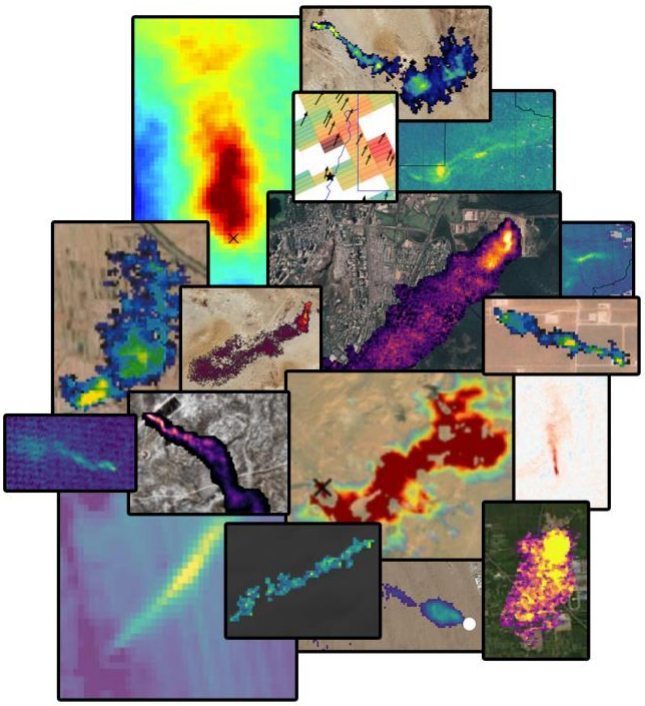
Intercomparison of high-resolution CO_2 satellites

WP900 - Deliverable D9.1 v2.2



ESA Contract No. 4000143908/24/I-LR

Authors

GHGSat
- Antoine Ramier
- Jason McKeever

Kayrros
- Julian Akani Guery
- Maxence Nevoret

SRON Netherlands Institute for Space Research
- Joannes D. Maasakkers
- Xin Zhang

Universitat Politècnica de València
- Luis Guanter
- Javier Roger Juan

Approved

Date: 2026-01-14 Signature: 	Date: Signature: Simon Pinnock Digitally signed by Simon Pinnock Date: 2026.03.30 16:05:46 +01'00'
Ilse Aben (SRON) MEDUSA Science Lead	Simon Pinnock (ESA) Technical Officer

Change log

Version	Date	Status	Authors	Reason for change
0.1	June 6, 2024	Draft	see title page	First version V0 delivered to the project lead.
0.2	June 7, 2024	Version for review by ESA. Accepted on 11 July 2024	see title page	First round of corrections
1.1	March 29, 2025	Version for review by ESA	see title page	First version V1 delivered to the project lead (March 21, 2025). New sections added to V0: section 5 and 6
2.0	6 October 2025	Version for review by the project lead	See title page	First version V2 delivered to the project lead. Added section 7 compared to V1.
2.1	13 October 2025	Version for review by the project lead	See title page	First version V2 delivered to the project lead (v2.0). Added section 7 compared to V1.
2.2	17 December 2025	Version for review by ESA	See title page	Addressing comments and recommendations from ESA

Table of Contents

1 Executive summary

This document reports on work package WP900 of the MEDUSA project, which focuses on performance analysis and intercomparison between high-resolution satellite instruments for detecting carbon dioxide (CO_2) point sources. Three instruments are included in this study: GHGSat-C10 (dedicated CO_2 imaging instrument), EMIT, and EnMAP (general-purpose scientific missions with sensitivity to CO_2). Data from the public missions were analyzed independently by three teams: Kayrros, SRON, and UPV. Performance metrics describing the detection limit and quantification accuracy of each instrument against stack-level measurements were calculated and are summarized in Table 1 below.

Table 1. Detection and quantification performance metrics for each instrument, aggregated over the multiple analysis teams for EMIT and EnMAP.

Instrument	Detection threshold (50% probability)	Regression slope	Regression R^2	Number of observations
EMIT	8.2 MT/yr	0.42	0.72	34
EnMAP	4.4 MT/yr	0.47	0.84	32
GHGSat-C10	18.6 MT/yr	0.84	0.90	33

The main conclusions of this report are:

1. The 3 instruments demonstrate clear potential for satellite-based monitoring of CO_2 emissions, with some overlap and complementarity in current capabilities.
2. EnMAP demonstrated the lowest detection threshold (~4.4 MT/year at 50% probability of detection), followed by EMIT (~8.2 MT/year) and GHGSat-C10 (~18.6 MT/year). The chosen set of sites included significant geographic and topographic challenges, and these estimates are likely conservative, as opposed to the best performance achievable under favourable conditions.
3. Regarding emission quantification, EMIT and EnMAP systematically underestimated emissions compared to ground truth data, while GHGSat was less biased. Limitations of the wind transport model, namely the need to account for source elevation and hot gas buoyancy, are thought to be important drivers of this underestimation, and should be addressed by using more realistic numerical simulations to calibrate the effective wind speed (U_{eff}), and/or using 100-meter wind speed. Additional investigations of the L0 to L2 processing are also recommended, for comparing matched filter outputs with those of inversion methods analogous to those used by GHGSat.

Reporting

The report is delivered in incremental steps as the project progresses:

- Version D9.1.v0 describes the proposed approach. We provide a list of 10 sites that will be observed 3 times by each of the 3 satellites over a 10-month period (a total of 90 observations). The sites are selected based on the availability of stack-level measured emission rates at an hourly resolution, operational instrument limitations (EMIT), and to provide a diverse set of emitter types over a range of geographic and topographic conditions.
- Version D9.1.v1 describes the mathematical methods and algorithms used to retrieve CO_2 concentration and emission rates from raw or intermediate satellite data and gives an overview of the retrieval results.
- Version D9.1.v2 (this deliverable) presents the final analysis of instrument performance, intercomparisons, and recommendations. The key aspects of the analysis include the probability of detection (detection threshold), quantification accuracy, and intercomparisons between instruments.

2 Context

The first greenhouse gas sensing satellites (e.g. GOSAT, OCO-2) had strong emphasis on CO_2 for the purpose of area flux mapping with global coverage. More recently we have seen satellites with higher-resolution imaging capabilities being applied to point source imaging and quantification, primarily for methane – hence the methane focus of the Climate-Space GHG Emissions Project MEDUSA. However, these types of high-resolution systems can also be applied to CO_2 – albeit with different performance levels and use cases. These satellites have the potential to give independent, well-validated emission estimates for industrial-scale CO_2 emitters worldwide.

CO_2 sensing at the facility scale is an emerging, less mature capability compared with methane. Previous work in the literature is limited to some studies with PRISMA [1,2] and a first publication on EMIT [3]. On November 11th, 2023, GHGSat added three new satellites to its constellation including GHGSat-C10 (also known as Vanguard) – the first satellite designed specifically for CO_2 point-source imaging. It is the first of its kind, namely a satellite dedicated to measuring CO_2 emissions at the facility level.

Here we study the performance of GHGSat-C10 and two public hyperspectral satellites with comparable spatial resolution – EnMAP and EMIT.

3 Scope and objective(s)

In line with ESA's request to explore equivalent capabilities (to methane) for monitoring CO_2 plumes, in preparation for future improved instruments (i.e. CO2M), this report studies the novel CO_2 point source imaging capability from a New-Space mission (GHGSat) with the objective to compare it to two research-based missions (EnMAP and EMIT). The latter were not designed to measure CO_2 but have shown they can detect the largest plumes under favorable conditions which makes it therefore interesting to include them in this study.

The objectives of this WP are the following:

- Estimate of CO_2 emission quantification accuracy for all three satellites
- Estimate of CO_2 detection limit for all three satellites

4 Approach and Setup [WP920]

While the main MEDUSA work on methane has a strong focus on controlled release validation, we cannot take this approach for CO_2 – since the relatively high detection limits would require unpractically massive release volumes. Rather, we propose an approach similar to Ref. [2], where we compare L4 emission rates retrieved from satellite data to independently measured and reported ground truth rates from industrial sites like power plants.

4.0 Satellites and team composition

We focus on three satellites for this study: GHGSat-C10, EMIT, and EnMAP. The teams selected for this sub-project consist of the greenhouse gas satellite retrieval and emission experts from GHGSat, Kayrros, SRON, and Universitat Politècnica de València (UPV). Consortium partners processed satellite data according to Table 2 below. As a note regarding previous versions of this deliverable, SRON and UPV were initially listed as “best effort” for some or all of the instruments, but in the end provided full data analyses for both EMIT and EnMAP. UPV was added as full partner in WP900 during the project through CCN-1.

Table 2. List and role of consortium partners.

	GHGSat-C10	EnMAP	EMIT
GHGSat	☑		
Kayrros		☑	☑
SRON		☑	☑
UPV		☑	☑

4.1 Considerations of site selection criteria

Most large CO_2 emitting facilities such as power plants, refineries, and cement plants are inventoried in publicly available databases (e.g. government portals, Climate Trace, etc.), along with corresponding yearly emissions estimated either from direct measurements or bottom-up estimates. The number of suitable sites is larger than can be considered within the scope of this study. A few (n=10) suitable sites are therefore selected based on the criteria listed in the following paragraphs.

4.1.0 Diversity of emitter types and geography

The selected sites should include a variety of emitter types that are representative of the largest contributors to global CO_2 emissions:

- Fossil-fuel power plants, including the 3 main types of fuels (coal, gas, oil)
- Refineries
- Cement plants

Additionally, sites should have some level of geographic diversity, including factors such as:

- Topography (mountains or flat)
- Land cover (agricultural fields, forests, coastal, deserts)
- Variety of countries

4.1.1 Satellite operational limitations

GHGSat and EnMAP are on sun-synchronous polar orbits and must be tasked specifically to target facilities of interest. They impose no major constraint on the choice of sites other than observation capacity, which is sufficient for the requirements of this program.

EMIT is mounted on the International Space Station (ISS) which has an 51.64° inclined orbit (see Fig. 1 below). Consequently, sites at high latitude cannot be observed and are excluded from the study. Additionally, the instrument is not tasked to specific sites and observation opportunities are not predictable over extended periods of time. We favored sites that had existing clear EMIT observations already available. This increases our confidence in future successful observations and provides the fallback option of using archive observations if fewer than 3 observations are taken within the duration of the MEDUSA project.

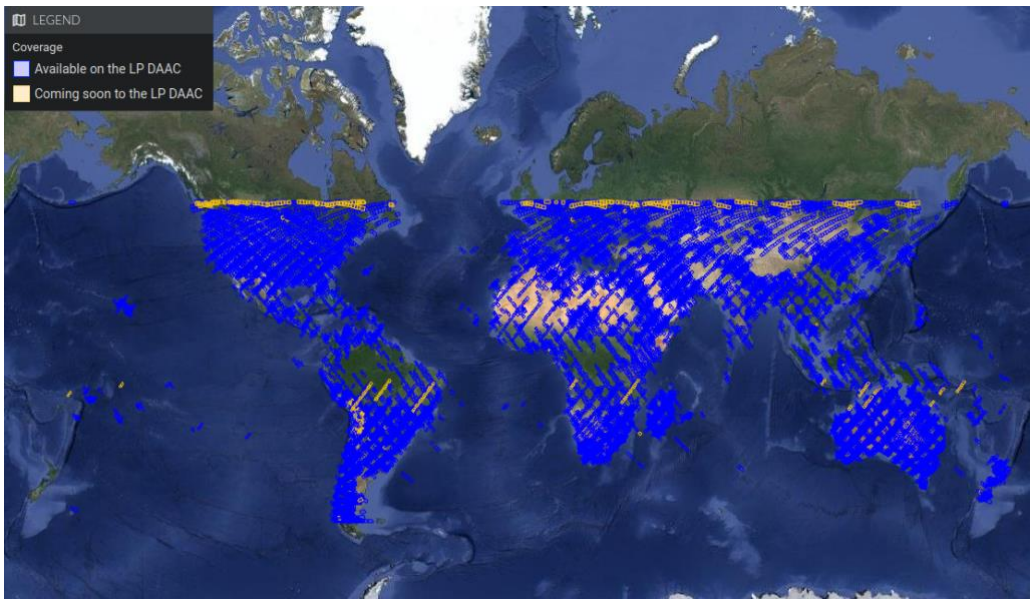


Figure 1. EMIT coverage from https://earth.jpl.nasa.gov/emit/data/data-portal/coverage-and-forecasts/

4.1.2 Range of expected emission rates

The approximate emission rates from the targets should span from below the anticipated detection limit of GHGSat-C10 (~100 t/hr) to well above the detection limit of the hyperspectral satellites. We selected sites with average emissions spanning between around 50 t/hr and 2,000 t/hr. Further variability beyond this range is expected due to hourly and seasonal changes in power production, number of active units, and number of flue stacks among which the emissions are divided.

4.1.3 Choice and availability of ground truth data

Continuous emission monitoring systems (CEMS) are the preferred and most direct source of "ground truth" emissions measurements. Many countries have such systems installed on some or most of their fossil-fueled power plants, including Japan, the US, the European Union (EU), South Korea, China, and Canada [4,5]. However, to the best of our knowledge, this information is not publicly available at the level of granularity required for our analysis (facility-level, hourly sampling), with the notable exception of the United States where the Environment Protection Agency (EPA) publishes the data through the Clean Air Markets Program Data (CAMPD, https://campd.epa.gov). As such, American power plants form the largest class of sites chosen in this study.

