods show important increases in the number of estimates compared to any single method but do not provide any significant improvement in terms of accuracy.
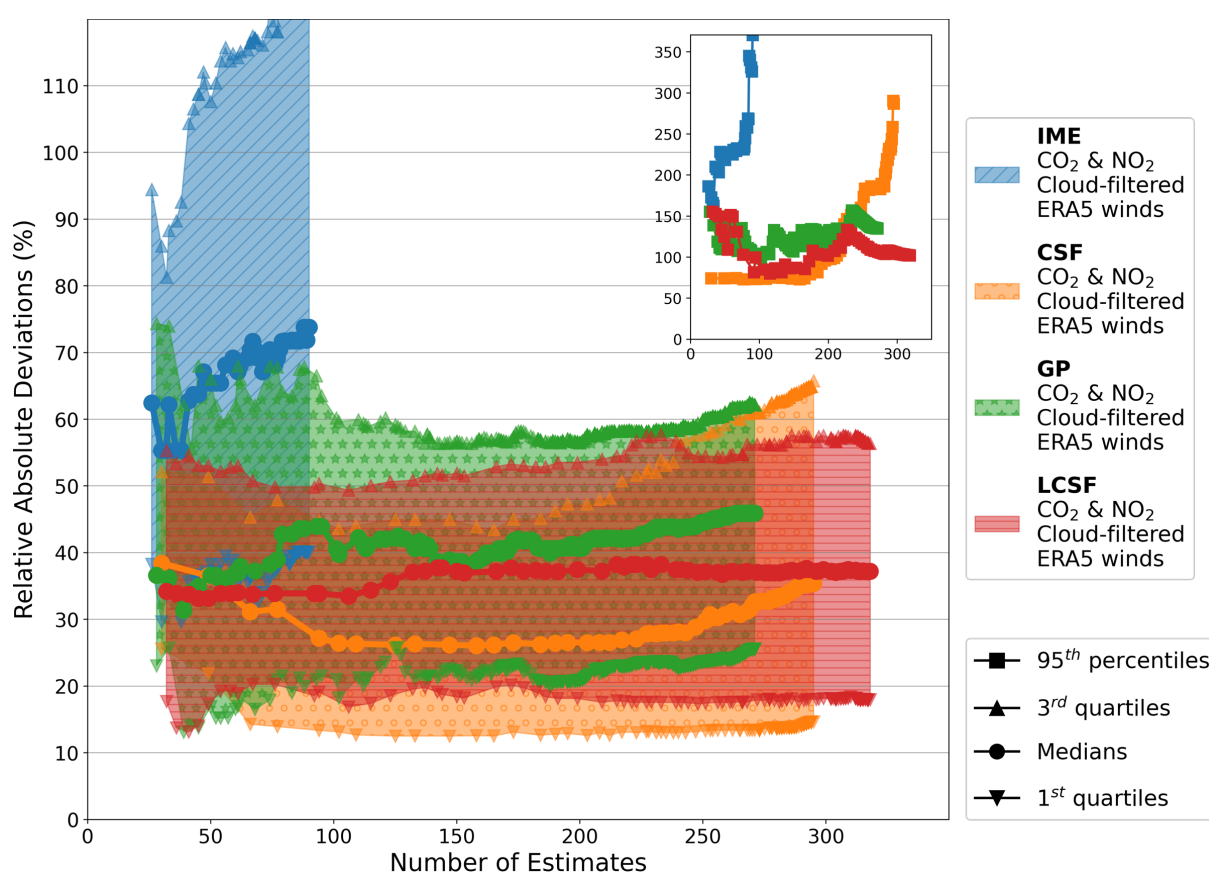
Figure 1: Accuracy vs. number of the instant estimates of emissions from cities and power plants when applying the light plume inversion methods of T4.2 to XCO$_2$ and NO$_2$ cloud-filtered images and using ERA5 winds. The number of estimates and accuracy vary with the applied uncertainty threshold. A stricter threshold results in a lower number of estimates but usually in improved accuracy per estimate. The filled areas represent the inter-quartile ranges of the distributions of the relative absolute deviations between the estimates and the actual emissions. Note that the relative absolute deviations of the CSF, GP and LCSF methods are characterized by IQRs whose first/third quartiles are at most equal to ~20 %/~60 %. The 90th percentiles of the distributions are shown in the inset (fig. included in Santaren et al., 2023; update of the results documented in D4.4).

The estimation of instantaneous emissions is generally more uncertain than the estimation of annual emissions. The seasonal cycles of the emissions were not reproduced with any of the methods. Further development is thus needed to improve the estimation of the intra-annual variability of the emissions.

All methods were also **tested with real data by considering one year of Sentinel 5P/ TROPOMI NO$_2$ observations** to derive NO$_X$ emission estimates for the Matimba-Medupi power plant area in South Africa (Hakkarainen et al. 2022, Hakkarainen et al. 2023b). Compared to Eskom-reported emissions (https://www.eskom.co.za/dataportal/ emissions/ael/, Fig.1), each of the methods strongly underestimated the magnitude of the source if the commonly used NO$_2$-to-NO$_X$ conversion factor of 1.32 was applied. High-resolution large-eddy simulations with the MicroHH model with a simple plume chemistry scheme showed that due to the non-linear chemistry, the **optimal NO$_2$-to-NO$_X$ conversion factors** are method-dependent and the derived values for these factors are substantially (more than 50%) higher than the commonly used value of 1.32. The MicroHH simulations covered periods of 48h, therefore it was not possible to derive optimal scaling factors for the Div method. By applying the updated conversion factors, the IME, LCSF and CSF methods yielded the most accurate emission estimates (within about 20-30% of the reported emissions) while GP and Div (with the non-optimised conversion factor) yielded poorer estimates (within 50-60% of the reported emissions). It is worth noting that despite the improved scaling factors, all the methods still underestimated the emissions.

All lightweight methods are based on simplifications and assumptions. Our tests indicated that **the GP, CSF and LCSF methods perform most robustly in CO$_2$ emission estimation**. The LCSF method appears to be the most convenient to maximize the number of instant estimates, the GP method to derive the most accurate annual estimates (when applying a strict uncertainty threshold), and the CSF method to derive the most accurate instant estimates when considering a moderate uncertainty threshold. However, the performance of the methods critically depends on the pre-processing techniques, like de-noising, background estimation, and plume detection. The poor scores of the IME method, for example, may have been impacted by the specific plume detection approach and may be further improved by better accounting for cloud cover. During the benchmarking activities, all the computationally light methods were progressively optimised and tuned, and their performances regularly improved. Thus, it is expected that further **optimization of the practical implementation of the methods and pre-processing techniques** could be achieved and could lead to improved robustness, uncertainty quantification and outlier detection.

At this stage, it would be important to continue the development and optimisation of all the methods and their uncertainty characterization and to consider if **ensemble methods** would eventually generate the most robust estimates, and not just more estimates, as they do now. Despite these new developments, keeping a high level of automation via a transparent and modular library such as ddeq will be critical for the operational application of proven and standardized methods.

