METHODOLOGY REPORT FOR THE REPRESENTATIVE DATASET

EO4SEM

Prepared by:	GMV FMI GHGSat LSCE ONERA Science Partners
Code:	D1 - Methodology Report for the Representative Dataset
Version:	2.0
Date:	2025/10/29
Internal code:	GMV 27399/24 V2/25

THIS PAGE IS INTENTIONALLY LEFT BLANK

APPROVAL

Title: Methodology Report for the Representative Dataset	
Issue Number: 2	**Revision Number: N/A**
Author(s): EO4SEM Team	**Date:** 2025/10/29
Approved by: GMV	**Date of Approval:** 2025/10/29

CHANGE LOG

Reason for change	Issue	Revision	Date
Update for EO4SEM MTR	2	1	2025/10/29

CHANGE RECORD

Issue Number:	**Revision Number:**		
Reason for change	**Date**	**Pages**	**Paragraph(s)**
N/A			

DISTRIBUTION

Name/Organisational Unit:

TABLE OF CONTENTS

1. INTRODUCTION 6
1.1. Scope 8
1.2. Reference documents 8
1.3. Acronyms 8
2. STAKEHOLDER'S NEEDS AND REQUIREMENTS 11
2.1. GHG policy frameworks 11
2.2. Stakeholder requirements 12
3. REPRESENTATIVE DATASET 13
3.1. Data gathering and processing 13
3.1.1. Ship Emission Model Data 13
3.1.2. TROPOMI Proxy Data 16
3.1.3. Hyperspectral Proxy Data 18
3.1.4. GHGSat Proxy Data 19
4. APPLICATION DEVELOPMENT PRINCIPLES 22
4.1. ESA APEx Platform 22
4.1.1. Geospatial explorer 22
4.1.2. APEx Product Catalogue 23
4.1.3. APEx Web Portal 24
5. REPRESENTATIVE DATASET PACKAGING 26
5.1. Data Exploration and analysis 26
5.2. Data Processing 27
5.3. Metadata File 28
5.4. Technical Monitoring using Gitlab 29
6. ANNEX A: BIBLIOGRAPHIC REFERENCES 30

LIST OF FIGURES

Figure 3.1: The working principle of STEAM (source: (Johansson et al., 2017)). Yellow indicates input data, green represents external factors, light blue denotes pre-processing and dark blue indicates individual model components........ 14
Figure .1: EO4SEM APEx cloud-based development platform diagram requested by GMV 22
Figure .: EO4SEM Geospatial explorer 23
Figure . Screenshot of EO4SEM Data Catalogue 24
Figure 4.4: Screenshot of EO4SEM project portal (more content coming soon) 25
Figure .1 a) Nox STEAM 2019 from FMI b) EO4SEM_AEMISSIONS_CH4_TROPOMI_WFMD_CAMSREGv8_201911 — ch4_posterior. The principle of the traditional atmospheric inversion approach is to correct a priori estimates of the emission maps, also denoted "prior emissions", to reduce differences between atmospheric observations and their simulations with a CTM fed with the estimates of emission maps. The prior estimates of NOx and CH4 anthropogenic emissions in this study are based on emission estimates from the CAMSREG-v8 gridded inventory for the year 2019. We assimilate observations from the TROPOMI RPRO-v020400 for NO2 and from TROPOMI-SRON or TROPOMI-WFMD for CH4. Two types of files will be available from LSCE/Science Partners for the regional inversions of CH4 and NOx maritime emissions. **c)** Plume detected by GHGSAT near the port of Athens. 26

Figure .2: Function to to generate JSON objects for adding layers/items to the Catalogue Viewer (Geospatial Explorer) 27
Figure .3 Adding coordinate reference system to input data 28
Figure .4 Data analysis showing not valid COGs....28
Figure .5 Conversion of data do valid COGs....28
Figure .6 Data description Excel file 29
Figure .7: EO4SEM GitLab repository....29

LIST OF TABLES

Table 1.1: Reference documents.... 8
Table 1.2: List of acronyms.... 8
Table 2.1: List of ports shared by EMSA....12
Table 3.1: Key parameters of GHGSat constellation 20

1. INTRODUCTION

The maritime transport sector is pivotal to global trade, facilitating 90% of goods movement worldwide and underpinning critical industries like tourism and fisheries. Within the European Union (EU), it supports 75% of international trade, making it integral to the EU's economic framework. However, the sector is a significant contributor to environmental challenges, responsible for 3-4% of global anthropogenic CO_2 emissions and 14% of the EU's transport emissions (~135.9 million tonnes in 2022).