In countries where CEMS data is not publicized, an alternative approach proposed in Ref. [4] is to use electric power generation statistics as a proxy. We have selected sites

from Europe and Australia where hourly to sub-hourly power generation data is available from the following sources:

- Europe: European Network of Transmission System Operators for Electricity (ENTSOE). https://transparency.entsoe.eu/
- Eastern Australia: National Electricity Market (NEM). http://nemweb.com.au/Reports/

Finally, for emitters that are not power plants (e.g. refineries, cement plants), intercomparison between different satellite instruments will be the primary cross-validation approach. Publicly available emissions data lack the temporal resolution to be used as ground truth and may only be available within over a year of time lag, but can be used to sanity-check the order of magnitude of retrieved emissions. The European Environment Agency () provides yearly CO_2 emissions data for regulated facilities across Europe which can be accessed through the following link: https://climate.ec.europa.eu/document/download/ebb2c20e-8737-4a73-b6ba-a4b7e78ecc01_en?filename=verified_emissions_2023_en_1.xlsx.

4.2 List of selected sites

Based on the considerations listed above, we selected the following list of sites to be included in this study:

Table 3. Name, location, and properties of selected sites

Site	Location	Type	Emission estimate (Mt/y)*	Terrain	Reference data
Gorazdze Cement Plant	Gorazdze, Poland	Cement Plant	2.5	Flat (forest, farms)	EEA
TotalEnergies Raffinerie Mitteldeutschland	Leuna, Germany	Refinery	1.6	Flat (forest, farms)	EEA
Platte	Grand Island, NE, USA	Coal PP	0.487	Flat (crops, lakes, urban)	EPA (CEMS)
High Desert Power Project	Victorville, CA, USA	Gas PP	1.746	Mix (desert)	EPA (CEMS)
Rockport	Rockport, IN, USA	Oil PP	3.389	Flat (farms)	EPA (CEMS)
Four Corners Steam Elec Station	Fruitland, NM, USA	Coal PP	7.853	Hills (desert)	EPA (CEMS)
James H Miller Jr	Quinton, AL, USA	Coal PP	18.109	Flat (forest)	EPA (CEMS)
Cartagena Escomberas Power Station	Cartagena, Spain	Gas PP + Refinery	1.429	Hills (coast)	ENTSOE (Power)
Stanwell Power Station	Stanwell, Australia	Coal PP	7.291	Hills (grass, bushes)	NEM (Power)
Callide (B, C)	Mount Murchison, Australia	Coal PP	2.049	Mix (mines, lake, bushes, farms)	NEM (Power)

*Source: Climate Trace (https://climatetrace.org/)

4.3 Planned observation schedule

This study aims to collect a total of 90 observations: 10 sites, observed 3 times by each instrument, 3 observations each. The paragraphs below describe tasking and data collection for each instrument.

4.3.0 EMIT

L1B products from EMIT will be downloaded from the Earthdata catalog (https://search.earthdata.nasa.gov/search?q=C2408009906-LPCLOUD). As there is no process to request observations, we used the number of acquired observations since the beginning of the mission (Aug. 2022) as a proxy of the number of images that will be acquired in the future. Based on that number, 10 target sites likely to reach the 3 observation threshold have been chosen. However, EMIT passed into an extended mission phase of operations since early 2024. Thus, the observation plan might be updated accordingly and lead to a lower frequency of acquisition over certain sites

compared to the past. If some sites do not reach the desired number of observations, we will switch to other target sites with sufficient archive images.

4.3.1 EnMAP

To see whether plumes have been detected in the past, we will take inventory of the available archived EnMAP L1 data over the 10 selected sites. Subsequently, we will set targets for the EnMAP proposal. The EnMAP IPS system (https://planning.enmap.org/) can target selected sites following four steps: submitting the proposal, waiting for review, submitting observation requests, and waiting for L1 products. Due to the proposal priority being around level 6, the limitation of tiles for one proposal is approximately 25. We plan to submit three EnMAP proposals to the IPS proposal portal sequentially over a period of five months. Initially, we will draft and submit a proposal for two observations per site within the first two months. The review process typically takes between half a month to one month. To increase the likelihood of targeting the sites with the clear-sky condition, we will set the maximum cloud coverage to 30%. After the two-month proposal period, we will assess whether there is a need to change the targets if no plumes are detected. Subsequently, two more proposals will be submitted, each spanning one and a half months, to continue the targeting process. The tile and cloud-coverage settings will remain the same as in the first proposal. By the end of this process, we aim to obtain at least three new (2024) clear observations for all 10 selected sites. If the number of newly acquired observations is not sufficient, we will add observations from the archive (if available).

4.3.2 GHGSat

GHGSat instruments will be directly tasked to target each of the 10 listed facilities. The revisit time for an individual instrument is 14 days, providing about 20 measurement opportunities (for isolated sites) during the 10-month period allocated for data collection. This provides sufficient margin to account for weather and opportunity conflicts that can arise for nearby sites in Europe or the US.

4.4 Timeline

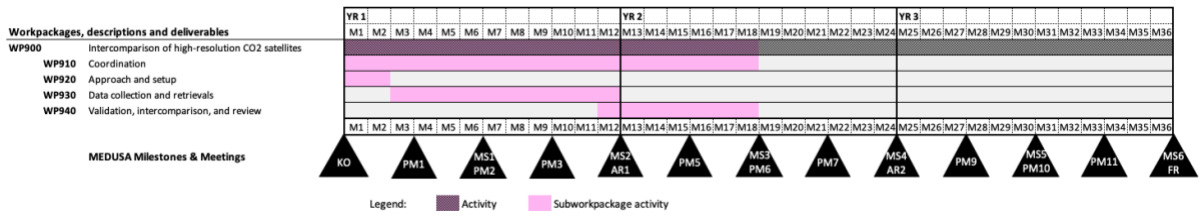


Figure 2. Gantt chart for the WP900 work package.

The schedule of this WP is planned to take 18 months. Acquiring satellite observations from the listed sites will occur over 10 months, starting at KO+2M (delivery of WP920 – approach and setup), up to April 2025 (KO+12M), leading to MS2 (WP930 – data

collection and retrievals). Finally, over the last 6 months of the activity, WP940 will focus on the validation, intercomparison & review, leading to MS3 and associated final report.

Quarterly WP900 internal meetings will be scheduled to track the progress of data collection and adjust the approach if necessary.

5 Data collection [WP930]

5.0 Satellite observations

A total of 164 clear-sky observations were collected across the 3 instruments and 10 sites, from which a subset of 100 has been selected for analysis. Retrievals from selected observations have been uploaded to the SRON SharePoint, and a summary is presented in Table 4 below.

Table 4. Number of clear-sky and selected observations for each site and instrument

Site	Reported Emission estimate (Mt/y)	EMIT clear-sky	EMIT selected	EnMAP clear-sky	EnMAP selected	GHGSat clear-sky	GHGSat selected
James H Miller Jr	18.109	6	4	2	2	5	4
Four Corners Steam Elec Station	7.853	5	4	5	4	9	4
Stanwell Power Station	7.291	2	2	6	4	4	4
Rockport	3.389	4	4	6	4	2	2
Gorazdze Cement Plant	2.5	7	4	2	2	4	4
Callide (B, C)	2.049	1	1	3	3	3	3
High Desert Power Project	1.746	4	4	6	4	5	4
TotalEnergies Raffinerie Mitteldeutschland	1.6	7	4	3	3	0	0
Cartagena Escomberas Power Station	1.429	3	3	5	4	8	4
Platte	0.487	9	4	3	3	5	4
Total	**100 selected**	**48**	**34**	**64**	**33**	**52**	**33**

In order to get a minimum of 3 clear-sky observations per site per satellite, the original data collection plan and satellite tasking was designed to measure each site as many times as possible (>3). At the end of the data collection period, this resulted in most sites having been oversampled by most instruments. However, a few site-satellite pairs had fewer successful measurements, mainly due to variability in cloud coverage. To compensate, we increased the number of selected observations from 3 to 4 per site-satellite pair whenever possible.

For cases where more than 4 observations were available, we decided which observations would be included in the study based on the following criteria:

- If measurements of the same site by two or more instruments in the same week were available, they were included. We hypothesize that emission rates might be more correlated over a shorter time lag, allowing a more direct intercomparison between the instruments of the same-week pair.
- If yet more than 4 observations could be chosen from, we favoured a wider seasonal coverage by selecting observations over the full duration of the data collection period (as opposed, for example, to selecting only the more recent ones).

We did not select observations based on the presence or absence of CO_2 plumes, in order to prevent biases in upcoming detection sensitivity analyses.

5.1 Reference data

5.1.0 EPA emissions measurements

The Environment Protection Agency (EPA) in the United States provides hourly CO_2 emissions measurements of power plants, which are published quarterly after a one-month reporting period. At the time of writing, emissions data is available before December 31, 2024. For the observations collected during Q1 of 2025, data is expected to be available in early May 2025, in time for the validation and intercomparison phase of the study (WP940).

Table 5. EPA facility ID and information for the 5 sites located in the US

EPA ID	Facility Name	State	Units	Stacks	Primary Fuel	Latitude	Longitude
59	Platte	NE	1	1	Coal	40.8538	-98.3481
2442	Four Corners Steam Elec Station	NM	2	1	Coal	36.69	-108.481
6002	James H Miller Jr	AL	4	4	Coal	33.6319	-87.0597
6166	Rockport	IN	4	1	Diesel Oil	37.9256	-87.0372
55518	High Desert Power Project	CA	3	3	Natural Gas	34.5953	-117.365

The emissions are obtained by querying the EPA's application programming interface (API), based on the facility identifiers listed in Table 5 and date of the observation. We

download data for all power generators over the relevant 24-hour periods, and interpolate emissions based on the local time of the satellite overpass. This process is illustrated in Figure 3 for one of the James H Miller Jr observations by EMIT.

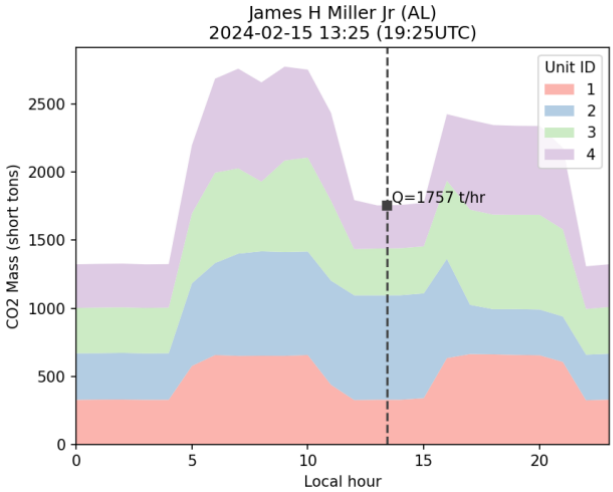


Figure 3. Example CEMS-measured CO_2 emissions from the EPA database, interpolated at the local time of the satellite overpass (indicated by the dotted line).

As a note, the flue stacks from which CO_2 and other combustion byproducts are emitted may be separated by distances greater than the spatial resolution of the instruments (30-60 m). Multiple plumes could therefore be distinguished, at least in principle, from the James Miller station where stacks are spaced by ~150 m (the distance is ~30 m for High Desert, and others have a single stack). However, this effect was not observed in the plumes identified at this point, possibly because of wind direction, viewing perspective, or because the plume tails merge too rapidly as they are diffused downwind. Further investigations into this topic are planned in the validation and intercomparison phase [WP940].

5.1.1 Power generation from Australian sites

The National Electricity Market (NEM) in Australia provides reports with sub hourly data of power generation by power unit for all power stations in the country (Table 6). The Australian National Greenhouse Accounts Factors [15] provides emission intensities per region. Both assets are in Queensland. These data will be used as proxy of CO_2 emissions. Example NEM data is shown in Figure 4.

Table 6. NEM information for the two sites located in Australia

Facility Name	Area	Units	Stacks	Primary Fuel Type	Latitude	Longitude	Emission Intensity

							(ton.CO2e/MWh)
Stanwell Power Station	Stanwell, Australia	4	1	Coal	-23.509	150.617	0.71
Callide (B, C)	Mount Murchison, Australia	2	2+2	Coal	-24.347	150.617	0.71

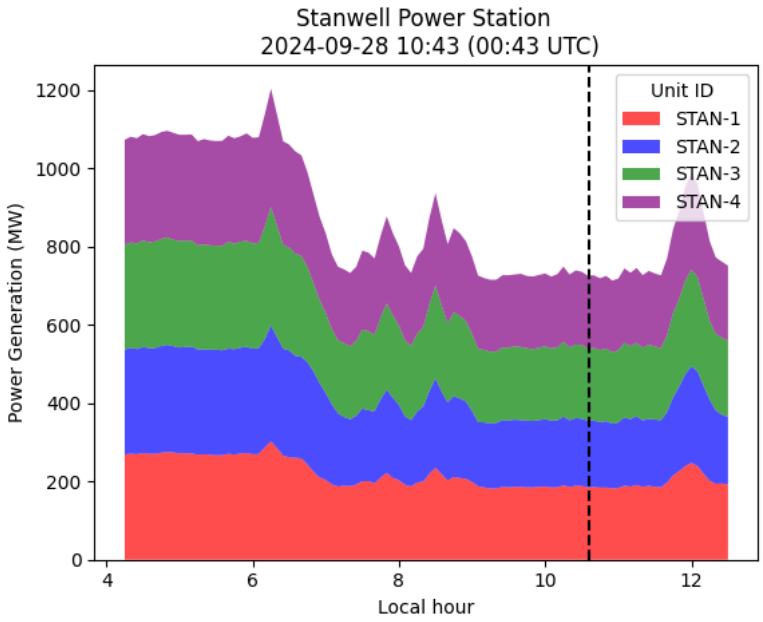


Figure 4. Example NEM Power Generation 5-min Data for the Stanwell Power Station (AUS). The dotted line represents the local time of the satellite overpass.