3.3 Conclusions from the plume inversion activities based on local scale transport models (T4.3)

Task T4.3 was dedicated to the development of atmospheric inversion approaches at the local scale exploiting the information from high-resolution atmospheric transport models for the quantification of city and industrial plant CO$_2$ emissions based on spaceborne images of their XCO$_2$ plumes. The observations assimilated by these systems potentially also included images of species co-emitted with CO$_2$ like CO and NO$_2$. The corresponding deliverable D4.5 documented three complementary studies, which provided valuable insights into different aspects of the use of atmospheric transport models:

- The development of an advanced processing of XCO$_2$ images for the **detection and inversion of plumes based on Convolutional Neural Networks (CNN) trained with high resolution model simulations** (Dumont Le Brazidec et al., 2023a and 2023b). The expected advantage of a CNN, when trained with realistic model simulations with exact knowledge of the position of the plume and of the emission strength, is that it can optimally exploit the skill of the transport model to simulate typical plumes under a wide range of transport conditions when processing the information contained within an image without being affected by model errors when modelling a plume at a given time (which are potentially large). This implicitly assumes that the model simulations used to train the CNN are not biased and represent well the distribution of plumes over wide ranges of dispersion conditions, while the model simulation of the direction and shape of a real plume at a given time often lacks precision. Thanks to exploiting the model skills, the CNN-based detection and inversion of plumes was expected to provide more accurate results than the simpler techniques tested in Task T4.2, especially for complex transport conditions, and if tackling reactive species in addition to CO$_2$. It was tested on the same benchmarking framework based on the pseudo CO2M SMARTCARB simulations as that used in the T4.2. However, the application of the CNN-based methods was restricted to test cases without gaps in the images.

- The development and application of a city-scale **Carbon Cycle Fossil Fuel Data Assimilation System - CCFFDAS** (Kaminski et al., 2022a) to pseudo data experiments for the city of Berlin and its surroundings. The aim of the analysis with this city-scale inverse modelling system was to assess the ability to solve for the **spatial or sectoral distribution of the emissions within the urban areas** at the resolution of CO2M using a high-resolution transport model that solves for the relationship between the emissions in each modelling grid cell and all observations. Since the CCFFDAS controls the parameters behind the emission models rather than the emissions themselves, a parallel objective was to assess the impact of the co-assimilation of the NO$_2$ observations from CO2M which may provide additional useful constraints on these parameters.

- The analysis of the **CO$_2$/CO local concentration enhancement ratios over emission hotspots in Europe** based on available satellite NO$_2$ (to define the location of the emission hotspots), CO and CO$_2$ observations (from TROPOMI and OCO-2) to assess the **reliability of the co-assimilation of co-emitted species for the local scale inversions**.

The general conclusions of these studies confirmed the potential of the use of transport models to strengthen the source estimates compared to the T4.2 light local scale inversion techniques (when using CNN-based approaches) and to infer emissions from specific plants or districts within a city (with CCFFDAS approaches, and when co-assimilating XCO$_2$ and NO$_2$ satellite data). The **extrapolation skill of the CNN-based approach**, i.e., the ability to infer emissions from sources that are not covered by its training set of simulations, was evaluated with first positive insights. It is a critical aspect of the method since its operational use should rely on a training over a limited set of simulations to cover wide sets of sources (mainly for computational reasons, since high resolution transport models are computationally expensive). After the delivery of D4.5, further improvement and analysis of the CNN-based inversion of the XCO$_2$ plumes, and in particular comparisons between the results obtained with the CNNs and with the CSF method that was tested in T4.2 helped strengthen the conclusions from this deliverable. They are documented in the publication by Dumont Le Brazidec et al. (2023b).

In D4.5, the **CNN-based detection of the plumes in XCO$_2$ images** (i.e. the CNN-based segmentation of the XCO$_2$ images) was showing better performances than the statistical threshold tests used in T4.2, even when tackling scenes with overlapping plumes or plumes from cities and power plants not included in the CNN training sets. These results were obtained without using wind-fields or NO$_2$ images in input of the CNN.

The proven effectiveness of **CNNs in inverting XCO$_2$ plumes from large point sources**, using pseudo XCO$_2$ and ERA-5 wind field re-analysis is noteworthy. In T4.3, image-to-scalar CNN models have been effectively employed for estimating power plant emissions, focusing on deducing the flux rate associated with anthropogenic plumes. Recent research, as highlighted in D4.5 and in Dumont Le Brazidec et al., 2023b, strongly supports the feasibility of developing a "universal" CNN model, trained on a selected group of power plants, and deriving highly accurate estimates for various plants. The generalisation capabilities of the CNN models used in T4.3 have been rigorously tested on unobserved images from different regions within the SMARTCARB domain, with a focus on plumes from the Boxberg, Lippendorf, and Turow power plants. Each model, trained with data excluding the target plant to which it is associated, has shown notable accuracy, as evidenced by median relative errors around 20–25% and median absolute errors significantly lower than those achieved with the CSF method for instant emission estimates based on individual XCO$_2$ images (Fig. 2). The use of the results from the CNN-based plume detection or of the NO$_2$ images co-registered with the XCO$_2$ images in input of the CNN models for inversion do not increase significantly this accuracy. This demonstrates that the CNN-based inversion problem could be tackled without a preprocessing step for the plume detection. However, in T4.3, the CNN-based inversion was not tested on plumes from cities due to the lack of training/testing datasets specific to such plumes.

The total training time of the CNNs for inversions in Dumont Le Brazidec et al. (2023b) is of approximately four hours (using an Nvidia Quadro RTX 5000 16GB GPU). The application time on an image is almost instantaneous. This makes the method **highly competitive with the light local scale inversion techniques from T4.2 for operational use**, provided that the approach has a high extrapolation skill.

However, the current version of the CNN-based inversion algorithm may require adjustments when applied to new datasets that significantly differ from the training data, such as those with vastly different topographies. **The level of automation of this algorithm is thus currently limited.** To align the CNNs with the new data characteristics, fine-tuning through retraining on a subset of the new data or similar data may be necessary. This process aims to rectify any mismatches between the original training data distribution and that of the new dataset. Acquisition of these new data could involve conducting targeted simulations or employing specific data augmentation techniques. While the existing CNNs are designed for robustness and generalizability, notable variances in the data, especially in the characteristics of the plumes, could compromise the accuracy of the results when directly applied to real-world scenarios.

It is thus important to acknowledge the **challenges in generalising CNN-based results to more complex scenarios**. These include plumes from urban environments, which are inherently more complex and exhibit lower signal-to-noise ratios than those from power plants, and areas with more intricate topography than Eastern Germany. These challenges are intrinsically linked to the size of the training dataset, raising questions about the extent of simulations required for global application. Moreover, the effectiveness of CNN models, trained on simulated data and applied to actual satellite images, warrants thorough scrutiny to ensure their generalisability and accuracy in a variety of real-world scenarios. Additionally, the challenges posed by the loss of pixels in the XCO$_2$ images due to cloud cover and the presence of systematic error patterns in these images are critical topics that need to be addressed.