To mitigate this impact, the EU has introduced ambitious regulatory frameworks under the Green Deal[1] and Fit-for-55[2] package. These include the extension of the Emission Trading System[3] (ETS) to maritime transport, the FuelEU Maritime regulation[4], and revisions to the Monitoring, Reporting, and Validation [5](MRV) regulation. Together, they aim to:

- Establish carbon pricing for maritime emissions starting in 2026.
- Reduce the Greenhouse Gas (GHG) intensity of ship energy use.
- Integrate stricter monitoring for GHG such as methane (CH_4) and nitrous oxide (N_2O).

Despite these initiatives, effective implementation hinges on robust, independent data collection and monitoring capabilities to ensure compliance and track progress towards emission reduction targets.

To this end, key priorities include monitoring individual vessel emissions (CO_2, CH_4, and N_2O) per voyage, aggregated emissions at regional and national levels, and fugitive emissions from ships. Additionally, there is a need to track other contributors to global warming, such as black carbon and Volatile Organic Compounds (VOCs), to support comprehensive enforcement of EU and international policies.

Given the existing regulatory context, in the context of EO4SEM, Earth Observation (EO) data is being leveraged to assess the potential of current and future EO instruments (and requirements for the future EO instruments) to support such an improvement of the knowledge on shipping emissions.

EO data is being used as a complementary source to existing engine-driven emission models, such as FMI's STEAM v4, which relies on AIS (Automatic Identification System) data streams to estimate emissions at ship level. While effective, these models are limited by AIS availability and cannot account for non-reporting ships, such as those with disabled AIS or naval vessels.

EO imagery can help bridge these gaps by providing independent data for emission monitoring and validation, and in the scope of this project, is being used with two interconnected goals in mind:

- TROPOMI (on-board Sentinel-5), PRISMA, EnMAP and GHGSat observations (together composing the reference dataset) are being processed with inversion models to refine emission estimates for pollutants and GHGs (NO_x, CO_2, and CH_4) over the sea. These emission estimates are being aggregated temporally and spatially to create accurate maps of ship-driven pollutants and GHGs across maritime corridors, local regions and ports.
- To create a reference dataset and the respective new developments and methodologies to be used as benchmark, helping to guide the development of technical specifications and consolidate the added value that the Copernicus Expansion Missions could represent in the future for this high priority area, more precisely CO2M and CHIME.

To support stakeholders' needs, the EO-driven approach will provide a benchmark for automated and scalable emission analysis, reducing manual effort while increasing reliability. It will also help assess emission patterns near ports and industrial areas, even if as a secondary objective. Finally, the entire service will be deployed on an ESA Network of Resources (NoR) platform, ensuring interoperability and ease of access for users.

---

[1] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en
[2] https://www.consilium.europa.eu/en/policies/fit-for-55/
[3] https://eur-lex.europa.eu/eli/reg/2023/957/oj
[4] https://eur-lex.europa.eu/eli/reg/2023/1805/oj
[5] https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets/monitoring-reporting-and-verification-eu-ets-emissions_en

1.1. SCOPE

This document, Deliverable 1 – Methodology Report for the Representative Dataset, outlines the methodology used to develop and validate the so-called "Representative Dataset".

This is Version 2 of the report, which will be updated again for the final review.

1.2. REFERENCE DOCUMENTS

The following documents, although not part of this document, amplify or clarify its contents. Non-academic documents are referenced in the form [RD.x]. Academic references are listed in Annex A.