5.1.2 Power generation from European sites

The European Network of Transmission System Operators for Electricity (ENTSOE) in Europe provides hourly power generation data by unit for all power plants, with less than a day lag (

Table 7). These data will be used as proxy of CO_2 emissions.

Table 7. ENSTOE information for the site located in Spain

Facility Name	Area	N Units	Primary Fuel Type	Latitude	Longitude	ENTSOE 2023 Power Generation (MWh)	EEA 2023 Emission Estimates (ton.CO2-e)	Emission Intensity (ton.CO2-e/MWh)
Cartagena	CTA, ES	3	Fossil Gas	37.572	-0.938	2708129 MWh	1055873 ton	0.3898

Example data are shown in Figure 5. Note this sample time series is an interesting example of a limitation of satellite-based CO_2 emissions monitoring, as the overpass was collected at a time of day when the power plant was not operating. Of the 100 observations collected in this study, 76 had non-zero reference emissions, and 24 had zero emissions. All 10 sites had some non-zero emissions, confirming they are all active facilities. The least active site at overpass times was Platte with 3/11.

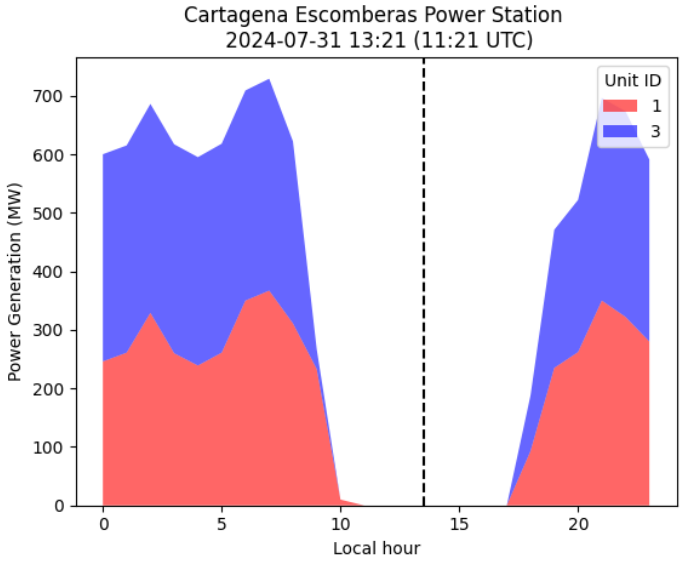


Figure 5. Example ENTSO-E hourly Power Generation Data for Cartagena Escomberas Power Station. The dotted line represents the local time of the satellite overpass.

5.1.3 Non power-generation sites

The two non-power generation sites selected are located in Europe. The dataset of verified yearly CO2 emissions, provided by the European Environment Agency (EEA) will be used as reference. This dataset has a low temporal granularity, but it can be used to sanity-check the order of magnitude of retrieved emissions. The 2023 dataset can be downloaded through the following link: https://climate.ec.europa.eu/document/download/ebb2c20e-8737-4a73-b6ba-

a4b7e78ecc01_en?filename=verified_emissions_2023_en_1.xlsx (see summary in Table 8). The European Commission announced that the 2024 dataset will be available on April 4, 2025.

Table 8. Emissions for the refinery and cement plant from European Environment Agency 2023 inventory.

EEA ID	Facility Name	**Type**	**2023 Verified Emissions (Mt)**	Latitude	Longitude
490	Goradze Cement Plant	Cement Plant	2.51	50.535	17.978
20	TotalEnergies Raffinerie Mitteldeutschland	Refinery	1.69	51.282	11.997

6 CO_2 retrieval algorithms [WP930]

Figure 6 below provides examples of CO_2 emission plumes detected by each of the instruments, and processed using the retrieval algorithms from each of the teams. The following subsections will describe the algorithms used to retrieve local CO_2 enhancements (L2 and L3 products), and to estimate the emission rate (L4 product).

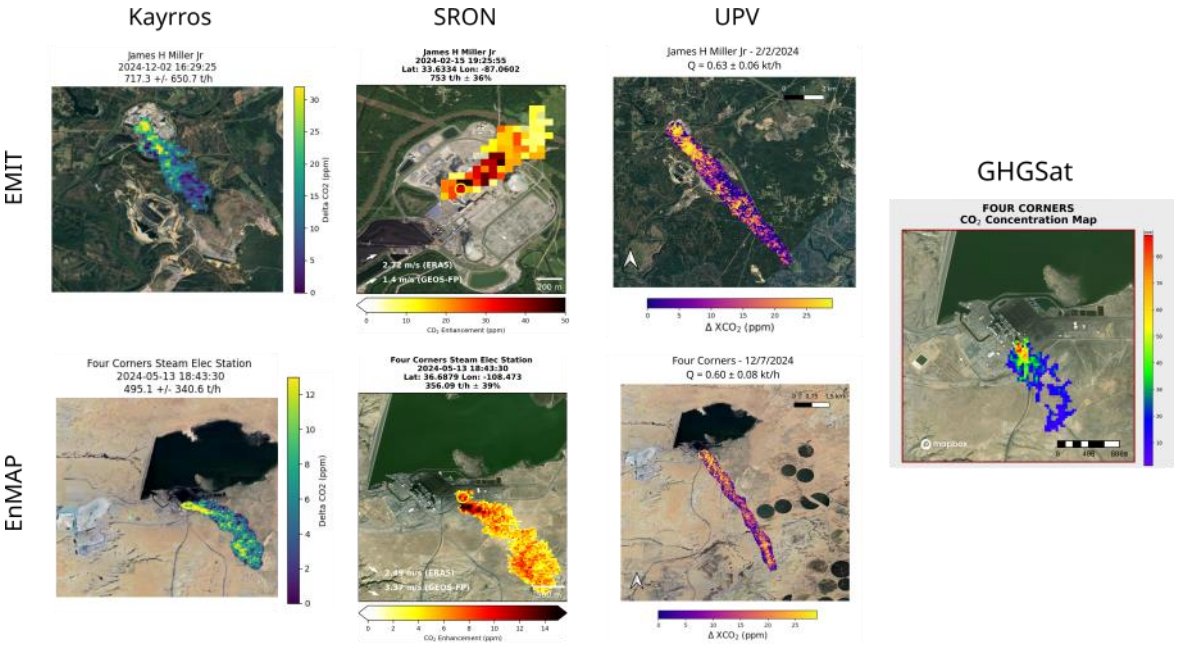


Figure 6. Example CO_2 emission plumes detected using EMIT, EnMAP, and GHGSat-C10, and processed by the consortium partners. The EMIT and EnMAP images do not all show data from the same date.

6.0 Level-2/3: Spatially-resolved CO2 enhancement

Level 2 and 3 retrieval algorithms broadly fall into 2 categories:

- For hyperspectral sensors (EMIT and EnMAP), all teams use variants of the matched filter method.
- For GHGSat-C10, nonlinear regression is used to simultaneously estimate instrument and atmospheric parameters.

6.0.0 Kayrros

Kayrros algorithm is based on the matched filter formulation described in Guanter et al., 2021, applied to the L1B hyperspectral product. The matched filter is performed on a per-column basis on the spectral ranges 1520 nm – 1670 nm and 1940 nm – 2105 nm where carbon dioxide has respectively weak and strong absorption features.

For every scene, a high-resolution reference spectrum is generated by integrating the Beer-Lambert law on an atmosphere model of homogeneous layers using:

- The observation parameters (viewing zenith angle, solar zenith angle) of the scene.
- The absorption coefficients from the HITRAN 2020 molecular spectroscopic database.
- The 1976 US standard atmosphere from LOWTRAN 7 atmospheric transmittance and background radiance model (Kneizys et al., 1988) is used for defining the composition of the atmospheric layers.

The unit absorption spectrum is eventually computed as the Jacobian of the reference absorption spectrum for an additional ppm of carbon dioxide.

6.0.1 SRON

SRON starts from L1 data for both instruments. For EnMAP the L1 data come from the EOWEB GeoPortal (https://eoweb.dlr.de/); for EMIT we download L1 data from NASA Earthdata (https://search.earthdata.nasa.gov/search?q=C2408009906-LPCLOUD). No pre-treatment is applied to the L1 data.

The SRON algorithm applies matched filters with two different wavelength windows, 1930–2200 nm and 1300–2500 nm, to retrieve the carbon dioxide enhancement and denoised carbon dioxide enhancement field, respectively. The original carbon dioxide enhancement field is used for emission quantification, while the denoised one is used for plume masking. The matched filter is applied per-column, with a land/water mask applied based on OpenStreetMap and ESA WorldCover data (https://hyp3-docs.asf.alaska.edu/water_masking/).

To determine the unit carbon dioxide absorption spectrum (k), we employ a forward model (Gloudeman et al., 2008) and convolve the radiance with the imager's central wavelength and FWHM. The k value (ppm^{-1}) for each band is calculated as the slope of the regression between the natural logarithm of radiance and carbon dioxide enhancements.

Plume detection is carried out by automasking and visual inspection. First, a Chambolle total variance denoising filter (Chambolle et al., 2004) is applied to obtain a smoothed carbon dioxide enhancement field from the less noisy carbon dioxide field (generated using the 1300 nm – 2500 nm window). Then, plume pixels are identified by a watershedding technique. It treats pixel values as a topographic surface and separates them into catchment basins. Threshold values of 2 and 3 standard deviations are used to identify multiple localized high-enhancement features and nearby areas with high enhancement values. Finally, plume pixels are removed from

the background estimation for the matched filter with the narrow window (1930–2200 nm) to make sure the sparsity assumption of the matched filter is not broken.

6.0.2 UPV

We use radiance L1 data for the EnMAP (https://planning.enmap.org/ips/app), and EMIT (https://search.earthdata.nasa.gov/search) missions. Pixels related to large water bodies that can interfere with our data-driven algorithms are filtered out in a supervised step.

We use the simple matched-filter (MF) (Guanter, 2021) for plume detection and the lognormal matched-filter (LMF) (Pei, 2023) for emission quantification. MF provides better precision, which helps to better distinguish plume pixels from background noise. On the other hand, LMF provides better accuracy, which is more suitable for emission quantification. For both methods, we carry out 2 iterations and in the second one we discard the potential CO_2 emission pixels to extract the required input parameters, i.e. mean and covariance. The unit gas absorption spectra (K) is obtained considering the acquisition specific conditions and a range of CO_2 mixing ratio length enhancements from 0 ppm·m to 60,000 ppm·m, which covers the typical plume enhancement values from the CO_2 retrievals showed in Thorpe (2023). Note that if the LMF retrieval does not show a clear plume, we also use the MF retrieval for emission quantification.

K is deduced using a LookUp Table (LUT) of radiance spectra obtained from combining HITRAN (Brown, 2013) and Libradtran (Mayer, 2005). Moreover, we use a default midlatitude summer atmospheric profile contained in the RFM (Dudhia, 2017) and scale it to current values. As in Foote (2021), gas enhancement is added to the first 500 m above the surface and the scaling height of 8 km is used to convert the ppm·m units to ppm (Thompson, 2016). Then, we can relate a set of enhancement values to a set of radiance spectra, which is first convolved to the instrument spectral response function and then a natural logarithm is applied. Finally, we perform a linear regression to extract the slope (K) that relates these two sets of values.

Our algorithms are applied in the 1960–2200 nm spectral range, where strong CO_2 absorption features are located. We set the lower limit at 1960 nm because radiance values at lower wavelengths can be nearly zero due to strong water vapor absorption. Moreover, the retrieval methods are applied in a per-column basis to avoid effects from the non-uniformity from the across-track detectors from the instruments.

Plume detection is made by visual inspection in those retrievals used for detection. The potential plume must 1) follow wind direction, 2) originate from a potential emitting source, and 3) have a plume-like shape.

For EnMAP's case, retrievals and related outputs are orthorectified using a combination of RPCs from the EnMAP L1 product metadata and a DEM extracted from SRTMGL1 (NASA JPL, 2013), which is fetched using Google Earth Engine and. For EMIT's case, the orthorectification is carried out using the GLTs attached in the L1 data.

6.0.3 GHGSat

GHGSat uses the same measurement concept and retrieval algorithms for CO2 as for CH4, which has been described in Ref. [7] of which we hereby provide a summary.

The GHGSat spectrometer is based on a wide-angle Fabry-Perot (WAF-P) imaging concept. The F-P element consists of two optical flats mounted within a mechanical enclosure. The two optical surfaces are positioned such that the inner surfaces, with reflectivity R and spaced a distance d apart, form an optical cavity. Light with wavelength λ and incident angle θ with respect to the F-P surface normal is transmitted according to:

$$T_{FP}(\theta,\lambda) = \frac{1}{1+\left(\frac{2\,\mathcal{F}}{\pi}\right)^2 \sin^2\left(\frac{2\pi n d\cos(\theta)}{\lambda}\right)} \quad (1)$$

where n is the index of refraction of the medium within the optical cavity and $\mathcal{F} = \pi\sqrt{R}/(1-R)$ is the reflectivity finesse.

In applications using F-P based spectrometers, the F-P gap spacing d is often dynamically scanned during the measurement. In contrast, the GHGSat instrument uses a fixed gap spacing and exploits the angular dependence of the m^{th} F-P transmission mode's spectral position $\lambda_m = 2\pi n d\cos(\theta)\,/m$ to measure the spectrum of the incident light. To perform a measurement, a programmable number of closely overlapping two-dimensional images are taken in which the atmospheric absorption spectrum is "imprinted" on the images in the form of spectral rings due to the angle-dependent Fabry-Perot transmission spectrum. During the observation sequence, the ground target traverses the field-of-view, sampling the full extent of the spectral information contained in the images.