In summary, while the use of CNNs in the context of T4.3 was highly promising and demonstrated effectiveness, there is a clear need for ongoing research and development to address these challenges and to fully assess the potential of CNNs in broader and more complex conditions.
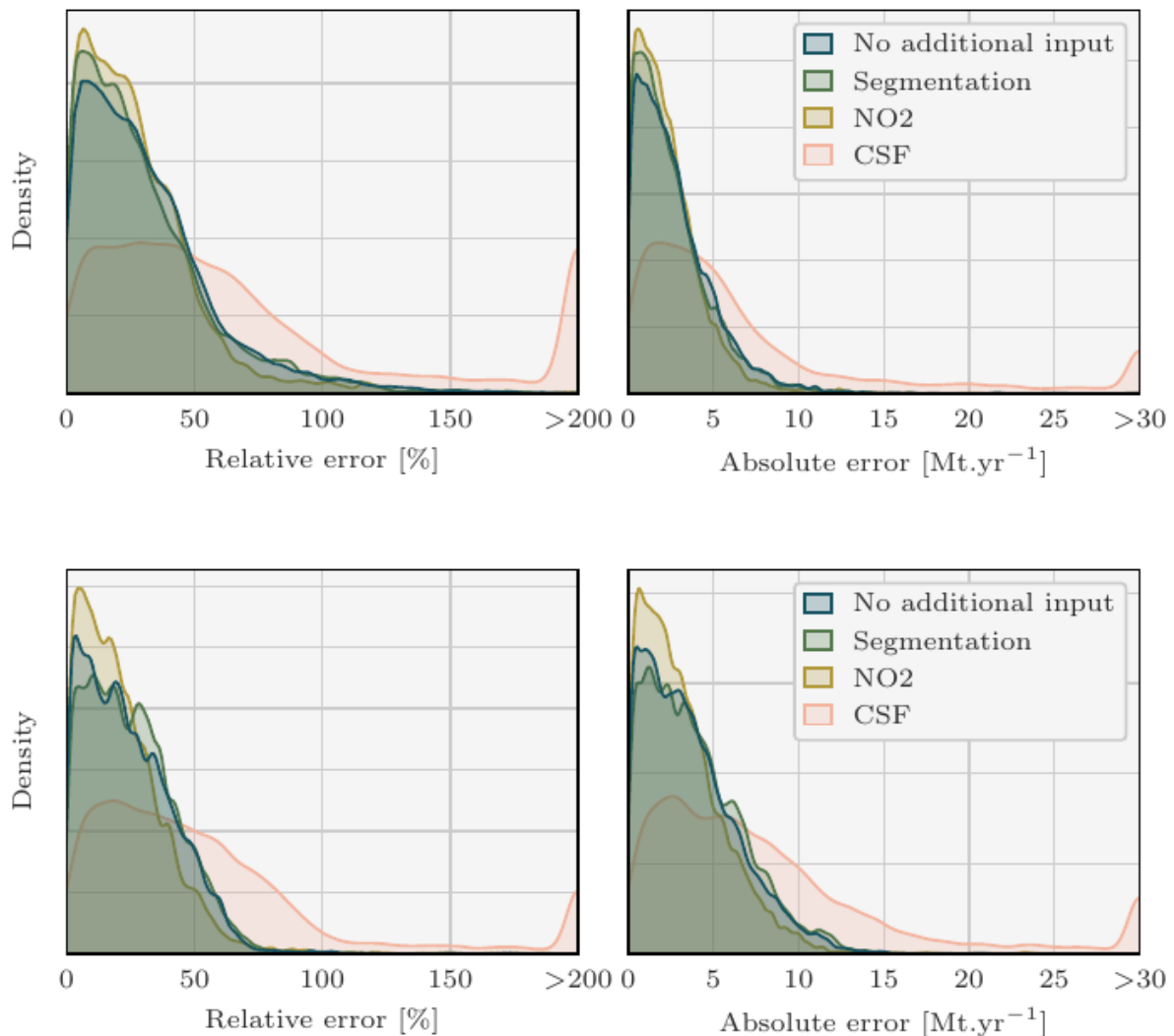
Figure 2: density plots of the absolute value of errors (relative to the true emissions of not) between the predicted and the true Lippendorf (top) and Turow (bottom) instant emissions (corresponding to satellite overpasses). Four sets of predictions are considered, corresponding to three CNN models with three different sets of inputs and the CSF method. Each CNN model is trained with the XCO$_2$ field and the winds field re-analysis as inputs. Two of the models additionally assimilate the NO$_2$ field or the predictions of the CNN-based segmentation model (for the plume detection). Predictions with relative errors greater (in absolute value) than 150% or absolute values of errors greater than 30 Mt/yr were set to 150% or 30 Mt/yr to increase the visibility in the figure. Figure from Dumont Le Brazidec et al., 2023b.

In a more general way, there is a **need to conduct tests in more realistic conditions to better characterize the skill, the cost, and the required level of complexity of the approaches based on local scale transport models**. This applies to the initial application of the CCFFDAS which is made, here, assuming that residual (after correction of biases, e.g. though rotation of plumes) transport modelling errors can be summarized as a random Gaussian noise without spatial or temporal structures (i.e. in a way that does not really challenge the local transport model-based inversions), to explore the full potential of XCO$_2$ observations. However, overcoming the actual transport model errors can require complex approaches as shown by the application of CNNs, and these approaches may somewhat reduce the constraint on the sectoral and spatial distribution of the emissions within cities. On the other hand, **the CCFFDAS approach has the potential to combine the atmospheric constraints from CO2M with** those from additional observational data streams on both fossil

and natural emissions, including **SIF from CO2M**. This extra potential has yet to be explored. It lies in the simultaneous constraint on natural fluxes (through SIF) on the one hand and on the combined fossil and natural atmospheric signal (through XCO$_2$) on the other hand from the same platform. This simultaneous observational constraint appears a promising approach to solve for the fossil emissions. This applies not only to the city-scale examined here. It is equally relevant at the global scale, where it could be explored with the system of Kaminski et al. (2022b). Furthermore, the CCFFDAS could also be operated at national to continental scale and enrich the ensemble of national-scale inversions operated in Task 4.4.

Furthermore, the analysis of the ratios between the local enhancements of CO$_2$ and CO over emission hotspots revealed the limited amount of spatial and temporal overlapping between CO and CO$_2$ observations over such hotspots (at least in Europe) when considering the existing datasets (here that of TROPOMI and OCO-2), and the high temporal and spatial variability of these ratios, and thus, implicitly, **the large variations and uncertainties in the underlying CO and CO$_2$ emission ratios**. The first problem should be overcome in the future by the availability of co-registered images of NO$_2$ and XCO$_2$ from CO2M. However, in the CCFFDAS OSSEs, the ability to infer some information about the spatial and sectoral distribution of the emissions within the city based on the co-assimilation of the pseudo NO$_2$ and XCO$_2$ images from CO2M was explored for a limited set of alternative assumptions regarding the uncertainties in the NO$_x$/CO$_2$ emissions ratios. The ability to exploit the NO$_2$ co-assimilation with the CCFFDAS should probably be assessed with further scenarios of emission ratio uncertainties.

3.4   Confronting the conclusions from tasks T4.1-3 on local scale inversions

Tasks 4.1 to 4.3 addressed the challenge of quantifying emissions from industrial point sources and cities using XCO$_2$ images from different angles. T4.1 evaluated how well the plumes from such sources can be simulated and how models should be set up to properly represent such plumes. T4.2 developed and benchmarked data-driven emission quantification methods that do not require expensive atmospheric transport simulations as in T4.1. T4.3, finally, investigated novel or improved emission plume inversion approaches in connection with transport simulations such as in T4.1.

The evaluation of models in T4.1 showed that atmospheric transport models are able to represent XCO$_2$ plumes as will be observed by CO2M in a very realistic way if they are run at a resolution of about 2 km or better. Typically, **the models simulate correctly the characteristic statistical properties** such as mean width and amplitude as a function of distance from the sources. However, the analysis of the simulations from T4.1 shows that **the plumes cannot be reproduced perfectly in space and time** due to the stochastic nature of turbulence and also due to uncertainties in the local wind.

The benchmarking of lightweight plume inversion methods performed in T4.2 largely relied on synthetic CO2M observations produced with one of the models tested in T4.1, the COSMO-GHG model operated at 1 km resolution. Knowing that the simulated plumes are a good representation of reality means that **the assessment of the methods in T4.2 is adequate** and relevant. However, the evaluation of the COSMO-GHG model in T4.1 indicated that this model is somewhat more dispersive than others and also more dispersive compared to the observations. As a result, real plumes will likely be more compact and more pronounced and therefore more easily detectable. The results of the benchmarking exercise in T4.2 may thus be a bit too pessimistic: for a given quality threshold (based on the diagnostic of uncertainties), more plumes might be detectable in reality and the uncertainty in the emission estimates per case might be a bit smaller. On the other hand, the generation of synthetic CO2M observations was based on very simple assumptions about the observation error characteristics, which might have led to somewhat too optimistic results potentially compensating the previous, pessimistic aspect.