Table 1.1: Reference documents

Ref.	Document ID.	Title	Rev.
[RD1]	ESA Contract 4000145881/24/I-DT-bgh – EO4SEM_signed	ESA contract No. 4000145881/24/I-DT-bgh with GMV Skysoft S.A. Sentinel Users Preparation (SUP) initiative - [SUP-1] - Applications Preparedness with stakeholder and end-users participation. Project: EO for Ship Emission Monitoring (EO4SEM).	NA
[RD2]	GMV 10933/24 V1/24	Full Proposal for EO4SEM Project – “EO for Ship Emission Monitoring”.	1.0

1.3. ACRONYMS

Table 1.2: List of acronyms

Acronym	Description
AIS	Automatic Identification System
AK	Averaging Kernel
APEx	Application Propagation Environment
API	Application Programming Interface
CAMS	Copernicus Atmosphere Monitoring Service
CDS	Copernicus Climate Data Store
CH_4	Methane
CMEMS	Copernicus Marine Environmental Monitoring Service
CNMF	Constrained Nonnegative Matrix Factorisation
CO	Carbon Monoxide
CO_2	Carbon Dioxide
COG	Cloud Optimized Geotiff
CRS	Coordinate Reference System
EMSA	European Maritime Safety Agency
ENMAP	Environmental Mapping and Analysis Program
EO	Earth Observation
ETS	Emission Trading System

Acronym	Description
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FMI	Finnish Meteorological Institute
GHG	Greenhouse Gas
GNSS	Global Navigation Satellite System
HTTP	Hypertext Transfer Protocol
IME	Integrated Mass Enhancement
IMO	International Maritime Organisation
LNG	Liquefied Natural Gas
MMSI	Maritime Mobile Service Identity
MRV	Monitoring, Reporting, and Validation
MSI	MultiSpectral Instrument
N_2O	Nitrous Oxide
.nc	NetCDF
NIR	Near Infrared
NoR	Network of Resources
NO_x	Nitrogen Oxides
NVOC	Non-Volatile Organic Compound
ODS	Ozone Depleting Substance
OGC	Open Geospatial Consortium
PRISMA	PRecursore IperSpettrale della Missione Applicativa
QA	Quality Assurance
RD	Reference Document
S2	Sentinel-2
S5P	Sentinel-5 Precursor
SNR	Signal-to-Noise Ratio
SO_x	Sulphur Oxides
SRON	Space Research Organisation Netherlands
STAC	SpatioTemporal Asset Catalogues
STEAM	Ship Traffic Emission Assessment Model
SWIR	Short-Wave InfraRed
SVOCs	Semi-Volatile Organic Compounds
TROPOMI	TROPOspheric Monitoring Instrument
TVCD	Tropospheric Vertical Column Density
UTM	Universal Transverse Mercator

Acronym	Description
UV	Ultraviolet
VOC	Volatile Organic Compound
WAF-P	Wide-Angle Fabry-Perot
WFMD	Weighting Function Modified Differential
XCH_4	CH_4 total column average mole fraction
XCO_2	CO_2 total column average mole fraction

2. STAKEHOLDER'S NEEDS AND REQUIREMENTS

All the work developed in the scope of EO4SEM is being done in collaboration with the project stakeholder – EMSA (European Maritime Safety Agency). EMSA is a European Union agency, established in 2002 to enhance maritime safety, security, and environmental protection across EU waters. This is done by providing technical expertise, operational support, and policy implementation assistance to EU Member States and the European Commission, having as primary role to ensure that maritime transport remains safe, sustainable, and efficient while addressing environmental and security challenges associated with the maritime industry.

EMSA's responsibilities span a wide range of activities, including monitoring ship traffic through satellite-based systems like SafeSeaNet[6], coordinating responses to maritime pollution incidents, and supporting compliance with EU maritime legislation.

In alignment with the **EU Green Deal**, EMSA is increasingly focused on **monitoring and reducing greenhouse gas (GHG) emissions from ships**. EO4SEM supports this mission by leveraging EO data to enhance EMSA's capabilities in the following areas:

- **Emission Hotspot Detection** EO data can help identify areas with high concentrations of ship emissions (e.g., NO_2, SO_2, CO_2) using satellite-based atmospheric sensors. This enables EMSA to prioritize inspections and interventions in critical zones.
- **Validation of Ship-Reported Data** Ships are required to report fuel consumption and emissions data under the **EU MRV (Monitoring, Reporting and Verification)** regulation that requires ships over 5,000 gross tonnage (GT) to monitor, report, and have their CO2 and other relevant emissions verified for voyages within, into, or out of European Economic Area (EEA) ports. EO data provides an independent source to **cross-validate** these reports, improving data integrity and compliance.
- **Temporal and Spatial Analysis of Emissions** EO data allows EMSA to track emissions over time and across regions, helping to assess the effectiveness of policy measures and identify trends in maritime pollution. This could range between hourly to yearly composites of data.
- **Support for Enforcement and Policy Development** EO-derived insights can be used to **support enforcement actions** against non-compliant vessels and to **inform future policy decisions**, such as emission control areas (ECAs) or low-emission zones.
- **User ready data** EMSA requires EO data to be user-ready format, ensuring seamless usability for operational teams. In this aspect, potential use of APEx (https://apex.esa.int/) along with its Geospatial platform should support this need.
- **Scalability and Sustainability** EMSA seeks EO solutions that are **scalable across EU waters** and **sustainable in terms of cost and data continuity**, ensuring long-term utility beyond the EO4SEM project.