The goal of the retrieval algorithm is to estimate the instrument and atmospheric state vector $\boldsymbol{x}$ from a measurement vector $\boldsymbol{y}$. The measurement vector is the photocurrent Φ collected by the pixel array of the SWIR spectrometer, in units of electrons per second per pixel. The state vector includes CO2 concentration as the main quantity of interest, expressed as the vertically integrated column density Ω_{CO2} in units of moles

per meter-squared, as well as other unknown variables that influence the measured signal, such as the ground albedo A and instrument parameters.

A forward model $\boldsymbol{F}(\mathbf{x})$ is constructed, which describes the light propagation through the atmosphere and instrument and predicts how much light intensity is sensed at the instrument's pixel array:

$$\boldsymbol{y} = \boldsymbol{F}(\mathbf{x}) + \boldsymbol{\epsilon}_y + \boldsymbol{\epsilon}_F \tag{2}$$

where $\boldsymbol{\epsilon}_y$ represents the measurement error and $\boldsymbol{\epsilon}_F$ represents error in the forward model.

The forward model represents our best knowledge of the instrument and atmosphere, with approximations used to evaluate the model more efficiently when performing retrievals. The camera signal $F_{i,j}$ at detector pixel (i,j) in photocurrent units [e^- s^{-1}] is given by:

$$F_{i,j}(\mathbf{x}) = \int L(\mathbf{x},\lambda) \cdot C(\lambda) \cdot QE(\lambda) \cdot T_{OSF}(\lambda) \cdot T_{FP}(\theta,\lambda) d\lambda \tag{3}$$

where $L(\mathbf{x},\lambda)$ is the spectral radiance as a function of the state parameter $\mathbf{x}$ and wavelength λ, $C(\lambda)$ is the radiometric conversion factor that converts spectral radiance to the number of photons on a pixel per unit time, $QE(\lambda)$ is the quantum efficiency with which the camera converts a photon to electric charge, $T_{OSF}(\lambda)$ is the transmission of the order-sorting filter that defines the spectral bandpass region, and $T_{FP}(\theta,\lambda)$ is the F-P transmission function.

The spectral radiance $L(\lambda)$ received by the instrument is the solution to a simplified radiative transfer equation where thermal emission, aerosol and molecular scattering have been neglected. The solution to this problem is a form of Beer-Lambert's law:

$$L(\lambda) = L_0(\lambda) \cdot \exp\left[-\sum_s \tau_s(\lambda)\right] \tag{4}$$

where $L_0(\lambda)$ is the spectral radiance in the absence of atmospheric absorption, and $\tau_s(\lambda)$ is the optical depth associated with gas species $s \in \{\mathrm{CH_4}, \mathrm{CO_2}, \mathrm{H_2O}\}$ defined as

$$\tau_s(\lambda) = \int n_{\mathrm{air}}\, x_s\, \sigma_s(\lambda)\, dl \tag{5}$$

Here, n_{air} is the number density of air, x_s is the volume mixing ratio of species s, and $\sigma_s(\lambda)$ is its molecular absorption cross-section calculated using HITRAN [14]. The integration is understood to be carried along the slanted light path in the atmosphere, which is the sum of the downwelling limb from sun to ground and the upwelling limb from ground to instrument. We perform the integral numerically using discrete layers that are evenly spaced in pressure.

The inverse problem of retrieving the unknown state parameters from known measurement is solved by non-linear regression using the Levenberg-Marquardt algorithm, minimizing the cost function:

$$\chi^2(\boldsymbol{x}) = \left(\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x})\right)^{\mathrm{T}} \mathbf{S_o^{-1}} \left(\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x})\right) \quad (6)$$

where $\mathbf{S_o}$ is the measurement covariance matrix.

Because $\boldsymbol{F}(\boldsymbol{x})$ is a nonlinear function of $\boldsymbol{x}$, we must solve for the state vector iteratively. This requires knowledge of the Jacobian $\mathbf{K}(\boldsymbol{x}) = \frac{\partial \boldsymbol{F}(\boldsymbol{x})}{\partial \boldsymbol{x}}$ to weight the state vector step $\boldsymbol{\Delta x}^i$ taken during the i^{th} iteration. Evaluating the forward model and Jacobian is computationally expensive and impractical for the large amounts of data collected by the instrument. Instead, we use a 3 step procedure:

1. Image alignment to map the "fixed-camera" coordinates to a "fixed-ground" referential (each fixed-ground cell has been seen at multiple FP angles and thus contains the full spectral information)
2. Scene-wide average retrieval using the full forward model to estimate the scene-wide average state vector $\hat{\boldsymbol{x}}$
3. Per-column retrieval done using a linearized forward model (LFM) evaluated at the linearization point $\hat{\boldsymbol{x}}$.

6.1 Level-4: emissions quantification

Emission rate estimations can be computed from CO_2 concentration or column densities using the integrated mass enhancement (IME) method [8,9]. IME measures the local accumulation of carbon dioxide, which is driven by two factors: the emission rate, and the rate at which the CO_2 dissipates in the surrounding atmosphere, which in turn is driven by wind speed. The source rate is calculated using the equation

$$Q = \frac{U_{\mathrm{eff}}}{\sqrt{A}} \mathrm{IME} \quad (7)$$

where Q is the emission rate, U_{eff} is an effective wind speed, A is the plume mask area, and IME is the integral of CO_2 enhancements within the masked plume, defined as the connected pixels with enhancement above measurement noise. The effective wind speed is an empirically calibrated function of the true local wind speed, which can be estimated from meteorological databases or using local measurements.

All consortium partners use the IME method to estimate emission rate, with some variations in terms of plume identification, input wind speed, effective wind speed calculation, and plume masking algorithm. Those differences are outlined in the following subsections.

6.1.0 Kayrros

The plume mask is defined through manual visual inspection of the carbon dioxide anomaly map. To be considered a potential plume, the observed anomaly must meet three criteria: align with the wind direction, emanate from a possible emission source and have a plume-like shape.

Every Kayrros detection is quantified using the Integrated Mass Enhancement (IME) method (Varon et al., 2018). The ERA5 10m wind speed reanalysis data (Hersbach et al., 2018) is used, assuming a default 0.5 m/s uncertainty on each component of the wind vector. The wind speed is converted into effective wind speed according to the formula from Guanter et al., 2021. *Ueff = 0.34 * U10 + 0.44.* A flux rate uncertainty is estimated by combining the uncertainty on the wind data and an empirical 1-σ estimate of the retrieval noise.

6.1.1 SRON

The algorithm uses the Integrated Mass Enhancement (IME) method to quantify the emission rate. We perform instrument-specific calibrations for effective wind speed (U_{eff}) based on large-eddy simulations. U_{eff} depends linearly on the 10-m wind speed (U_{10}):

$$\text{EMIT: } U_{eff} = 0.35U_{10} + 0.44$$
$$\text{EnMAP: } U_{eff} = 0.33U_{10} + 0.45$$

The primary choice for wind is the ERA5 10-m wind speed. However, the GEOS Forward Processing (GEOS-FP) data is used in cases where the ERA5 wind direction differs from the plume direction by more than 90 degrees. If both the ERA5 and GEOS-FP wind data fail to accurately capture the wind direction, the algorithm defaults to using the ERA5 wind data. The length of the plume is calculated as the square root of the plume's area.

We consider several sources of uncertainty in our emission estimations: wind speed error, random error of the retrieval, and uncertainty in the IME calibration. The wind error is set as 1.5 m s^{-1} for wind speeds higher than 3 m/s. For wind speeds lower than 3 m/s, a relative wind error of 50% is applied following Varon et al. (2018). To quantify the effects of retrieval random error, the plume mask is applied to non-plume pixels across the entire scene and the standard deviation of the resulting emission rates is calculated. To capture the uncertainty in the calibration, the average fit residual is applied for point sources. Finally, the total uncertainty is defined as the square root of the sum of the squares of the individual uncertainties.

The EnMAP L2 product is orthorectified to mitigate image distortion caused by variable topography by combining Rational Polynomial Coefficients (RPCs) and a Digital Elevation Model (DEM), while the EMIT L2 product is orthorectified using a geometric lookup table (GLT). We provide our output in NetCDF format for spatial variables like the plume mask and in CSV files for the source rate and its uncertainty (given in ton CO_2/hr).

6.1.2 UPV

Wind speed values are obtained from GEOS-FP reanalysis product (Molod et al., 2012), from which we obtain U10, the wind speed at 10 m above the surface. If this product is not available, the same parameter is extracted from ERA 5 - Land (Muñoz, 2021) through Google Earth Engine.

Plume masking is performed through a supervised procedure in which the user defines a polygon by clicking on the retrieval displayed on the computer screen, capturing the plume's shape.

To obtain the effective wind speed, we carried out a calibration with simulated CO_2 plumes, using typical flux rate (Cusworth, 2023) and retrieval noise values. For EMIT, we found Ueff = 0.3 · U10 + 0.74 and for EnMAP we obtained Ueff = 0.24 · U10 + 1.01. Finally, the flux rate emission uncertainty is obtained by the square root of the quadratic sum of the errors related to the IME and the Ueff values. The latter error is driven mainly by the error related to U10, from which we assume an error of 50 % (Guanter, 2021).

6.1.3 GHGSat

Emissions are identified by visual inspection of the L2 retrieval (map of CO2 column density), based on the strength of the enhancement relative to the background noise. To distinguish real enhancements from artifacts, the operator assesses the spatial structure of the plume (general morphology, alignment with the wind direction), and ensures the enhancement is not correlated with features present in the albedo or its gradient. Knowledge of the infrastructure (or lack-thereof) also factors in the decision process.

Once plumes are identified, a binary mask of the plume is computed by successively applying a spatial band-pass filter, intensity threshold, and rejecting isolated pixels disjoint from the main body of the plume. This mask identifies which connected ground cells have a column density above the measurement uncertainty and originate from the defined source.

To quantify emissions, GHGSat uses a variant of the IME method, where the effective wind speed is defined as a linear function:

$$U_{\text{eff}} = \alpha_1 \, \mathrm{U}_{10} + \alpha_2 \tag{8}$$

where U_{10} is the wind speed at 10 m above ground obtained from a third-party database, and $\alpha_1 = 0.34$ and $\alpha_2 = 0.42$ are model parameters obtained through numerical simulations [9].

Each emission quantification Q is associated with a measurement uncertainty estimate ΔQ, which accounts for variability in observation conditions such as ground reflectance, sun illumination, terrain, and wind. The uncertainty is defined as the sum in quadrature of 3 individual error components which are assumed to be weakly or non-correlated:

$$\Delta Q = \sqrt{\Delta Q_{\text{meas}}^2 + \Delta Q_{\text{wind}}^2 + \Delta Q_{\text{mod}}^2} \tag{9}$$

The first error component ΔQ_{meas} is the measurement error associated with imperfect CO_2 concentration retrieval, which include shot noise, camera read noise, and other ground-correlated noise sources. ΔQ_{wind} is the error on the wind speed, which arises from using a coarse resolution meteorological database to estimate the local 10-m wind speed U_{10}. Its value is an empirical function of wind speed obtained by comparing the database U_{10} against precise measurements from weather stations, and it is typically on the order of 35%. ΔQ_{mod} is a model error which is associated with the use of the same effective wind speed model (conversion from U_{10} to U_{eff}) for plumes of varying shapes and sizes in different meteorological conditions. A constant value of 7% is used, obtained from numerical simulations [9].

7 Validation, Intercomparison & Review [WP940]

7.0 Number of observations and detections

Table 9 below provides a summary of the total number of observations and detected plumes for each of the 10 monitored sites. Note that for public hyperspectral missions (EMIT and EnMAP) 8 of the 66 combined observations had conflicting detection assessments between the processing teams, which are listed in Table 10. The plume counts in Table 9 are the average number of detections across the 3 teams.

In terms of the number of emissions successfully detected, EnMAP performed best (49% of observations contained plumes), followed by EMIT (25% plumes/observations) and GHGSat (15% plumes/observations). In terms of the differences between providers (Kayrros, SRON, UPV), SRON was generally more conservative as they assessed that for some observations, correlations between CO_2 enhancements and albedo were too strong to reliably detect and report a plume.

Table 9. Total number of observations and detected plumes for each site.

Site	EMIT obs	EMIT plumes	EnMAP obs	EnMAP plumes	GHGSat obs	GHGSat plumes
Callide (B, C)	1	0.00	3	3.00	3	0
Cartagena Escomberas	3	0.67	4	0.00	4	0
Four Corners	4	1.00	4	3.33	4	2
Gorazdze Cement Plant	4	1.33	2	1.00	4	0
High Desert Power Project	4	0.00	4	0.67	4	0
James H Miller Jr	4	3.00	2	2.00	4	2
Platte	4	0.00	3	0.00	4	0
Rockport	4	2.00	4	2.67	2	0
Stanwell Power Station	2	0.67	3	3.00	4	1
Total Mitteldeutschland	4	0.00	3	0.00	0	
Sum	**34**	**8.67**	**32**	**15.67**	**33**	**5**

Table 10. EMIT and EnMAP observations where Kayrros, SRON and UPV made different emission assessments (presence or absence of plume). N.D. indicates non-detection or non-reported detection (for example due to concerns over potential correlations with albedo) while

N.Q. indicates detection without quantification as the assessment was made that a reliable emission rate quantification could not be performed due to interference from a lake or clouds.