In a more general way, due to their realism, simulations such as that of T4.1 are highly useful in the context of Observing System Simulation Experiments (OSSEs), where the potential of CO2M (or other) observations for estimating emissions is tested. In order to represent the uncertainty in atmospheric transport, however, it is advised to apply ensemble methods, i.e., ensembles of simulations with perturbed meteorological initial and boundary conditions. The synthetic observations could then be generated from one ensemble member and the inversion methods could be tested with the remaining members.

The fact that plumes can be simulated realistically but not with the same turbulent structure as the real plumes and not necessarily at the right location due to errors in local wind direction, has important implications for inversion approaches as applied in T4.3. This **confirms the assumption made in the rationale for the use of CNNs trained with model simulations for the plume inversion** in this task. This also confirms that t**he current assumptions on the model error made for the application of the city-scale CCFFDAS is too optimistic**. Due to unavoidable mismatches between simulated and real plumes, classical inversion methods (including the current version of the CCFFDAS) minimizing the differences between a single forward model run and the observations are not suitable. We recommend to explore the treatment of the results of such model error. One possibility to reduce spatial mismatches is to rotate the simulated plumes, but a likely better approach is also to apply a non-local metric as proposed by Vanderbecken et al. (2023), which avoids the critical double-penalty problem of traditional approaches. Alternatively, model simulations (as in T4.1) were used to train the CNN-based inversion developed in T4.3. Again, the fact that such model simulations can be considered to represent reality quite accurately means that such a training is meaningful.

As already introduced in section 3.3, Dumont Le Brazidec et al., (2023b) compared the CSF method from T4.2 and the CNN method from T4.3 for the estimation of emissions from power plants. As reported in Figure 2, the CNN approach yields considerably more accurate predictions of power plant emissions compared to the CSF method. For Lippendorf, for example, the CNN model shows a median absolute value of the relative error of approximately 20 % and a median absolute error of around 3 Mt.yr^{-1} (the average emissions for Lippendorf are 15.2 Mt.yr-1). In comparison, the CSF method exhibits a higher median absolute value of the relative error of around 40 %, and the absolute value of the error is approximately double, at 6 Mt.yr-1. Conclusions are similar for the two other power plants studied in Dumont Le Brazidec et al. (2023b). **Given an appropriate dataset and training, and when considering full XCO$_2$ images, the CNN method is likely to outperform the CSF and other light plume inversion methods** in terms of accuracy and efficiency on all types of hotspots and scenarios.

However, current CNN models, trained and tested exclusively on full XCO$_2$ images (without clouds) from the SMARTCARB dataset over East German power plants, may have significant biases in retrieving the emissions of hotspots and cities in other geographical locations. A too limited training not sufficiently representing the plumes in other meteorological situations and other locations of the world (e.g., coastal cities, cities in more variable topography, etc.) can be a critical issue for the approach. Handling the gaps in images due to cloud cover and transferring the CNN model trained on simulated data to real satellite observations will also not be trivial steps. The differences in the characteristics of the measurement uncertainties between reality and those assumed in the synthetically produced data could raise critical issues. The training with simulated observations is a promising way forward, but it should be extended to increasing challenging problems (e.g., overlapping plumes, cloud coverage, other geographical and meteorological settings).

Given the open questions regarding appropriate training of CNNs, **the light plume detection and inversion techniques analysed in Task T4.2 keep on being the most suitable approaches to process images for a large number of target sources (point sources and cities) in the near term**. For operational applications, we thus currently recommend the use of these light techniques. However, **the CNN-based approach could quickly reach a good level of maturity and should keep on being developed and extensively tested**.

3.5 Recommendations regarding the development and integration of a specific module to tackle emission hot spots in the multi-scale inversion prototype

The findings from the previous sections stress the challenge of simulating the plumes and inverting the emissions from local sources in the prototype of the CO2MVS multi-scale inversion system under development, which relies on the IFS transport modelling at 9 km resolution. First, the spatial resolution of the IFS is too coarse to properly model individual plumes as indicated in section 3.1. Second, the inversions in the multi-scale inversion prototype have to rely on the comparisons between the model simulations and the observations at the observation time (since it is the basis for tackling the large scales simultaneously), which, as discussed in section 3.4, may be highly impacted by the transport modelling errors at local scale.
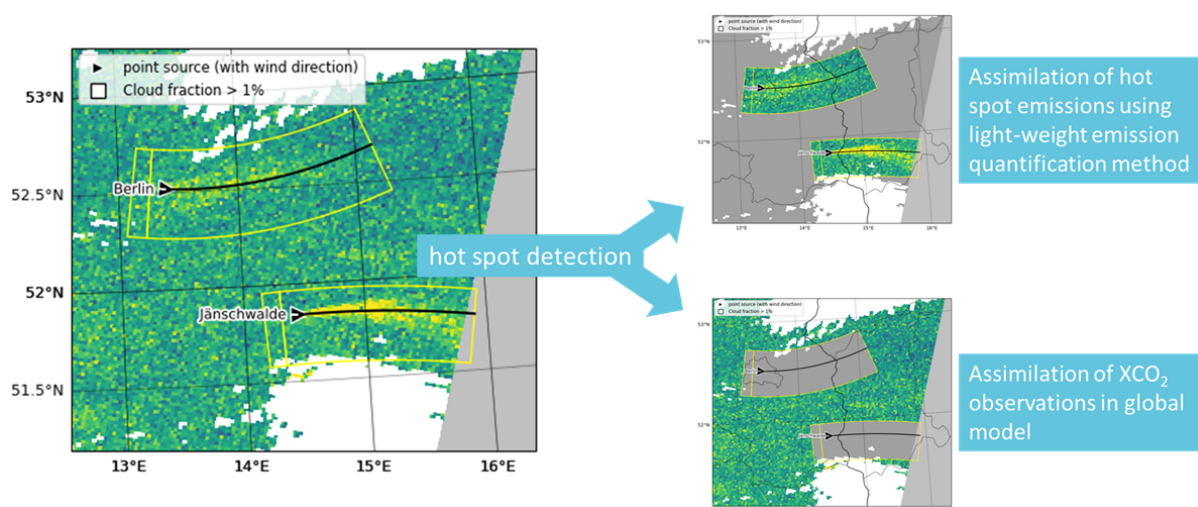

Figure 3: schematic of a potential strategy to split the operational process of the satellite XCO$_2$ observations between a specific branch dedicated to local inversions of plumes, and the main multi-scale global inversion framework

Therefore, in the near term, there will be a need for **a distinct module (a specific branch of the operational system) handling the local inversions based on XCO$_2$ spaceborne imagery**, with the identification of signals associated with plumes from individual sources or emission hotspots, and their processing via plume inversion approaches. As indicated in section 3.4, such a module would probably have to rely on the light plume detection and inversion methods investigated in T4.2, and maybe on CNN-based inversions if significant progress is made quickly in testing these approaches on complex cases. In order to connect or to impose the results obtained with such a module to the multi-scale inversion estimates, the two branches would have to be coupled (see section 5). The branch for local scale inversions could focus on instant estimates for local emissions based on individual images, leaving the temporal and spatial extrapolation of these estimates to the multi-scale inversion branch. Alternatively, the local scale inversion branch could exploit a priori information from inventories to deliver full instant to annual estimates of the emissions from urban and point sources. A partitioning of the XCO$_2$ satellites images between the two branches as illustrated in Figure 3 may be the simplest solution to ensure that the same observations are not assimilated twice when coupling these two branches. But advanced coupling techniques (such as the one detailed in section 5.2) may allow to exploit the full XCO$_2$ images in the multi-scale inversion branch (see section 4.1), even when part of these images are already exploited in the local scale inversion branch.