2.1. GHG POLICY FRAMEWORKS

The critical interest of EMSA in GHG emissions is related to the fact that the Agency supports the extension of the EU Emission Trading System (ETS) to maritime transport, a key regulatory initiative which introduces carbon pricing for CO_2 emissions from ships, gradually transitioning to cover 100% of maritime emissions by 2026.

Also, EMSA facilitates the implementation of the FuelEU Maritime regulation, which aims to reduce the GHG intensity of fuels used by ships and promote the adoption of sustainable alternatives. This regulation includes targets for energy efficiency and requirements for the use of shore-side electricity in ports to minimise emissions during docking.

Considering this regulatory framework, there is a need from EMSA's side to assess the trustworthiness of emissions reported and the sector's progress towards both EU and global targets. Monitoring individual ship emissions per voyage, as well as total emissions on monthly, seasonal, and yearly scales at European, regional, and local levels, is essential for effective implementation and enforcement of regulatory measures.

[6] https://www.emsa.europa.eu/ssn-main.html

2.2. STAKEHOLDER REQUIREMENTS

In line with the context presented above, EMSA has identified a set of high-level requirements that are being used to frame the technical developments of the project. These requirements include monitoring individual vessels' GHG emissions per voyage (CO_2, CH_4, and NO_2), tracking monthly, seasonal, and yearly total emissions from the maritime transport sector at various scales, and detecting CH_4 leaks/slips along routes and at ports. Additional focus areas include monitoring other global warming contributors such as black carbon, ozone-depleting substances (ODS), and both volatile and non-volatile organic compounds (VOC and NVOC) from maritime sources.

To support these objectives, EO4SEM is concentrating on three key pollutants: NOx, CO_2, and CH_4. For CH_4, the emphasis will be on LNG carrier routes, particularly those connecting to regasification plants.

For NOx, the focus is on comparing emissions within designated NOx Emission Control Areas (NECAs) such as the Baltic Sea and the North Sea, against non-NECA regions like the Mediterranean Sea.

Regarding CO_2, the strategy involves identifying segments of major shipping routes—such as the corridor from the Kattegat through the English Channel, down to the Gibraltar Strait, and across to the Suez Canal—where ship-derived CO_2 signals can be isolated from land-based anthropogenic emissions.

EMSA has also referred to their interest on monitoring port activities, namely potential methane leaks from liquified natural gas (LNG) ships at anchor due to a combination of boil-off and uncombusted methane from ship engines, and leaks during LNG transfer either from shore-to-ship or ship-to-shore (the more likely scenario in Europe). In this regard, a list of ports in which EMSA has an interest have been shared by the agency:

Table 2.1: List of ports shared by EMSA

Name	Locode	Lat	Lon
Marseille	FRMRS	43.3982	4.9977
Fos-sur-Mer	FRFOS	43.4087	4.8902
Port-de-Bouc	FRPDB	43.4029	4.9776
Barcelona	ESBCN	41.3422	2.1632
Algeciras	ESALG	36.1475	-5.4281
Huelva	ESHUV	37.1594	-6.8959
Piraeus	GRPIR	38.0248	23.5936
Agioi Theodoroi	GRAGT	37.9273	23.1422

3. REPRESENTATIVE DATASET

This chapter details the collection, processing, development, and validation of the various dataset that are being included in the project's Representative Dataset. As a foundational component of the project, this dataset aims to enhance the measurement of GHG emissions from maritime sources. It will be updated and expanded throughout the project.

The Representative Dataset integrates data from TROPOMI, PRISMA and EnMap hyperspectral, Sentinel-2, and GHGSat to assess the potential added value of future satellite missions under the Copernicus Expansion programme, specifically CO2M and CHIME.