Instrument	Site	Date	Ref. rate (t/hr)	Kayrros (t/hr)	SRON (t/hr)	UPV (t/hr)
EMIT	Cartagena Escomberas	2023-04-23	83	170	N.D.	175
	Gorazdze Cement Plant	2024-04-07	286	330	N.D.	530
	Gorazdze Cement Plant	2024-08-01	286	184	N.D.	325
	Stanwell Power Station	2024-11-03	523	144	N.Q. (clouds)	36
EnMAP	Four Corners	2024-10-10	606	101	N.Q. (lake)	794
	Four Corners	2024-10-22	458	179	N.Q. (lake)	622
	High Desert Power Project	2024-08-26	221	154	N.D.	39
	Rockport	2025-01-15	2347	N.D.	850	255

7.1 Detection threshold

Detection thresholds are derived from a probability of detection (PoD) model, which is fitted to the ensemble of detection events (detection or no-detection) via binary regression. The model uses a sigmoid-shaped function that can represent the smoothly increasing PoD as a function of "ground truth" emission rate Q and a vector of fitted parameters $\boldsymbol{\beta}$. The logistic model was used in this analysis, with a probability of detection expressed as:

$$p(Q; Q_{50}, s) = \frac{1}{1 + e^{-\frac{Q - Q_{50}}{s}}}$$

where Q_{50} is the emission rate that would be detected with 50% probability (detection threshold), and s is a transition width parameter. The regressions are performed separately for each instrument / measurement provider pair.

Detection results and the corresponding fitted probability models are shown in Figure 7 (using only samples with CEMS-measured emission rates), and Figure 8 (using samples with all available reference rates). The fitted results are similar in both cases, which provides some confidence that the approach for benchmarking CO_2-monitoring satellites based on EPA CEMS data can be generalized. The detection threshold at

50% probability is included in the figures and additionally summarized in Table 11. Detection thresholds range from around 2 MT/yr (EnMAP/UPV) to 18 MT/yr (C10/GHGSat). EnMAP was found to have the lowest detection threshold of the 3 instruments, at 4.4 MT/year when combining results from the 3 teams (CEMS only).

Table 11. Summary of detection thresholds at quantified using binary regression. The thresholds are specified at 50% probability of detection.

Instrument	Provider	N samples all refs	Q50 all refs (MT/yr)	N samples EPA CEMS	Q50 EPA CEMS (MT/yr)
EMIT	Kayrros	34	7.2	20	8.2
	SRON	34	8.5	20	8.2
	UPV	34	7.2	20	8.2
	All	**102**	**7.7**	**60**	**8.2**
EnMAP	Kayrros	32	3.5	17	4.8
	SRON	32	3.3	17	5.3
	UPV	32	2.0	17	1.4
	All	**96**	**2.9**	**51**	**4.4**
GHGSat-C10	**GHGSat**	**33**	**17.9**	**18**	**18.6**

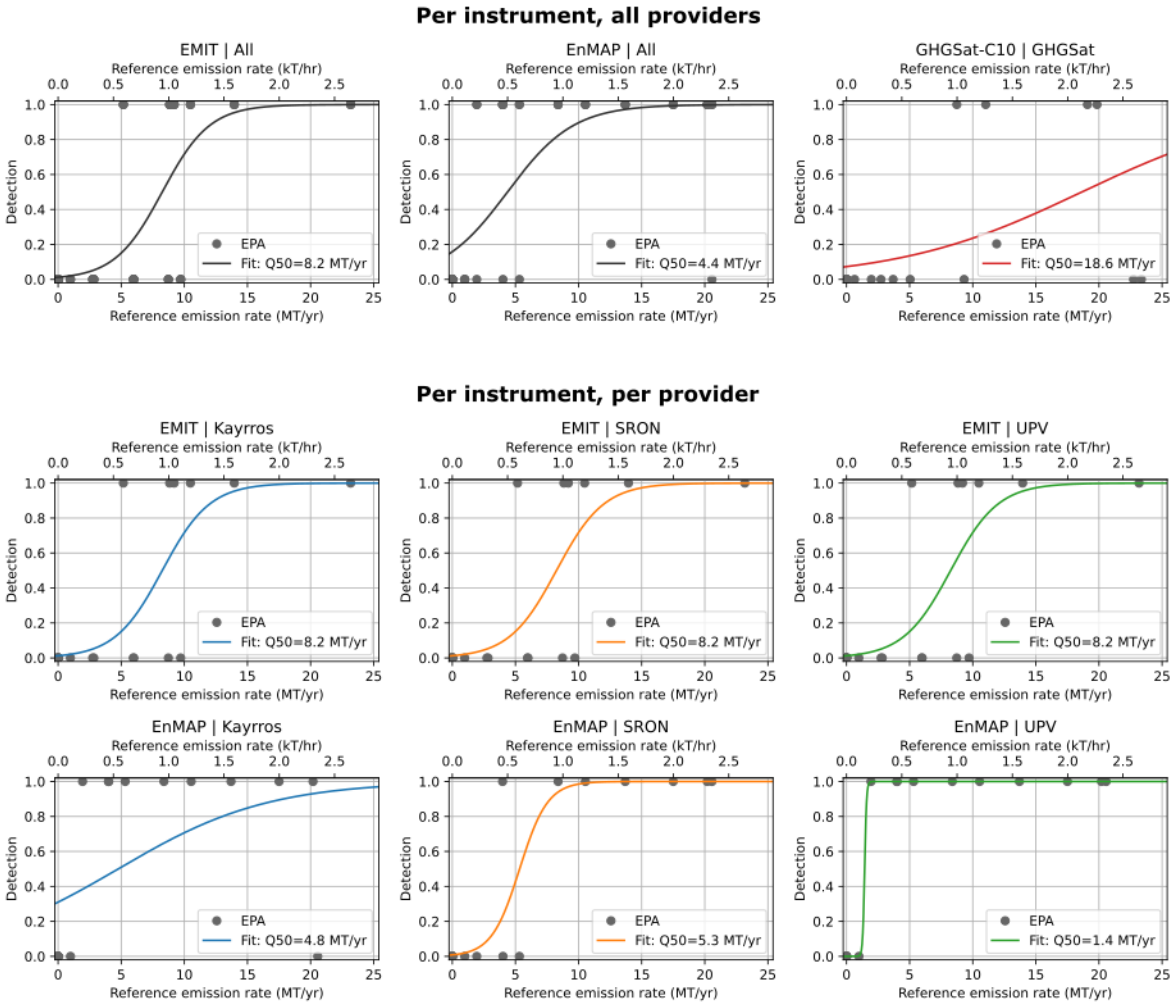


Figure 7. Binary detection results and probability of detection model fit for sites with CEMS measurements available.

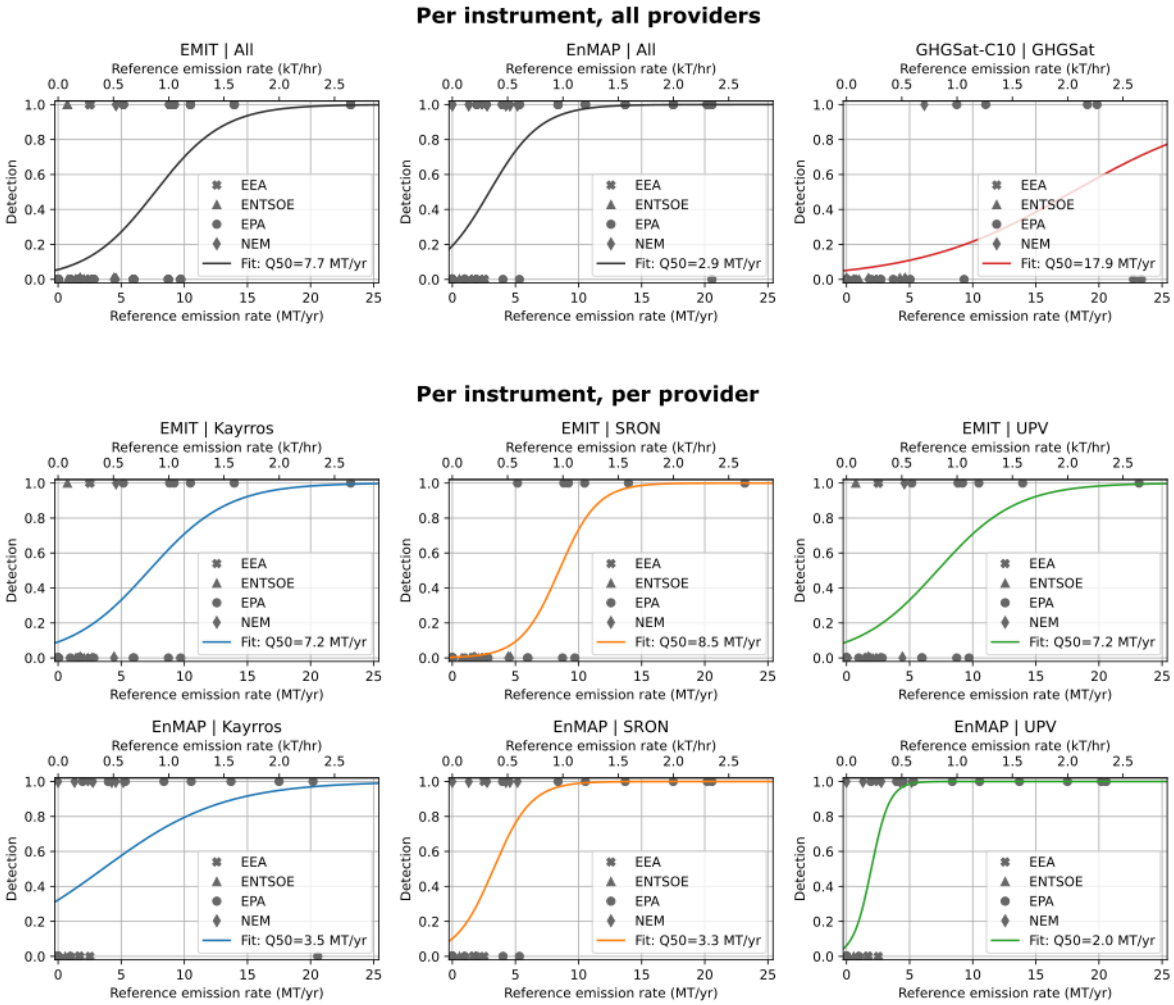


Figure 8. Binary detection results and probability of detection model fit for all sites.

A few notes must be made about the detection probability results. the relatively small sample size for each instrument results in high uncertainty of the detection probability, and high sensitivity to individual samples especially at high rates or in the rising portion of the probability curve. The difference in team-to-team variability between EMIT and EnMAP provide a clear example: whereas for EMIT, all teams reported the same detections (thus the perfect overlap in fitted PoD curves in Figure 7), a few differences for EnMAP detections resulted in completely different PoD curves.

There may also be differences in the level of confidence required to declare an emission detected for a given retrieved signal. The alternative hypothesis to explain team-to-team difference would be that there were significant differences in SNR for level 2 retrievals of corresponding scenes, but those were not seen in practice.

Wind speed *U* drives the dissipation of emitted gas plumes in the surrounding atmosphere and therefore has a strong influence on detection probability. Controlled release experiments performed for benchmarking methane-sensing satellites and airborne sensors generally included anemometers positioned close to the release point, which provide a robust way of explicitly accounting for wind speed in detection and quantification performance analysis[16,17]. The CO_2 emissions used in this analysis originate from real sites where such data is not available. This aspect limits the granularity in modelling which factors influence detection threshold, but do not invalidate the results in any way. Indeed, by performing observations over an extended period of time (a few months to a few years depending on site) and using a diversified set of geographic locations and topographies (as well as different meteorological situations and stack heights varying the atmospheric rise of the plume) effectively aggregates the effect of wind variability over a wide range of real-world conditions. In this sense, this approach is perhaps more representative of practical applications compared with controlled releases performed in a limited set of conditions.

7.2 Quantification accuracy

Quantification accuracy is assessed by comparing the emission rates estimated from satellite measurements against the reference rates derived from CEMS or other sources, described in Section 4.2.4. In the cases where more than one distinct plume was visible from the same site, the total of their emission rates was used and compared against the total reported emissions from all power generating units. Regarding time integration and interpolation, hourly data from EPA and ENTSOE represent the integrated emissions during the hour of day, i.e. from HH:00 to HH:59, and the rounded-down hour of the observation timestamp was used for indexing. For sub-hourly data from NEM, a 30-minutes integration window was used, starting 30-min prior and ending at the observation timestamp.

Quantification accuracy metrics are represented graphically in Figure 9 to Figure 12, and summarized in Table 12. Parity plots show the retrieved rates Q_m on the y axis as a function of the reference rate Q_r on the x axis. A linear regression going through the origin was fitted (ordinary least-squares) for each satellite / measurement provider pair, which can be compared against the 1:1 line (dotted grey). As was the case for the detection threshold analysis, a set of figures was generated with only sites where hourly CEMS measurements were available (US power plants), and another set with all available data including emission estimates based on proxy data. Vertical error bars are the retrieval quantification errors as described in the methodology section. Horizontal error bars were added for hourly CEMS data as the variation of the rate over a 2-hour period centered on the measurement timestamp. This error is meant to capture uncertainty of the time averaging kernel for the detectable integrated mass

enhancement. Finally, in addition to the parity plots, we plotted the error ratio $\varepsilon_r = (Q_m - Q_r)/Q_r$ as a function of the reference rate as an alternative way of visualizing the same results (Figure 11 and Figure 12). The mean relative error was computed, as well as a linear regression (ordinary least squares) to express a rate-dependent quantification bias.

Table 12. Summary of quantification accuracy metrics, averaged over all measurement providers (Kayrros, SRON, UPV) for EMIT and EnMAP.