The **use of standard tools** such as the ddeq Python library developed in the frame of T4.2 should facilitate the development of a branch dedicated to local scale inversions and its coupling to the multi-scale inversion prototype. The ddeq library already provides a high

degree of generalization and all light plume inversion methods could be applied to upcoming XCO$_2$ satellite images to quantify the emissions for a provided list of source locations at the global scale. This requires only that the data input is provided in the ddeq's data format. The ddeq library can also be applied to other trace gases (CH$_4$, CO and NO$_2$), but additional research is required to accurately account for the NO$_x$ chemistry inside plumes when analysing NO$_2$ images. Additional development in ddeq would still be necessary to improve the use of a priori information from bottom-up inventories and for implementing robust quality controls, i.e., estimating the uncertainties in a more accurate way and filtering problematic inversion cases reliably (e.g., under low wind speed conditions of when analysing overlapping plumes).

The **computational costs** for the light-weight approaches are very low (see section 3.2): the computation time for estimating CO$_2$ emissions of 16 sources using one month of cloud-free data from three CO2M satellites varied between 30-120 min depending on the method. We expect that CO2M will be able to quantify the emissions of up to 1000 hot spots globally (Kuhlmann et al., 2021), and the computational costs for 1000 sources would be only of 30-125 hours per month. This highlights the feasibility and relatively low cost of a specific operational branch dedicated to the XCO$_2$ plume inversions.

On the longer-run, if the spatial resolution of the IFS can be refined to 2-km or less, the multi-scale inversion prototype may play a stronger role for the local inversions: to train the CNN-based inversion approaches, or via large local ensemble approaches and the use of more suitable metrics for the model-observation comparisons (Vanderbecken et al., 2023) that could allow for a sufficient account of the transport model uncertainties at local scale (see section 3.4). A complex definition of the observation vector could be required to split the observation into the part that would feed suitably the local scale constraint, and that kept for the large-scale constraint on the flux estimates.

4 Synthesis from the national scale activities in T4.4

4.1 Summary of the conclusions from the inter-comparison of national scale inversions (T4.4)

In Task T4.4, 11 national scale inversions systems were developed and 10 of them were applied for the estimation of the CO$_2$ and CH$_4$ emissions from European countries (and to conduct tests with synthetic data over the USA). Deliverable D4.6 details these inversion systems and the results from these inversions. This **major ensemble of inversions** was complemented by the derivation of the national scale CO$_2$ anthropogenic emissions based on European scale NO$_x$ and CO inversions, as part of WP6 and of the CoCO2 report to GST (see deliverables D6.4, D6.5 and D6.6). The 11 systems developed in T4.4 control separately the CO$_2$ or CH$_4$ anthropogenic and natural fluxes. Their relatively high spatial resolution ranges from 0.5° resolution down to 10 km (for the control of the fluxes) or to 5 km (for the transport modelling), with both Eulerian and Lagrangian transport models, and variational, ensemble and analytical inversion approaches. Three of the systems assimilate surface and/or satellite CO$_2$ observations, the others assimilating surface observations only. The aim of this major inverse modelling effort was to document the current strategies and capabilities to monitor the CO$_2$ (and, to a lesser extent, CH$_4$) fluxes at national scale, and, in particular, to monitor the CO$_2$ anthropogenic emissions. It also aimed to provide support and guidance for the design of national operational systems and of the regional scale configuration of the multi-scale inversion system for the CO2MVS.

Relatively novel regional inversion capabilities were developed and tested to meet the objectives of T4.4 but the results demonstrate the lack of maturity of some of the corresponding components: mainly the separate control of the CO$_2$ anthropogenic emissions, the co-assimilation of co-emitted species, the co-assimilation of surface and satellite CO$_2$ observations, and the underlying characterization at fine resolution of the uncertainties in the

inventories of the anthropogenic emissions of CO$_2$ and co-emitted species used as prior estimate of the inversions.

The major result from this set of inversions is **the lack of control of the CO$_2$ anthropogenic emissions at the annual to monthly and national scales when using the existing CO$_2$ observations**. The co-assimilation of CO and NO$_2$ data does not increase this control significantly. In parallel, there is a large spread of the estimates of CO$_2$ natural fluxes (and of the CH$_4$ anthropogenic and natural fluxes) across the different systems or when assimilating surface or satellite observations.

Based on the analysis of the results, deliverable D4.6 identifies the lack of control of the anthropogenic emissions at large scales as a lack of ability to connect large scale variations in the CO$_2$ fields (dominated by the signal from the CO$_2$ natural fluxes) to the anthropogenic emissions. Therefore, D4.6 promotes the use of spatial resolution finer than 10 km for the transport modelling and for the control of the fluxes if targeting the CO$_2$ anthropogenic emissions, and the assimilation of observation signals associated with specific emissions hotspots: data from ground stations dedicated to specific urban areas, portions of the XCO$_2$ spaceborne observations corresponding to plumes from local sources etc. in addition to the other CO$_2$ observation datasets. There is currently a lack of such observations in the existing observation networks, but the situation should dramatically change once CO2M will be in orbit, and with the promotion of the CO$_2$ urban networks e.g. via the ICOS Cities project. The aim would be to **extrapolate the information on local sources to the larger scales relying on accurate characterizations of the spatial structures of the uncertainties in the inventories** used as prior information of the national scale inversions. This, in turn, would require a dramatic improvement of the knowledge of these uncertainties, and of the ability to model them appropriately.

D4.6 also promotes **a systematic analysis of the inverse modelling components that are responsible for the spread of the results for the CO$_2$ natural fluxes and CH$_4$ emissions** across the inversions. As illustrated in D5.3, modular inversion platforms such as the Community Inversion Framework, allowing a consistent use of several transport models (seven so far, covering most of the transport models used in T4.4), with a large range of inversion set-ups, would be a natural tool to use in the future for attributing the spread of the results, assessing errors in inversions and benchmarking systems and applications.

4.2 Perspectives and recommendations regarding the configuration of the main component of the multi-scale inversion prototype

The main outcome of the analysis in T4.4, i.e. the lack of control of the anthropogenic CO$_2$ emissions by the inversions using the existing CO$_2$ and co-emitted observations at annual to monthly/national scale and the high uncertainties in the CO$_2$ terrestrial ecosystem fluxes limits the potential to use these results to promote the development of national-scale operational systems. However, the systems developed here lay **the technical basis for the development of such systems**. For example, the ICON-ART system developed by DWD corresponds to the system which will be used for the German operational GHG monitoring which is now under development (and the situation could be similar in France with the CIF-CHIMERE inversion configuration).

Furthermore, the detailed results of the different national-scale inversion systems in T4.4 bring insights for **the configuration of the multi-scale inversion systems of the CO2MVS**. In particular, it highlights the need for this system to rely on the capability of the IFS to model the atmospheric transport at spatial resolutions finer than 10 km but also to control the fluxes at spatial resolutions finer than 10 km to tackle the CO$_2$ anthropogenic emissions even when targeting national-scale budgets. Such a high-resolution configuration must be accompanied by the assimilation of data from observation networks dedicated to specific emission hotspots, and of the local scale information in satellite data (typically the information on CO$_2$ plumes

downwind of local sources or groups of local sources) in addition to the observations dedicated to the characterization of large-scale variations in the CO$_2$ concentrations. An alternative is to leave the full control of the anthropogenic emissions to local scale inversion (focused on the plumes for the local sources) with a worldwide-coverage, in the frame of the coupling between the multi-scale CO2MVS system and a specific branch for such local scale inversion (see section 3.5).

In both cases, currently, the lack of observations dedicated to the CO$_2$ anthropogenic emissions, and in particular to specific emission hotspots, will maintain a strong limitation for the monitoring of the CO$_2$ anthropogenic emissions. The situation will change once the CO2M data will become available, but these conclusions also support the deployment of ground stations better focused on the anthropogenic emissions than the typical stations e.g., of the ICOS network in Europe.