Its primary goal is to evaluate the potential performance of these missions in detecting and quantifying GHG emissions compared to established ship emissions reference datasets (e.g., STEAM). This assessment will help refine satellite technology and improve future monitoring capabilities.

The idea to proceed in this direction is the usage of the existing EO data (previously referred missions) and use it as a proxy for regional and local-scale analyses of CO_2, CH_4, and NO_x emissions. In more detail:

- **CO2M mission**: the analysis of TROPOMI and GHGSat data will provide key insights into CO2M's capabilities, and the specifications needed for future missions to effectively monitor shipping emissions.
- **CHIME mission**: CHIME's data characteristics - swath width, spectral resolution, and spatial resolution - will be simulated over port complexes and open-sea regions using debris reflectance products from PRISMA and Sentinel-2. The COMANCHE radiative transfer model (Miesch et al., 2005) will generate CHIME-like radiance at the top of the atmosphere, allowing for a sensitivity analysis of maritime traffic emission plumes. Where relevant, other simulated products will be developed for the studied areas, incorporating atmospheric spatiotemporal variability linked to emissions modelling and observations.

3.1. DATA GATHERING AND PROCESSING

3.1.1. SHIP EMISSION MODEL DATA

Overview of the Ship Emission Model data.

The Ship Traffic Emission Assessment Model (STEAM), developed by the Finnish Meteorological Institute (FMI), is a state-of-the-art tool for evaluating fuel consumption and estimating ship emissions with high spatial and temporal resolution (Jalkanen et al., 2009, 2012, 2016; Johansson et al., 2013, 2017). STEAM combines AIS (Automatic Identification System)–derived activity data with technical insights into the global fleet and fundamental naval architecture principles (Jalkanen et al., 2009, 2012, 2014; Johansson et al., 2017).

AIS data encompasses key ship attributes, including International Maritime Organisation (IMO) and Maritime Mobile Service Identity (MMSI) numbers, name and ship type, latitude, longitude, heading and speed over ground, course over ground, allowing precise tracking of vessel activity. When AIS messages show the presence of vessels with newly assigned IMO numbers, a query is directed to S&P Global. If the vessel appears in the database, its details are obtained and inserted into the internal ship database prior to executing the emission estimation programme.

STEAM's internal database consists of detailed ship specifications, which enable calculations such as instantaneous power use, travel routes, ship speed, engine load, emission factors, emissions, use of shaft generators, alongside information on abatement technologies, fuel types, and specific fuel oil consumption (Jalkanen et al., 2009, 2012). To reduce uncertainties arising from operational variability, the model incorporates environmental factors like wind, sea currents, waves, and sea ice, enhancing its capacity to estimate the impact of ambient conditions on vessel performance.

In principle, the model operates in three stages as illustrated in Figure 3.1.