Instrument	Reference data	Mean number of emissions	Parity regression slope	Parity regression R^2	Mean ε_r	Pearson's correlation (ε_r VS rate)
EMIT	All refs.	8.7	0.43	0.71	-26%	-0.59
	EPA CEMS	6.0	0.42	0.72	-39%	-0.60
EnMAP	All refs.	14.7	0.51	0.74	-5%	-0.53
	EPA CEMS	8.7	0.47	0.84	-41%	-0.40
GHGSat	All refs.	5	0.84	0.90	-21%	0.25
	EPA CEMS	4	0.84	0.90	-21%	0.33

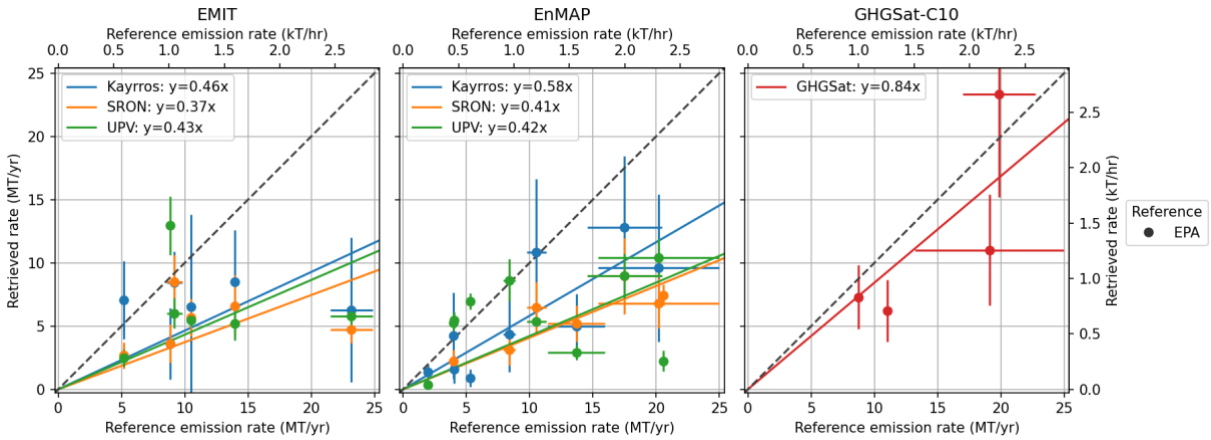


Figure 9. Quantification parity plots comparing retrieved emission rates with CEMS measurements.

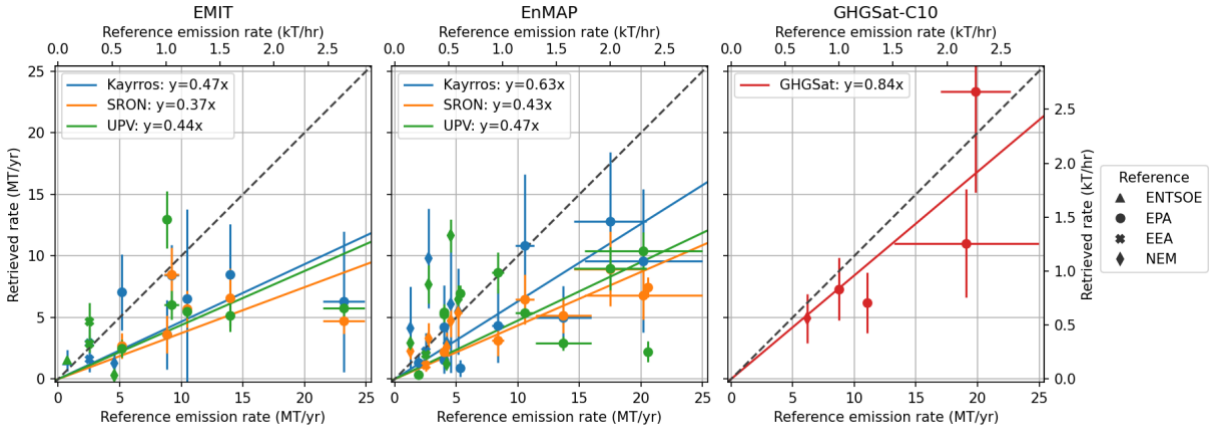


Figure 10. Quantification parity plots comparing retrieved emission rates with reference emissions from all sources.

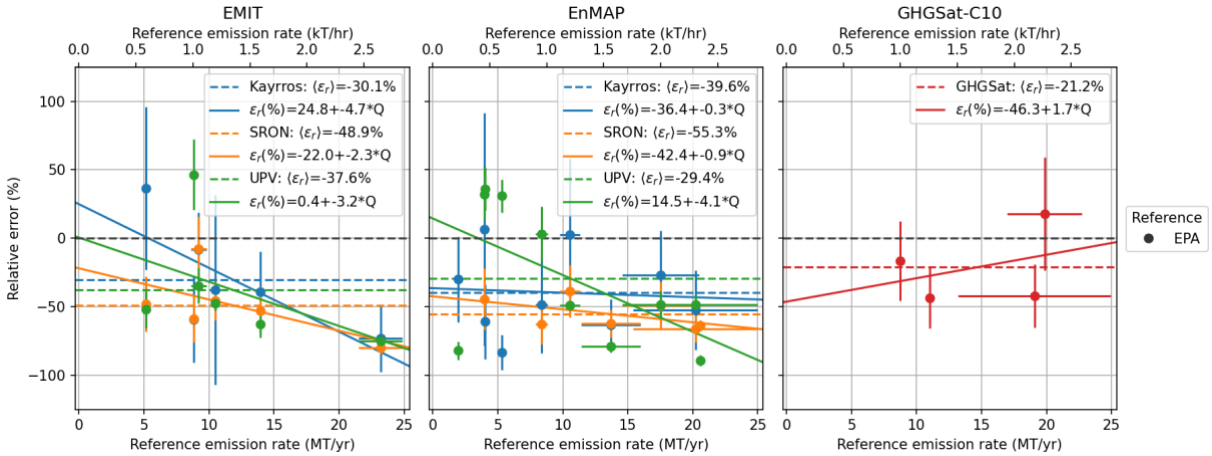


Figure 11. Relative quantification error compared with CEMS measurements. Dashed colored lines represent the mean bias, and full lines express the dependence of the bias on the emission rate via linear regressions (OLS).

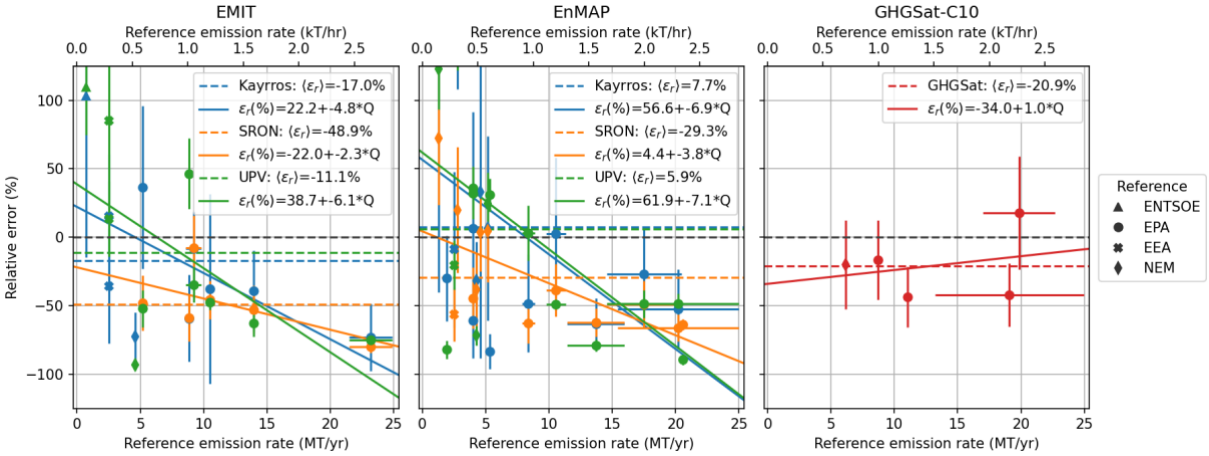


Figure 12. Relative quantification error compared with reference emissions from all sources. Dashed colored lines represent the mean bias, and full lines express the dependence of the bias on the emission rate via linear regressions (OLS).

Even though all regressions are based on a limited number of points and do not account for non-detections, we note significant discrepancies between retrieved emissions and the reference values. With EMIT and EnMAP measurements, the reference emissions were underestimated by around 35-65% based on the linear regression (mean slope = 0.45 for EMIT and EnMAP for CEMS-based references). The systematic underestimation was most pronounced for larger emissions (1.5 kT/hr or more), which can be observed by the negative correlation between the relative error and the source rate. GHGSat-C10 underestimated rates by a smaller margin (slope of 0.83 with all references), but the regression is only based on five data points.

Additionally, we evaluate the consistency of satellite-based measurements by comparing the emission rates retrieved by different teams for the same satellite observation. This intercomparison is presented in Figure 13 as parity plots. Both EMIT and EnMAP are included in the analysis. The bias between providers is quantified by fitting a line going through the origin using orthogonal distance regression (ODR), as opposed to ordinary least squares (OLS), because significant uncertainty is present on both the X and Y axes.

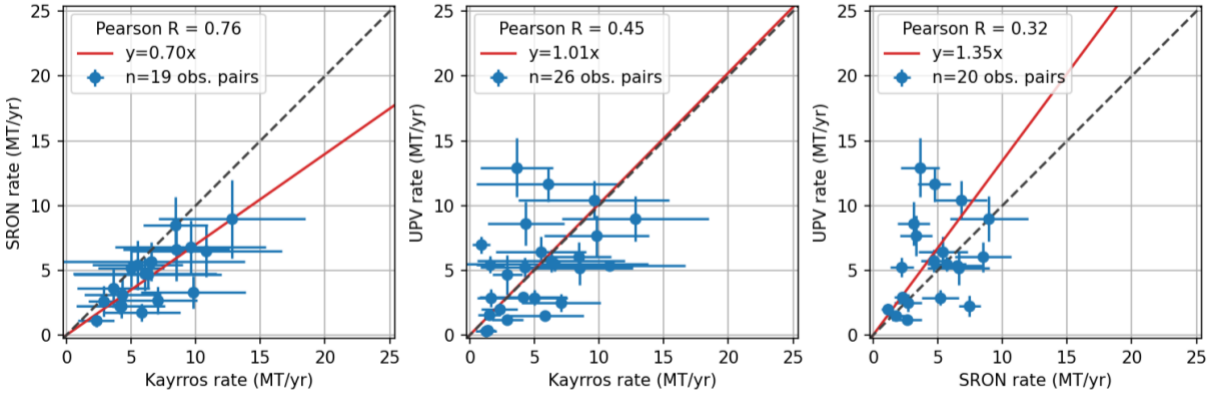


Figure 13. Parity plots comparing retrieved emission rates by different providers for matching observations. The full red line represents the orthogonal distance regression (ODR) for a line going through the origin. Error bars are the measurement uncertainty

We find a better correlation between rates measured by Kayrros and SRON (Pearson's R = 0.76) than with combinations involving UPV (R = 0.45, R = 0.32). This is most likely because UPV uses a different variant of the matched filter method for quantification (lognormal matched filter), whereas Kayrros and SRON use a linear matched filter. We also note that many of the estimated emissions differ by more than the uncertainty bars, which highlights the fact that discrepancies can exist despite the estimates being based on the same raw data. A discussion of possible causes based on inspection of the level 2 products is presented in section 7.3.1.

Using a similar methodology and plotting form, we compare emission rates retrieved by different satellites for the same site, shown in Figure 14. For this analysis, because observations were not synchronized between instruments and due to the significant uncertainty of each measurement, the rates are averaged for each site over all observations and measurement providers when applicable (Kayrros, SRON, and UPV for EMIT and EnMAP retrievals). Only sites where both instruments detected at least one plume were included in the plot and non-detections were not accounted for.

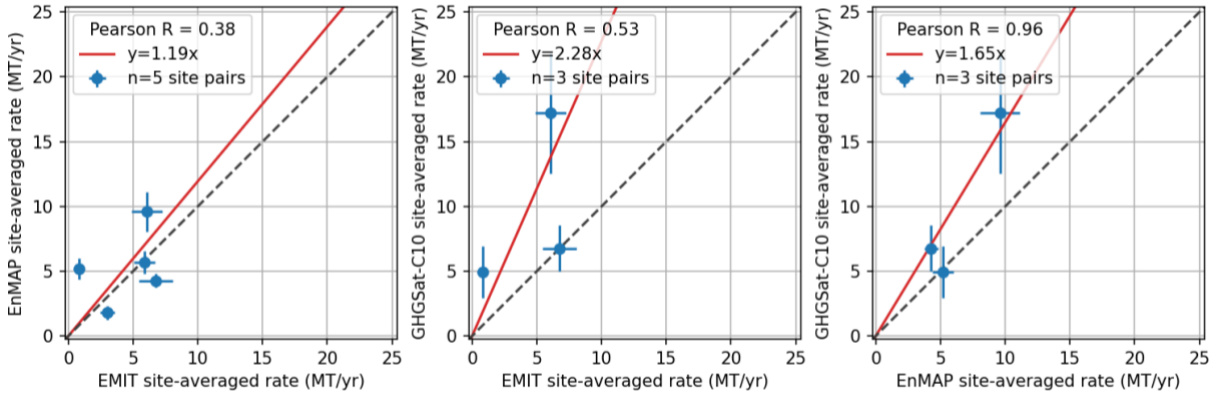


Figure 14. Parity plots comparing site-averaged retrieved emission rates from different instruments. Error bars are the standard error on the mean.