The results from T4.4 also stress the need to progress on the use of co-emitted species, in particular to decrease the uncertainty in the top-down information from co-emitted species (in both the observations and the chemistry transport modelling), and to progress on the characterization of the uncertainties in the gridded inventories of the emissions of CO$_2$ and co-emitted species at fine spatial and temporal resolution. Such an effort has been conducted in WP2 of this project and it is now resumed in the frame of the Horizon Europe CORSO project.

5 Coupling the systems and scales

5.1 Motivation

The initial motivation for discussing the coupling between national and local scale inversions in WP4 was the **ability to control the CO$_2$ anthropogenic emissions at both scales, and the reconciliation of the corresponding estimates**, the national scale inversion catching large scale gradients due to these emissions, and the local scale inversion tackling plumes from individual sources or clusters of sources (emission hotspots).

By construction, the local scale inversion approaches analysed in WP4 were focused on the anthropogenic emissions. The signal from local natural fluxes may not be large enough to be detected and processed appropriately with the current observation and inversion systems. However, in WP4, the local scale inversion approach was mainly tested on pseudo images covering very large local sources of CO$_2$. When scanning in an automatic way the existing satellite observations from OCO-2/3 over more than 8-years using the LCSF method assessed in T4.2, few plume transects are identified over Europe (see D4.4 and WP6 D6.4, D6.5 and D6.6). This limits the ability to assess the potential of ingesting the information from the corresponding plume inversions into national scale inversions. Further, the analysis of the national scale inversion results revealed the lack of control of the CO$_2$ anthropogenic emissions at large scales when using the current observation networks (see section 4.1). Therefore, the initial objective of reconciling the national and local scale estimates of the CO anthropogenic emissions when using the existing CO$_2$ data had to be reconsidered in T4.5.

The coupling between the national and local scale inversions from WP4 would rather correspond to t**he coupling between the estimate of the CO$_2$ anthropogenic emissions at local scale and the estimate of the CO$_2$ natural fluxes at regional scale**. As demonstrated in T4.4, the national scale inversion of the CO$_2$ natural fluxes is sensitive to the estimate of the CO$_2$ anthropogenic emissions despite the lack of control of these emissions. On the other hand, the fine scale variations of the background concentration field behind the plumes downwind to the local sources has been shown to be a significant source of uncertainty for local scale inversions when using both light techniques such as in T4.2 or techniques relying on atmospheric transport models such as in T4.3. However, the national scale inversion may hardly provide a reliable simulation of the local variations associated to the natural fluxes. The diagnostic of the background concentrations in the satellite images in the frame of the light

plume detection and inversion techniques may reveal to be more robust than the information from large scale inversions. A mix of both sources of information may however help tackle situations for which the loss of data in the satellite images due to cloud cover or high aerosol loads prevent from getting a good data-driven characterization of the background concentration field.

Furthermore, as discussed in section 4, the increase of the resolution of the national scale inversions and the future availability of the observation from CO2M (and from other observation networks or platforms more sensitive to the anthropogenic emissions than the current network) may lay down a new basis for considering the control of the CO_2 anthropogenic emissions at different scale, and thus for a coupling between local and national scale inversions.

Finally, section 3.5 discusses the specific coupling between the CO2MVS multi-scale inversion system and an operational branch dedicated to local scale inversion. More generally, the perspective to **couple local to national scale inversions to the multi scale inversion prototype** has been discussed in the CoCO2 project in view to exploit in the global CO2MVS system the information from systems covering limited areas with optimal configuration locally or regionally. The option currently envisaged for such coupling consist in the assimilation of the results from local and regional scale inversions into the multi-scale inversion system. Specific investigations have been conducted to develop such an option as part of WP5 and WP6 in connection with WP4. In the frame of deliverable D4.7, ensembles of local scale and national scale inversions estimates have been provided by WP4 to WP6 to test this option.

5.2   The coupling of systems to the multi-scale inversion prototype via data assimilation techniques

The corresponding concept is to assimilate the emission estimates from the high-resolution regional inversion system and from the light local plume inversion techniques as observations into the global multi-scale inversion system based on the Integrated Forecasting System (IFS). The approach, described in more details in the deliverable D5.1, requires two steps:

1. The derivation of the averaging kernel matrix and of the associated retrieval errors from the emission inversion methodology using ensembles from a Monte Carlo approach with the local and national scale inversions.

2. The integration of the local and regional emission estimates into the global IFS by merging the IFS Ensemble of Data Assimilation (EDA) output with an ensemble Kalman filter approach.

In that process, it is important to distinguish between several categories of products depending on the regional and local scale inversions:

1. **For regional inversions based on Ensemble Kalman Filters**: the required ensemble information is a by-product of the system.

2. **For regional inversions based on Variational inversion systems**: the ensemble can be produced based on an EDA approach.

3. **For regional inversions based on Analytical inversion systems**: the low-dimensional averaging kernel matrices and associated retrieval errors can be provided explicitly.

4. **For non-Bayesian local plume inversions**: the averaging kernel matrix used is a Dirac function. The error matrix **E** is obtained using uncertainty estimation techniques tailored to the local inversion method.

Figure 4 depicts the layered strategy employed to merge data from regional and local inversion analyses into the comprehensive global IFS inversion process, and further, to mimic a two-way nesting between these systems via the provision of boundary conditions from the IFS to the local and regional scale inversions system.
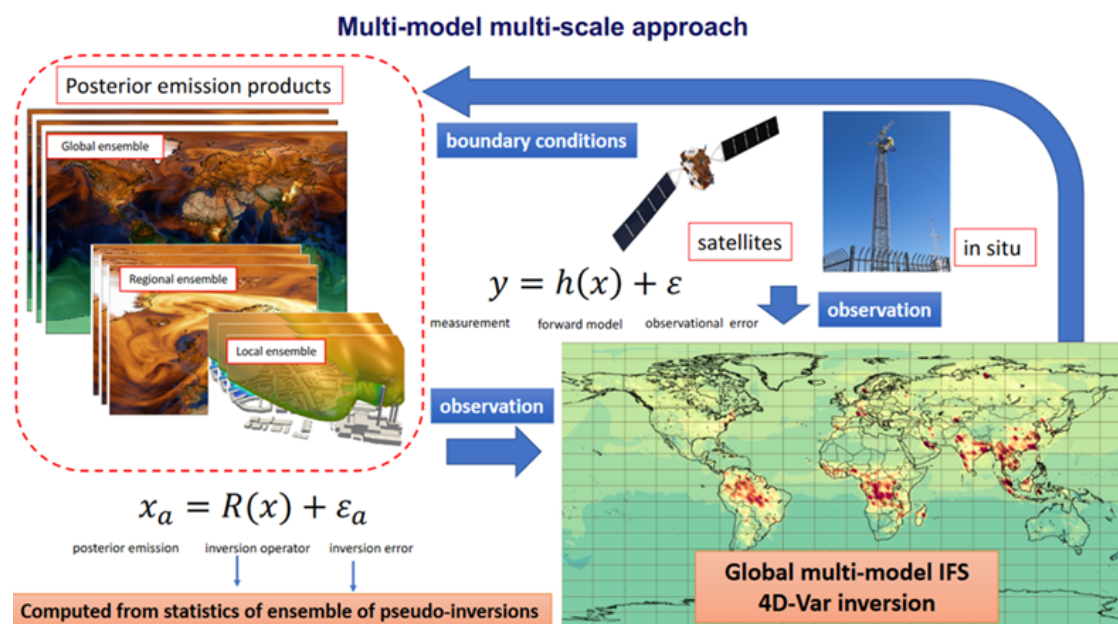

Figure 4: Schematic of the multi-scale approach for the integrated global IFS inversion system. The IFS global inversion system provides boundary conditions to the regional and local inversion systems. The high-resolution posterior emissions obtained from those systems are in turn assimilated as observations into the global IFS model, enabling a two-way propagation of information.