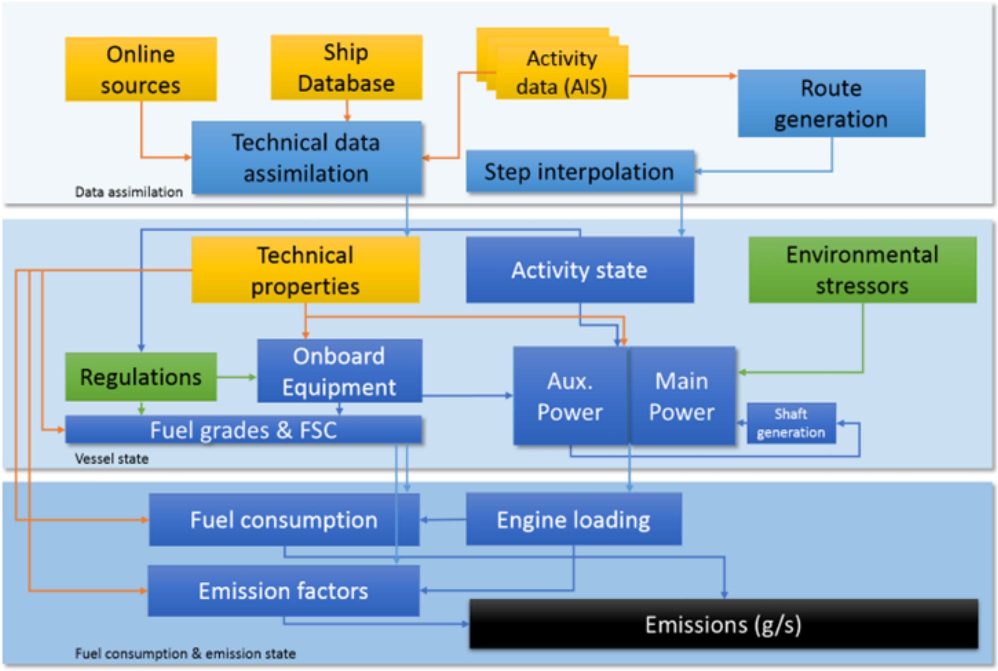

Figure 3.1: The working principle of STEAM (source: (Johansson et al., 2017)). Yellow indicates input data, green represents external factors, light blue denotes pre-processing and dark blue indicates individual model components

In the initial stage, relevant data is collected from multiple sources, combined, and pre-processed. Ship activity data is organised chronologically, with route-generation algorithms addressing spatial and temporal gaps. Missing technical details in the vessel data are replaced with information from online databases, static AIS data, or algorithms estimating values based on similar vessels.

The second stage evaluates the vessel's operational state on a minute-by-minute basis to estimate engine power requirements, fuel types, and sulphur content. Factors such as location, speed, operational mode, and environmental influences (e.g., waves, sea currents, ice cover) are integrated to determine power demands. Auxiliary power needs are evaluated based on configurations like shaft generators or diesel-electric systems, considering the operation mode of each vessel. In addition, local emission regulations and the use of abatement technologies, such as open-loop scrubbers, are incorporated into fuel-grade selection and sulphur content assessment.

In the final stage, engine load, fuel consumption, and emissions are calculated. The model determines which engines are active, their load distribution, and computes instantaneous fuel usage and exhaust emissions based on engine specifications, fuel type, and abatement technologies. Instantaneous quantities are integrated over space and time to generate a fully dynamic description of emissions.

A more comprehensive description of the working principles of STEAM, including the assessment of technical specifications of vessels and the evaluation of AIS data and processing, can be found in (Johansson et al., 2017) and (Johansson, 2011), respectively.

For this project, STEAM (version 4.3) integrates AIS data with technical specifications of over 120,000 ships to provide ship emission reference data and evaluate emission of key pollutants, including carbon dioxide (CO_2), methane (CH_4), nitrogen oxides (NO_x), at high spatial and temporal resolutions. The number of active AIS transmitters exceeds the number of vessels in the global commercial shipping fleet because small boats may also carry AIS. In those cases, smallest vessel category is assumed in STEAM.

The STEAM v4.3 model has shown high accuracy in aggregated annual fuel consumption predictions, with a mean prediction error of 9 %, reducing to 2% in large-scale comparisons. For individual vessels, the average accuracy of annual fuel consumption prediction is around 20%, based on the Monitoring, Reporting and Verification data over the last five years. However, uncertainties remain in regional-scale emission inventories due to gaps in AIS data, variability in fuel sulphur content, and the chemical characterisation of particulate matter.

Within the EO4SEM project, STEAM outputs serve as an important reference dataset to test and validate EO-derived emission estimates, thus, establishing a robust foundation for advancing satellite-based emissions monitoring.

Ship Emission Model data gathering and processing.

The STEAM model is employed to estimate emissions by leveraging vessel activity data obtained from the Orbcomm Ltd. AIS data archive. Additionally, the technical specifications of the global fleet are sourced from the S&P Global databases. These datasets provide essential information for modelling ship power load, fuel consumption, and emissions, forming a reliable foundation for assessing maritime pollution.

To enhance the accuracy of emissions estimates, the model incorporates the influence of various environmental factors, including wind patterns, sea waves, sea ice coverage, and ocean currents taken from the Copernicus Climate Data Store (CDS) and the Copernicus Marine Environmental Monitoring Service (CMEMS). The integration of these factors enables a more precise assessment of how varying sea and wind conditions impact fuel consumption and subsequent emissions from shipping activities.

The specific datasets used for ambient conditions include daily sea ice thickness and 6-hourly updated sea currents direction and speed, obtained from Global Ocean Physics Analysis and Forecast (CMEMS). The 10m wind direction and speed are obtained from ERA5, as are the 3-hourly wave direction and wave height information.