As previously discussed when comparing with CEMS-based reference rates, analyses of EMIT and EnMAP result in similar rates, which are biased low compared to GHGSat-C10. We observe varying levels of correlation between instruments (Pearson's R between 0.38 and 0.96). The correlation coefficients are slightly higher compared with the team-to-team intercomparison, which was done on individual estimates.This suggests that averaging multiple measurements may be required to accurately quantify the yearly emissions of a facility. Note that those conclusions have some uncertainty due the small number of sites where plumes were seen by at least 2 instruments (3 to 5).

7.3 Investigation of potential quantification error sources

7.3.0 Impact of wind altitude on quantification

We question whether the 10-meter wind speed (U10) that was used by all teams is the best wind speed to quantify CO_2 plumes, due to the height of the flue stacks from which emissions are released, and to the buoyancy of heated gases. Wind tends to be faster at higher altitudes, dissipating CO_2 to lower visible enhancements and resulting in source rate underestimation if a lower wind speed is used in the IME formula.

To estimate the magnitude of the underestimation, we collected reanalysis wind speeds at both 10- and 100-meter elevation from the ERA5 model for each of the MEDUSA WP900 observations. The gridded data were interpolated linearly to the site coordinates and observation timestamp. The distribution of the wind speed magnitude at both elevations is compared in Figure 15. As expected, the 100-m wind speed was generally higher than the wind at 10-m.

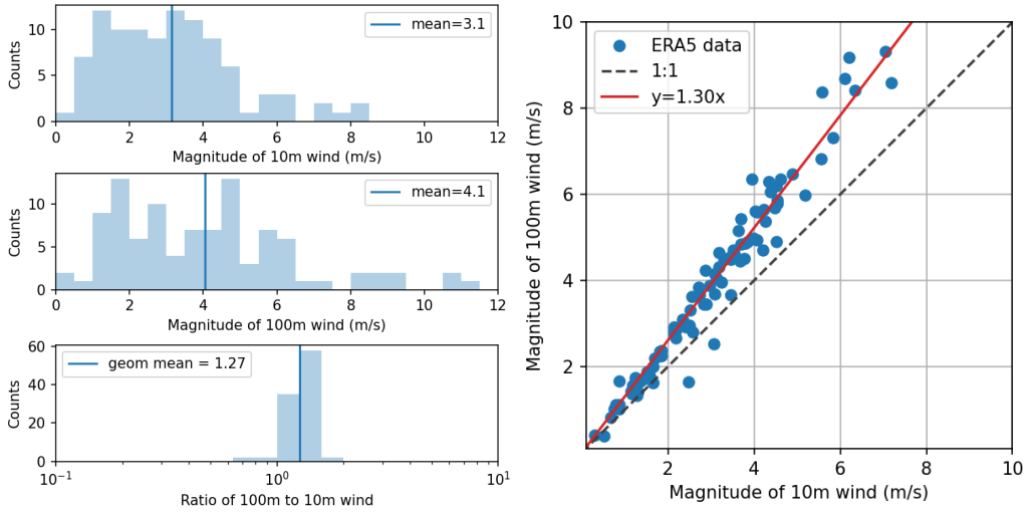


Figure 15. Comparison of wind speed at 10 m and 100 m above ground level, based on the ERA5 reanalysis product at location and timestamp of the MEDUSA CO_2 observations.

The relationship between U10 and U100 was quantified with 2 methods: the average ratio U100/U10, and the slope (OLS regression) of the U100 VS U10 parity plot. Both methods result in similar values: 1.27 and 1.30 respectively. In other words, using the 100-meter wind speed in the source rate retrievals instead of 10-meter wind would result in estimates that are about 30% higher (ignoring the offset term of the effective wind speed formula). This can therefore explain about half of the quantification biases observed on EMIT and EnMAP, while GHGSat would shift from a ~15% underestimation to a ~15% overestimation.

However, any global scaling factor between the wind speed "seen by the plume" and the available wind speed wind speed estimate should be lumped into the effective wind speed formula, and therefore not result in any bias, whether U10 or U100 is used. For this statement to apply, the plume dispersion simulations used to calibrate Ueff must include a realistic height of the release point and plume buoyancy due to the high temperature of released gases.

All teams use plume simulations in which the source is assumed to be located at surface levels and only considers U10. However, most of the sources observed with satellite data release CO2 at a height higher than 10 m. Therefore, the assumption of U10 as representative of the wind speed magnitude might lead to inaccurate wind speed calibrations for our observations, having an impact on the emission estimate. The effective wind speed calibration could therefore be improved with more CO_2-specific simulations.

7.3.1 Level 2 retrieval comparison

This section examines a few selected examples of the L2 products to seek possible causes of discrepancies.

Figure 16 shows a 2.3 kT/hr emission observed by EnMAP where all teams reported relatively similar rates, but underestimated the CEMS-reported emissions by around a factor 2. A histogram below each concentration map shows the distribution of column enhancements within the plume mask. A plume observed by GHGSat which has a similar CEMS-reported rate is also shown. The UPV rate is slightly higher than Kayrros, which is higher than SRON despite a similar plume mask. The SRON to Kayrros difference is likely explained by differences in retrieved enhancements. The higher UPV rate may be due to a combination of the mask including more downwind enhancements, and higher enhancements probably originated from the lognormal matched filter and the spectral window selection. The GHGSat retrieval has higher enhancements in the plume mask as shown in the histogram, and therefore results in a higher emission rate that is close to the CEMS reference. Note that the GHGSat retrieval isnoisier and contains more albedo-correlated artifacts, but those were excluded from the plume mask.

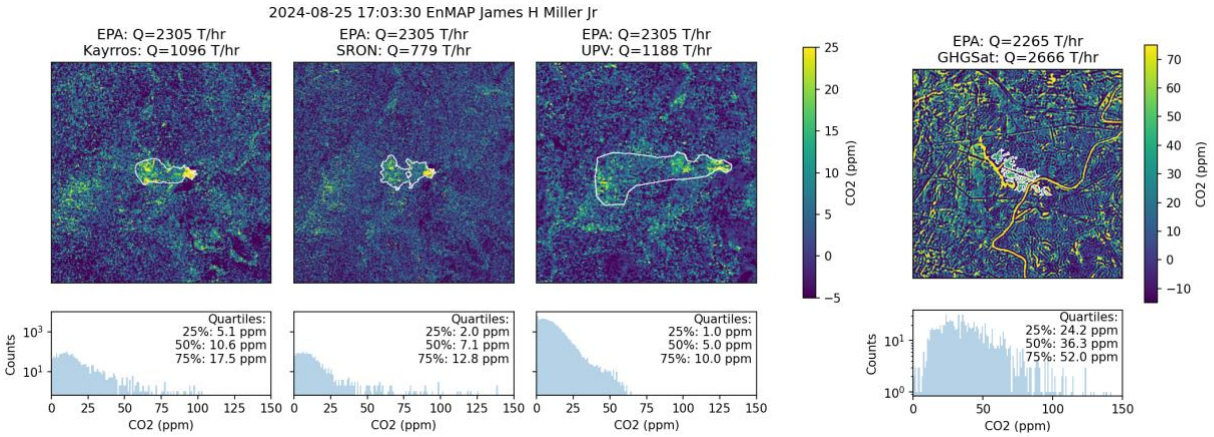


Figure 16. Example retrieved CO_2 for an emission observed by EnMAP, and a GHGSat plume of similar rate.

A second example shown in Figure 17 is a 1.0 kT/hr emission (CEMS-measured), observed by EMIT and a plume of similar magnitude observed by GHGSat. In this example, Kayrros and SRON underestimated the rate ($Q_{Kayrros/SRON}$ ~ 0.4 kT/hr), while UPV overestimated the rate (Q_{UPV} ~ 1.5 kT/hr). GHGSat-C10 produced a modest underestimation (Q_{GHGSat} ~ 0.8 kT/hr). The large discrepancies among the EMIT-measured rates are likely driven by the L2 enhancements being higher for UPV, as can be noted both in the enhancement heatmap and the histogram. Note that even

though the LMF retrieval from UPV seems more prone to albedo-related artifacts, these have limited impact on the CO_2 emission quantification if plume and artifact pixels do not overlap. Moreover, while the choice of retrieval method (MF or LMF) explained the largest fraction of the CO_2 retrieval overestimation at L2 by UPV , basic matched filter values from UPV are still higher than other groups. One possible explanation for this is the spectral range selection. While UPV used 1960 – 2200 nm, Kayrros used 1940 – 2105 nm, and SRON used 1930 – 2105 nm.

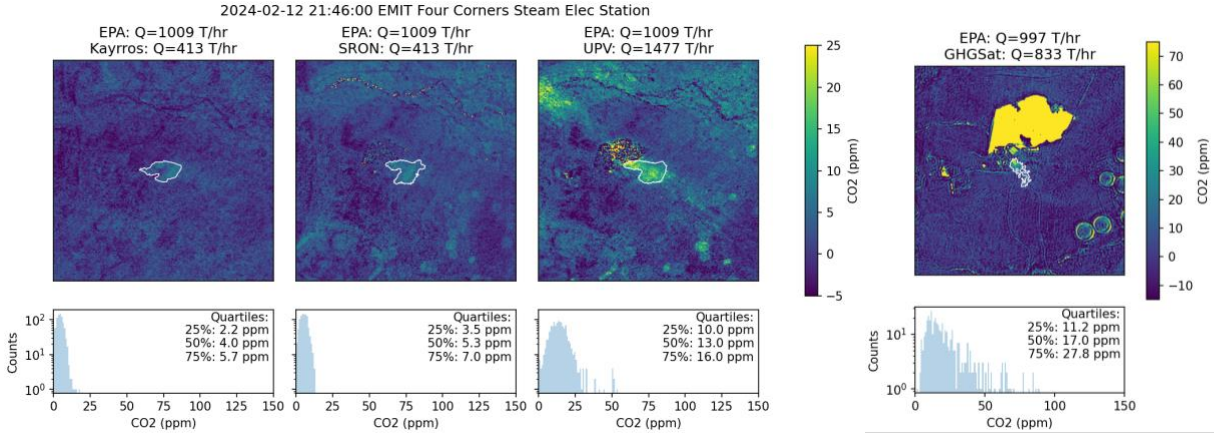


Figure 17. Example retrieved CO_2 for an emission observed by EMIT and a GHGSat plume of similar rate. Note that in the GHGSat retrieval, the bright yellow feature above the emission is an artifact from a water body.

7.3.2 Comparison of linear and lognormal matched filter

As described in section 6.0.2, the UPV team used two different variants of the matched filter method for ΔXCO2 retrieval: the simple matched-filter (MF) is used for plume identification, and lognormal matched-filter (LMF) is used for quantification, as it better models the nonlinearity of the optical signal variation with respect to CO_2 concentration. This section investigates the magnitude of this effect in the context of emissions quantification. For this purpose, we integrate simulated enhancements into L1 data. We choose a homogeneous scene from EMIT data and implement simulated enhancements into pixels from the same row. Then, we compare the retrieved values using both MF and LMF methods.

In Figure 18, we show the EMIT scene (left panel) that we selected to carry out the analysis. It shows the radiance band around 1960 nm as a reference, which captures a desert area in China. Here, we integrate randomly selected enhancements in 300 pixels from the same row. The row subset is selected because the corresponding columns (framed in red) captured surface homogeneity. We first implement enhancement values from 0 to 150 ppm to study the extent of the MF (center panel) and LMF (right panel) agreement with simulated data. Afterwards, we repeat the same

exercise between 0 and 15 ppm, which is a representative range of values observed in CO_2 plumes by UPV.

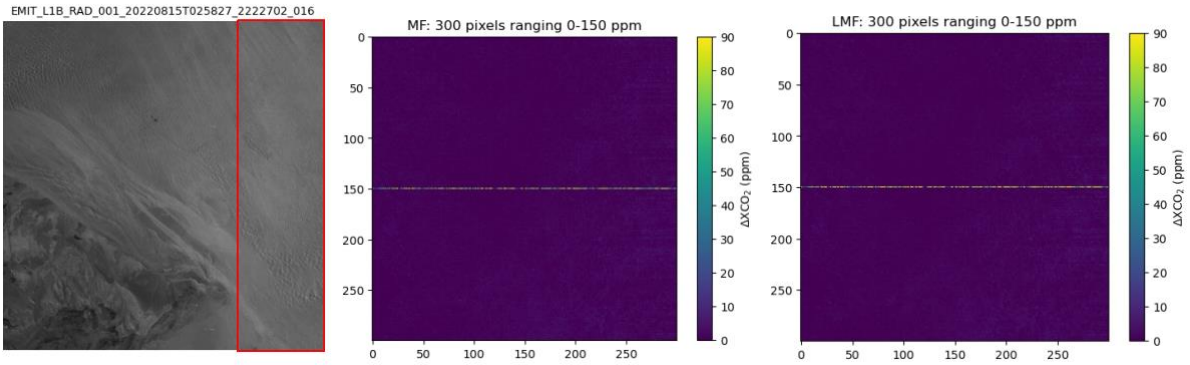


Figure 18. In the left panel, a ~1960 nm radiance band from an EMIT acquisition captured on 2022/08/15 in a desert area in China. Framed in red, the selected columns where the analysis is carried out. In the center and right panels, the MF and LMF retrievals, respectively, with a zoom to the line where the simulated enhancements were implemented.

The results from the 0-150 ppm range are shown in Figure 19, where we observe that the underestimation starts at 70 ppm with the LMF (right panel) and at 50 ppm with the MF (left panel). The underestimation is more pronounced in the MF case due to the linearization assumption (Pei et al, 2023). The reason for the underestimation in LMF needs further research. Thus, for both methodologies, a pronounced underestimation should not occur at typical ΔXCO_2 plume values.