In an operational context, the EDA or posterior IFS ensemble will be produced routinely and available at any time. The inherent flexibility of the proposed ensemble assimilation approach for the regional and local emission product will make possible their integration in the global IFS system in a continuous and seamless manner.

6 Benchmarking strategies: lessons from the local and national scale activities

The conclusions from T4.1-3 reveal that the benchmarking test cases, and in particular those based on the SMARTCARB simulations, were extremely useful **to develop and optimize the local scale methods, and to provide a general ranking among them**. It can clearly be used to assess new methods or local scale inversion configurations.

However, the conclusions from T4.1-3 also highlighted the fact that **these benchmarking test cases were relatively simple and did not fully challenge the methods**. For example, the differences between the COSMO and ERA-5 wind fields was used to characterize the uncertainty in the wind fields, which was probably not realistic enough. Furthermore, the focus on large and relatively isolated power plants and cities over the flat part of Eastern Germany and far from the coast in T4.2 and T4.3 provided favourable conditions for the method evaluation. The methods were not challenged with a realistic representation of the systematic error in the satellite XCO_2 observations. The simulation of the errors on these observations was restricted to Gaussian white noise, albeit with realistic statistics using error parametrization depending on solar zenith angle and surface albedo. Furthermore, as

mentioned in sections 3.1 and 3.4, the COSMO-GHG simulated fields in the SMARTCARB benchmarking dataset were likely more dispersed compared to the MicroHH simulated fields, while the latter fields showed a better match to in-situ observations. Hence, the COSMO-GHG data has slightly wider plumes of lower amplitude than what one might observe in real data.

Light-weight approaches (including machine-learning methods if those methods definitely have high extrapolation skills) are likely necessary to monitor the emissions of hot spots in an operational system due to their low computational costs. So far, the methods have mainly been developed, trained and validated with synthetic observations. How well the methods will perform with real satellite observations is therefore not fully assessed. To obtain reliable emission estimates with lightweight methods, it is therefore necessary to evaluate their performance with real observations as soon as they become available. One option is a **validation with measurements of point sources with well-known emissions (e.g. instrumented power plants) in measurement campaigns**. For individual sources, comparison with emission estimates from inversions using high-resolution models is also an option. It would be desirable to conduct regular validation campaigns to cover different source types, meteorological conditions, background chemistry, seasons and locations.

The **common open source *ddeq* Python library environment** that was used for the benchmarking of data-driven local scale emission quantification methods turned out to be very useful. The common benchmarking environment allowed direct and transparent comparison of the methods. The open-source environment is also flexible for optimizing further the current methods (including the pre-processing methods) or implementing new approaches.

In task T4.2, as well as in all benchmarking activities, **the choice of the metrics used for comparing the methods** is a crucial step. We chose to focus on the difference between estimated emissions and the true emissions. We studied both their distributions and number of 'successful' estimates (i.e., which passed selected thresholds on the diagnostic of uncertainty) as well as goodness in estimating monthly and yearly emissions. Results were analysed considering also the emission strength. Overall, the selected metrics led to satisfying analysis and ranking. However, we faced some challenges in implementing thresholds for 'successful' estimates since the diagnostic of the uncertainties (as a natural choice of quality indicator) of some methods was not realistic, or focused too much on a specific source of uncertainty. As mentioned in section 3.2, further work is needed to improve the uncertainty characterization but meanwhile, other ways of implementing practical thresholds on quality indicators may be needed.

Overall, the benchmarking activities and the shared *ddeq* environment on the ICOS-CP Jupyter Server were found very useful in CoCO2 Task T4.2, with the following advantages:

- it revealed bugs in implementing methods
- it supported the optimization of the methods
- it allowed sharing pre-processing steps
- it allowed transparent and direct comparisons
- it allowed comparing computational times in a rigorous way
- it allows flexibly extensions to test and benchmark: the most important need of extension is now the inclusion of synthetic XCO₂, NO₂ but also aerosol CO2M observations from pole-to-pole, including realistic simulations of the systematic errors on these observation retrievals
- it allows further optimization of the methods and implementation of new methods.

For national-scale inversions, the inter-comparison protocol was relatively loose, due to the high number of systems to be developed and tested, and in many cases, due to the lack of ability to impose general constraints for the characterization of the uncertainties in the prior estimates of the fluxes and of the regional boundary conditions, for the transport model and observation errors, etc. This stems for the diversity of the types of transport models, control and observation vectors, of the process of the boundary conditions etc. across the systems.

Many of the inversion components are hard-coded in the traditional inversion systems, so that they can hardly be adapted to common features. However, the results revealed the **need to investigate systematically and in a standardized way the weight of each inversion component in the spread of the results** for the CO$_2$ natural fluxes and CH$_4$ emissions. The **joint use of community and modular platforms** handling the variety of inversion approaches, of control vectors, of assumptions regarding the prior, observation and model error statistics, and ultimately, the variety of (Lagrangian and Eulerien) transport models used, such as the Community Inversion Framework (CIF) should support such a detailed analysis, as shown by the first exercises in WP5: T5.3 and T5.6. This would support a more exhaustive inter-comparison and analysis, to identify the best options for the inversion components, and thus standard and proven methods. The joint use of community tools would also increase the transparency of the inversion protocol and confidence in the results.

7 Conclusion

The synthesis of the analysis in WP4 provides

- conclusions regarding the spatial resolutions that are required for the transport modelling and control of the fluxes for both local and regional scale inversions of the CO$_2$ anthropogenic emissions
- a detailed assessment of the local scale inversion techniques to process the XCO$_2$ space-borne images of plumes downwind to CO$_2$ emission hotspots
- an incentive to develop a specific branch dedicated to local scale inversion of plumes for the operational process of the future XCO$_2$ imagery, in the near-term, based on light plume detection and inversion techniques
- an encouragement to keep on developing and testing plume inversion techniques based on machine learning
- a motivation to couple machine learning with city scale inversion frameworks (such as CCFFDAS) to bypass the atmospheric transport modelling errors at the observation times
- indications of a need to increase the ability to co-assimilate observations of co-emitted species (CO, NO$_2$), mainly via a better handling of the top-down information from these observations and from the chemistry-transport models, and a better characterization of the correlation of uncertainties in the inventories for CO$_2$ and co-emitted species at fine spatial and temporal scales
- an incentive to explore the potential in the simultaneous constraint on natural fluxes (through SIF) on the one hand and on the combined fossil and natural atmospheric signal (through XCO$_2$) on the other hand by measurements from the same platform (CO2M) in a CCFFDAS operated at scales from city to national/continental (enriching the ensemble of national scale systems operated in Task 4.4) to global
- discussions on the rationale and frameworks for the coupling between local and regional scale inversions, and for the coupling of local and regional inversions to the multi-scale IFS inversion system of the future CO2MVS
- a positive assessment of the benchmarking protocols used at local scale, with recommendations to upgrade their level of complexity (up to that of the real observation) and extend their geographical coverage
- insights on how to tighten the inter-comparison of national scale inversions to identify the inversion components that are responsible for a large part of the uncertainties in the estimate of the CO$_2$ natural fluxes and CH$_4$ emissions
- a promotion of open source, modular and community codes for local and national scale inversions such as those that have been used in WP4 and WP5 to strengthen the transparency, reliability, robustness and operativeness of the inversions, the ability to

identify the main source of uncertainties, the ability to improve the system configurations and the establishment of standard proven approaches

Overall, the initial objectives of WP4 were ambitious and were not completely fulfilled. In particular, we aimed to provide guidance on the detailed configuration of the national / regional inversions for the large-scale inversion of the CO$_2$ anthropogenic emissions, owing the large number of models developed in the frame of T4.4. However, the large-scale inversion of the CO$_2$ anthropogenic emissions using the current observation networks was a relatively novel development for CO$_2$ regional-scale inversions, and was limited by the current lack of CO$_2$ observations dedicated to the anthropogenic emissions. That being said, the consistency of the results regarding the CO$_2$ anthropogenic emissions across the different national scale inversions raised general guidance regarding the configuration of the future national systems and for the CO2MVS. The large-scale inversions of the anthropogenic CO$_2$ emissions keep on being exploratory, and may finally have to connect to local scale inversions via the coupling of systems or scales within the multi-scale inversion prototype, or via the gradual increase in the spatial resolution of the national scale systems.