Emission factors within STEAM have been refined to reflect advancements in engine technology, fuel types and emission mitigation strategies. The model now includes updated factors for nitrogen oxides (NO_x), carbon monoxide (CO), carbon dioxide (CO_2), sulphur oxides (SO_x), and particulate matter with diameters of 2.5 µm or less (PM2.5). Details of emission factor assignments can be found in (Grigoriadis et al., 2021). Furthermore, the emission parameters have been expanded to accommodate alternative marine fuels, such as methanol and ammonia. The model also provides more detailed emission factors for vessels equipped with abatement technologies, as detailed in (Majamäki & Jalkanen, 2022). These enhancements ensure that STEAM remains aligned with contemporary regulatory and technological developments in maritime emissions monitoring.

For the first analyses, the Finnish Meteorological Institute (FMI) generated, using the STEAM model, a high-resolution dataset of maritime emissions for 2023. This dataset includes comprehensive emission inventories for carbon dioxide (CO_2), methane (CH_4), and nitrogen oxides (NO_x) across the European maritime domain (15.25°W to 40°E; 25°N to 74.25°N). The outputs were produced at hourly temporal resolution and included both gridded binary data and point-source emissions data. For regional-scale simulations, emissions were resolved at a spatial resolution of 0.5° grid cells, while point-source data were georeferenced to specific vessel locations using GPS coordinates. geolocated to individual vessel positions based on GPS data. Upon request, additional simulations were conducted for the year 2023 at finer spatial resolution (0.05° × 0.1°) to assess the benefits of higher spatial detail.

Higher-resolution (0.05° × 0.1°) STEAM gridded datasets for 2022 were also produced, excluding small ships. FMI will subsequently provide both point-source data excluding small ships and gridded data including small ships for 2022.

The overall objective is to run the STEAM model at a finer spatial resolution of 0.05° × 0.1° for the period 2019–2023 to support detailed evaluation and comparison of emission datasets derived from Earth observation.

The integration of STEAM outputs into the EO4SEM project will not only provide a robust reference layer but will also aim to enhance the scientific validity of EO-based emissions detection. By combining these datasets, the project is addressing key challenges in maritime emission monitoring and associated uncertainties, paving the way for accurate, scalable solutions that align with regulatory and environmental goals.

Challenges, gaps, and lessons learnt.

A key challenge concerns the generation of consistent high-resolution datasets that include both large and small vessels. While STEAM can provide high-resolution (0.05° × 0.1°) gridded emission datasets covering all ships, the point-source data will include only IMO-registered vessels, excluding small ships. This requires additional STEAM simulations to ensure compatibility across datasets.

Another challenge relates to maintaining data confidentiality while preserving analytical integrity. For the point-source dataset, FMI will run STEAM with full ship identities and subsequently replace IMO numbers with the same random identifiers for the 2019–2023 period. This process, supported by an automated script, ensures traceability and reproducibility while safeguarding sensitive vessel information.

3.1.2. TROPOMI PROXY DATA

Overview of the TROPOMI instrument and data.

The TROPOspheric Monitoring Instrument (TROPOMI) was launched onboard the Copernicus Sentinel-5 Precursor (S5P) satellite in October 2017. It flies at an altitude of 824 km in a sun-synchronous orbit that crosses the Equator at approximately 13:40 LT. The instrument is an imaging spectrometer that measures in the Ultraviolet (UV)-visible, the Near Infrared (NIR) and the shortwave infrared range (2314 – 2382 nm). This imaging spectrometer supports the derivation of NO_2 Tropospheric Vertical Column Densities (TVCDs), SO_2 total columns, and CH_4 total column average mole fractions (XCH_4), among observations for other species.

The nominal footprint of the TROPOMI ground pixels at nadir for NO_2 and SO_2 was about 7×3.5 km^2 before 6^{th} August 2019 and 5.5×3.5 km^2 after 6^{th} August 2019. For XCH_4, the nominal footprint of the TROPOMI ground pixels at nadir was about 7.2×7.2 km^2 before 6^{th} August 2019, and 5.6×7.2 km^2 after this date. With a swath of about 2600 km on the ground, TROPOMI provides daily global coverage for NO_2, SO_2 and CH_4.