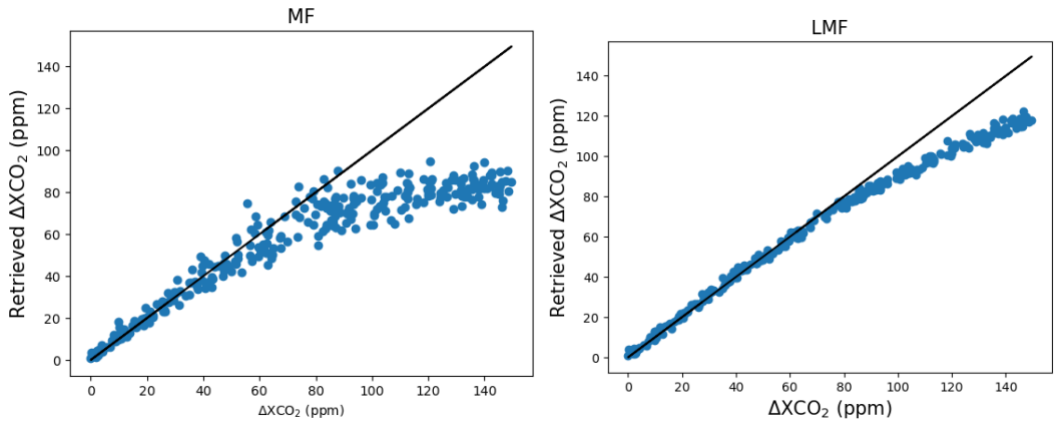


Figure 19. Scatter plots comparing the MF (left panel) and LMF (right) retrieval values from synthetic enhanced pixels against the true enhancement values considering the 0-150 ppm range.

The results from the 0-15 ppm range are shown in Figure 20, where we observe that these points are highly scattered and biased to higher values. However, those groups using matched-filter algorithms (SRON, UPV, and Kayrros) systematically obtained

lower flux rate values in comparison to the CEMS values (see Figure 9 to Figure 12). The poor retrieval precision from the retrieval values can indeed partially explain the scattered flux rate values, but not their systematic underestimation. Although GHGSat exhibits a lower number of points (only four plumes are quantified that align with CEMS-reported values), it seems to provide more accurate results in reference to CEMS values.

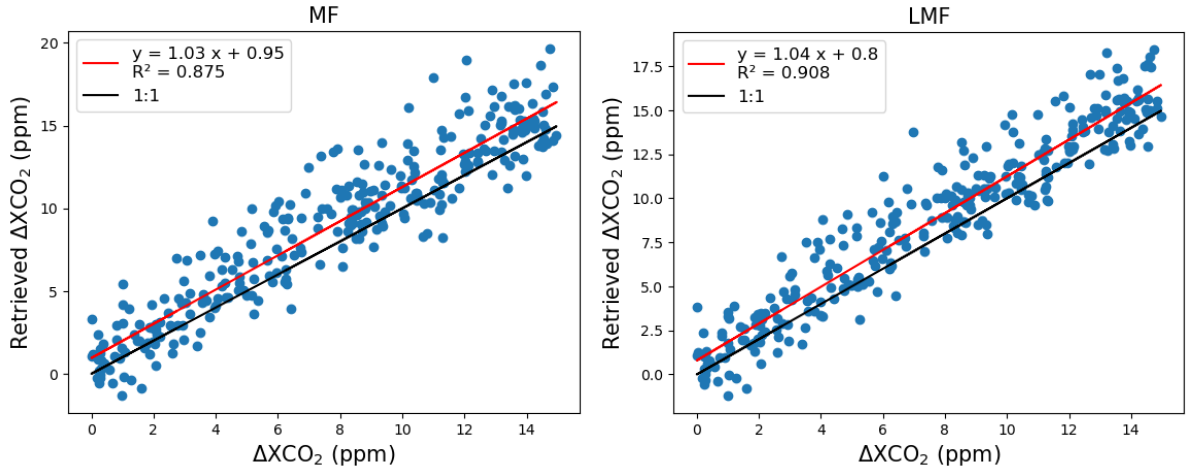


Figure 20. Scatter plots comparing the MF (left panel) and LMF (right) retrieval values from synthetic enhanced pixels against the true enhancement values considering the 0-15 ppm range.

Assuming these results can be generalized (EnMAP instrument, diverse range of surface types), the linearity assumption of the linear matched filter methods cannot explain the systematic underestimation in the emission estimates from public hyperspectral missions (EMIT, EnMAP). This, however, does not definitively rule out that other aspects of the matched filter method may produce biases in the L0 to L2 retrieval. Comparing the matched filter outputs with those of standard Differential Optical Absorption Spectroscopy (DOAS) inversion methods, which are used by GHGSat, could be an interesting follow-up investigation.

8 Conclusions and recommendations

This report evaluates the potential of 3 satellites (EMIT, EnMAP, GHGSat-C10) with high spatial and spectral resolution, to detect and quantify carbon dioxide emissions from large point sources. A set of 10 sites including fossil fueled power plants, cement factories, and refineries, located in geographically diverse areas, were surveyed for a total of 100 observations (10 sites, 3 satellites, and around 3 observations each). Data from the two public hyperspectral missions (EMIT and EnMAP) were independently processed by teams from 3 organizations (Kayrros, SRON, UPV), whereas GHGSat-C10 data was processed by GHGSat.

Based on this sample, we conclude that the 3 instruments demonstrated clear potential for satellite-based monitoring of CO_2 emissions, with some overlap and complementarity in current capabilities, but also significant limitations that should be addressed in the future to harness the full potential of this technology.

EnMAP demonstrated the lowest detection threshold (~4.4 MT/year at 50% probability of detection), followed by EMIT (~8.2 MT/year) and GHGSat-C10 (~18.6 MT/year), based on the subset of observations where high-fidelity stack-level measurements were available (US-located power plants). The chosen set of sites included significant geographic and topographic challenges, and those estimates are likely conservative, as opposed to the best performance achievable under favourable conditions.

This study also revealed limitations of the L4 methods (emission quantification), which were developed for methane emissions estimation and applied with minor modifications to the carbon dioxide case. EMIT and EnMAP systematically underestimated emissions by 35–65% compared to ground truth data, with regression slopes around 0.45. GHGSat-C10 showed better agreement (slope ~0.84), though based on fewer data points (n=5 quantified emissions). The use of 10-meter wind speed is thought to be a major driver of this underestimation, which should be addressed by calibrating the effective wind speed (U_{eff}) using WRF-LES simulations that account for stack height and plume buoyancy, and possibly expressing U_{eff} as a function of the 100-meter wind speed.

Note that these limitations of the effective wind speed calculation cannot explain the entirety of the underestimation, nor the fact that GHGSat quantifications were less biased than those of public hyperspectral missions. Investigations of the L0 to L2 processing are recommended as follow-up work, including for example a comparison of the matched filter outputs with those of standard DOAS inversion methods, analogous to those used by GHGSat.

We also note that the conclusions of this study may be more representative of power plants, especially coal-fired ones, from which we saw more plumes due to their larger emission rates. Expanding the dataset to improve statistical confidence and generalizability should be an additional follow-up step.

As a closing remark, we wish to highlight the general value of conducting this type of validation and intercomparison studies, and the overall success of the approach in identifying the shortcomings of current retrieval methods and the key areas of investigations for future work. References

9 References

[1] Cusworth, D. H., Duren, R. M., Thorpe, A. K., Eastwood, M. L., Green, R. O., Dennison, P. E., et al. (2021). Quantifying global power plant carbon dioxide emissions with imaging spectroscopy. AGU Advances, 2, e2020AV000350.

[2] Cusworth, D.H., Thorpe, A.K., Miller, C.E., Ayasse, A.K., Jiorle, R., Duren, R.M., Nassar, R., Mastrogiacomo, J.-P., Nelson, R.R., 2023. Two years of satellite-based carbon dioxide emission quantification at the world's largest coal-fired power plants. Atmos. Chem. Phys. 23, 14577–14591. https://doi.org/10.5194/acp-23-14577-2023

[3] Thorpe, A., Green, R., Thompson, D., Brodrick, P., Chapman, J., Elder, C., Irakulis Loitxate, I., Cusworth, D., Ayasse, A., Duren, R.,Frankenberg, C, Guanter, L., Worden, J.R., Dennison, P.E., Roberts, D.A., Chadwick, K.D., Eastwood, M.L., Fahlen, J.E. and Miller, C., Attribution of individual methane and carbon dioxide emissions sources using EMIT observations from space, Sciences Advances, doi/10.1126/sciadv.adh2391 (2023)

[4] Couture, H.D., Alvara, M., Freeman, J., Davitt, A., Koenig, H., Rouzbeh Kargar, A., O'Connor, J., Söldner-Rembold, I., Ferreira, A., Jeyaratnam, J., Lewis, J., McCormick, C., Nakano, T., Dalisay, C., Lewis, C., Volpato, G., Gray, M., McCormick, G., 2024. Estimating Carbon Dioxide Emissions from Power Plant Water Vapor Plumes Using Satellite Imagery and Machine Learning. Remote Sensing 16, 1290. https://doi.org/10.3390/rs16071290

[5] Tang, L., Xue, X., Qu, J., Mi, Z., Bo, X., Chang, X., Wang, S., Li, S., Cui, W., Dong, G., 2020. Air pollution emissions from Chinese power plants based on the continuous emission monitoring systems network. Sci Data 7, 325. https://doi.org/10.1038/s41597-020-00665-1

[6] A. Berk, P. Conforti, R. Kennett, T. Perkins, F. Hawes and J. van den Bosch, "MODTRAN® 6: A major upgrade of the MODTRAN® radiative transfer code," *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Lausanne, Switzerland, 2014, pp. 1-4, doi: 10.1109/WHISPERS.2014.8077573.

[7] Jervis, D., McKeever, J., Durak, B.O.A., Sloan, J.J., Gains, D., Varon, D.J., Ramier, A., Strupler, M., Tarrant, E., 2021. The GHGSat-D imaging spectrometer. Atmos. Meas. Tech. 14, 2127–2140. https://doi.org/10.5194/amt-14-2127-2021

[8] Frankenberg, C., Thorpe, A.K., Thompson, D.R., Hulley, G., Kort, E.A., Vance, N., Borchardt, J., Krings, T., Gerilowski, K., Sweeney, C., Conley, S., Bue, B.D., Aubrey, A.D., Hook, S., Green, R.O., 2016. Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region. Proc. Natl. Acad. Sci. U.S.A. 113, 9734–9739. https://doi.org/10.1073/pnas.1605617113

[9] Varon, D.J., Jacob, D.J., McKeever, J., Jervis, D., Durak, B.O.A., Xia, Y., Huang, Y., 2018. Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes. Atmos. Meas. Tech. 11, 5673–5686. https://doi.org/10.5194/amt-11-5673-2018

[10] Guanter, L., Irakulis-Loitxate, I., Gorroño, J., Sánchez-García, E., Cusworth, D. H., Varon, D. J., Cogliati, S., & Colombo, R. (2021). Mapping methane point emissions with the PRISMA spaceborne imaging spectrometer. *Remote Sensing of Environment, 265*, 112671. https://doi.org/10.1016/j.rse.2021.112671

[11] Roger, J., Guanter, L., Gorroño, J., and Irakulis-Loitxate, I.: Exploiting the entire near-infrared spectral range to improve the detection of methane plumes with high-resolution imaging spectrometers, Atmos. Meas. Tech., 17, 1333–1346, https://doi.org/10.5194/amt-17-1333-2024, 2024.

[12] Ouerghi, Elyes & Ehret, T. & Facciolo, Gabriele & Meinhardt, E. & Morel, Jean-Michel & de Franchis, Carlo & Lauvaux, Thomas. (2023). Methane Plumes Detection on Prisma L1 Images with the Adjusted Spectral Matched Filter and Wind Data. 7598-7601. 10.1109/IGARSS52108.2023.10282211.

[13] Pei, Z., Han, G., Mao, H., Chen, C., Shi, T., Yang, K., Ma, X., & Gong, W. (2023). Improving quantification of methane point source emissions from imaging spectroscopy. *Remote Sensing of Environment, 295*, 113652. https://doi.org/10.1016/j.rse.2023.113652

[14] Kochanov, R. V, Gordon, I.E., Rothman, L.S., Wcisło, P., Hill, C., Wilzewski, J.S., 2016. HITRAN Application Programming Interface (HAPI): A comprehensive approach to working with spectroscopic data. J. Quant. Spectrosc. Radiat. Transf. 177, 15–30.

[15] Department of Climate Change, Energy, the Environment and Water (DCCEEW). (2024). *Australian National Greenhouse Accounts Factors*. Canberra: Commonwealth of Australia. https://www.dcceew.gov.au/climate-change/publications/nationalgreenhouse-accounts-factors

[16] Conrad, Bradley M., David R. Tyner, and Matthew R. Johnson. “Robust Probabilities of Detection and Quantification Uncertainty for Aerial Methane Detection: Examples for Three Airborne Technologies.” Remote Sensing of Environment 288 (April 2023): 113499. https://doi.org/10.1016/j.rse.2023.113499.

[17] Sherwin, Evan D., Jeffrey S. Rutherford, Yuanlei Chen, et al. “Single-Blind Validation of Space-Based Point-Source Detection and Quantification of Onshore Methane Emissions.” Scientific Reports 13, no. 1 (2023): 3836. https://doi.org/10.1038/s41598-023-30761-2.

***** End of Document *****