The benchmarking of local scale inversion approaches led to more detailed conclusions, with a ranking and the assessment of the level of accuracy of the different methods depending on the targeted criteria for the monitoring of CO$_2$ emission hotspots. The exploitation of the new generation of high (sub-km) spatial resolution imagers (whose concept is promoted by various private or institutional initiatives) should probably rely on the same techniques as those for the process of the CO2M 2 km resolution images. New benchmarking experiments (with simulations from models such as the LES model used in T4.1) will be required to test and improve the specific configurations for such high-resolution images, but the inter-comparisons and development conducted in WP4 could serve as a basis for such an exercise.

WP4 brought a new list of expectations of upgrade, increase or improvement of the observational systems (complementing CO2M), of the spatial resolution of the transport models and inversions, of the characterization of the uncertainties in the inventories used as prior estimates of the inversions etc. and it let many important questions for local and regional scale inversions open. However, it also brought new and concrete insights for the development of the CO2MVS multi-scale inversion system and, more generally, for the atmospheric monitoring of the CO$_2$ anthropogenic emissions.

8 References

Dumont Le Brazidec, J., Vanderbecken, P., Farchi, A., Bocquet, M., Lian, J., Broquet, G., Kuhlmann, G., Danjou, A., and Lauvaux, T.: Segmentation of XCO2 images with deep learning: application to synthetic plumes from cities and power plants, Geosci. Model Dev., 16, 3997–4016, https://doi.org/10.5194/gmd-16-3997-2023, 2023a.

Dumont Le Brazidec, J., Vanderbecken, P., Farchi, A., Broquet, G., Kuhlmann, G., and Bocquet, M.: Deep learning applied to CO$_2$ power plant emissions quantification using simulated satellite images, Geosci. Model Dev. Discuss. [preprint], https://doi.org/10.5194/gmd-2023-142, in review, 2023b.

Hakkarainen, J., Ialongo, I., Koene, E., Szelag, M. E., Tamminen, J., Kuhlmann, G., and Brunner, D.: Analyzing Local Carbon Dioxide and Nitrogen Oxide Emissions From Space Using the Divergence Method: An Application to the Synthetic SMARTCARB Dataset, Front. Remote Sens., 3, https://doi.org/10.3389/frsen.2022.878731, 2022.

Hakkarainen, J., Ialongo, I., Oda, T., Szelag, M. E., O'Dell, C. W., Eldering, A., and Crisp, D.: Building a bridge: characterizing major anthropogenic point sources in the South African Highveld region using OCO-3 carbon dioxide snapshot area maps and Sentinel-5P/TROPOMI

nitrogen dioxide columns, Environ. Res. Lett., 18, 035003, https://doi.org/10.1088/1748-9326/acb837, 2023a.

Hakkarainen, J., Kuhlmann, G., Koene, E., Santaren, D., Meier, S., Krol, M. C., van Stratum, B. J. H., Ialongo, I., Chevallier, F., Tamminen, J., Brunner, D., and Broquet, G.: Analyzing nitrogen dioxide to nitrogen oxide scaling factors for computationally light satellite-based emission estimation methods: a case study of Matimba/Medupi power stations in South Africa. Submitted to Atmospheric Environment: X, 2023b.

Kaminski T, Scholze M, Rayner P, Houweling S, Voßbeck M, Silver J, Lama S, Buchwitz M, Reuter M, Knorr W, Chen HW, Kuhlmann G, Brunner D, Dellaert S, Denier van der Gon H, Super I, Löscher A and Meijer Y (2022a) Assessing the Impact of Atmospheric CO$_2$ and NO2Measurements From Space on Estimating City-Scale Fossil Fuel CO$_2$ Emissions in a Data Assimilation System. *Front. Remote Sens.*3:887456. doi: 10.3389/frsen.2022.887456

Kaminski, T., Scholze, M., Rayner, P., Voßbeck, M., Buchwitz, M., Reuter, M., et al. (2022b). Assimilation of Atmospheric CO$_2$ Observations from Space Can Support National CO$_2$ Emission Inventories. Environ. Res. Lett. 17, 014015. doi:10.1088/1748-9326/ac3cea

Krol, M., van Stratum, B., Anglou, I., and Folkert Boersma, K.: Estimating NO$_x$ emissions of stack plumes using a high-resolution atmospheric chemistry model and satellite-derived NO2 columns. Submitted to Atmospheric Chemistry and Physics, 2023.

Kuhlmann, G., Broquet, G., Marshall, J., Clément, V., Löscher, A., Meijer, Y., et al. (2019). Detectability of CO$_2$ emission plumes of cities and power plants with the Copernicus Anthropogenic CO2 Monitoring (CO2M) mission. Atmospheric Measurement Techniques 12, 6695–6719. doi:10.5194/amt-12-6695-2019

Kuhlmann G, Henne S, Meijer Y and Brunner D (2021) Quantifying CO$_2$ Emissions of Power Plants With CO$_2$ and NO$_2$ Imaging Satellites. *Front. Remote Sens.* 2:689838. doi: 10.3389/frsen.2021.689838

Kuhlmann, G., Koene, E. F. M., Meier, S., Santaren, D., Broquet, G., Chevallier, F., Hakkarainen, J., Nurmela, J., Amorós, L., Tamminen, J., and Brunner, D.: The ddeq Python library for point source quantification from remote sensing images (Version 1.0). Submitted to Geosci. Model Dev., 2023.

Santaren, D., Hakkarainen, J., Kuhlmann, G., Koene, E., Chevallier, F., Ialongo, I., Lindqvist, H., Nurmela, J., Tamminen, J., Amoros, L., and Broquet, G.: Benchmarking data-driven inversion methods for the estimation of local CO$_2$ emissions from XCO$_2$ and NO$_2$ satellite images. Submitted to Atmospheric Measurement Techniques, 2023.

Vanderbecken, P. J., Dumont Le Brazidec, J., Farchi, A., Bocquet, M., Roustan, Y., Potier, É., and Broquet, G.: Accounting for meteorological biases in simulated plumes using smarter metrics, Atmos. Meas. Tech., 16, 1745–1766, https://doi.org/10.5194/amt-16-1745-2023, 2023.

Document History

Version	Author(s)	Date	Changes
	Name (Organisation)	dd/mm/yyyy	
1	Grégoire Broquet (CEA), Dominik Brunner, Erik Koene, Gerrit Kuhlmann (Empa), Joffrey Dumont Le Brazidec (ENPC), Johanna Tamminen (FMI), Thomas Kaminski (iLab), Nicolas Bousserez (ECMWF), Julia Marshall (DLR), Diego Santaren, Antoine Berchet (CEA) and the WP4 team	22/12/2023	

Internal Review History

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

Estimated Effort Contribution per Partner

Partner	Effort
Organisation	effort in person month
Total	**0**