The regional scale analysis of NO_x (and subsequently CO_2, indirectly derived from the NO_x analysis and co-emission ratios associated with fossil fuel combustion) and CH_4 emissions, and the local scale analysis of NO_x (and subsequently CO_2) in the project will exploit the NO_2 and XCH_4 observations from TROPOMI. Some local scale analyses of the SO_2 observations (co-emitted with NO_x and CO_2 by fossil fuel combustion), in support of the derivation of CO_2 emissions estimates, might also be undertaken during the project. However, the current TROPOMI data gathering is focused on the NO_2 and XCH_4 products.

The spatial resolutions of the TROPOMI products are finer than those of previous satellite instruments dedicated to monitoring NO_2 TVCDs and XCH_4 observations. Several missions observing NO_2 and CH_4 from space with a large regular coverage that is suited for the regional scale analysis of the emissions and for the local scale analysis of the NO_x emissions have been launched after TROPOMI. However, the first geostationary missions (TEMPO and GEMS), to our knowledge, do not yield high quality NO_2 data yet and do not cover Europe, and the GOSAT-GW, Sentinel-4 and Sentinel-5 missions have just been launched in summer 2025 and do not yield NO_2 and XCH_4 observations yet. Currently, TROPOMI remains the reference for global multi-scale atmospheric monitoring of NO_x and CH_4 emissions. These resolutions are coarser than the 2×2 km^2 resolution of the future observations of NO_2, CO_2 total column average mole fractions (XCO_2) and XCH_4 of the CO2M mission. However, both TROPOMI and CO2M observations correspond to instruments with resolutions of a few kilometres. Therefore, investigations on the TROPOMI data will provide insights into the potential of the future CO2M observations at local and regional scales.

TROPOMI data gathering and processing for regional and local atmospheric inversions of NO_x (and subsequently CO_2)

TROPOMI-v02.04 to v02.06 NO_2 data has been collected, covering the period from 2019 to 2023, including TVCDs, and the corresponding retrieval errors and Averaging Kernels (AKs), retrieval pressure levels, and quality flags globally, even though the main analysis are focused on the maritime area of Europe. The first regional inversions have been performed for the year 2019, and the period has recently been extended to 2019-2023. The local scale analysis also covers the 2019-2023 period. If relevant, the version of the TROPOMI product used for these inversions will be updated regularly during the project.

The selection of the TROPOMI data to be assimilated in the inversions follows the criteria of van Geffen et al. (2022), with selected observations having a Quality Assurance (QA) value higher than 0.75. For the regional inversions, only observations with retrieval errors below 100% were selected.

The local scale analysis is performed at the spatial resolution of the observations. However, the regional scale inversions rely on comparisons between the observations and simulations with a configuration of the chemistry transport model CHIMERE for Europe at 0.5° resolution (this resolution may be refined during the project, and analyses using CHIMERE over other areas may also be considered).

To make suitable comparisons the simulated vertical profiles of NO_2 from CHIMERE are interpolated on the TROPOMI retrieval pressure levels. Then, the averaging kernels are applied to the simulated profiles to account for the vertical variations of the sensitivity of the satellite retrievals to the NO_2 concentrations within vertical tropospheric columns (van Geffen et al., 2022)).

As the spatial resolution of the TROPOMI data is finer than that of the CHIMERE configuration, the selected TROPOMI TVCDs are aggregated into "super-observations" at the model spatial resolution, as recommended by Rijsdijk et al. (2025). In order to associate the super-observations to an actual AK profile for a given CHIMERE 0.5°×0.5° grid-cell, the super-observations are taken as the observation (TVCD, retrieval uncertainty, pressure levels, and AKs) corresponding to the TVCD value closest to the mean of the TROPOMI TVCDs within the 0.5°×0.5° grid-cell and within the CHIMERE physical time step of about 5-10 minutes.

TROPOMI data gathering and processing for regional atmospheric inversions of CH_4

Different TROPOMI products are available for the regional scale analysis of CH_4 emissions overseas, including:

i) The operational TROPOMI product developed by the Space Research Organisation Netherlands (SRON), updated with iterative improvements of the beta research product (Lorente et al., 2021, 2022, 2023).

ii) The Weighting Function Modified Differential (WFMD) product from the Bremen University (Schneising, 2023, 2024).

However, there is a need for investigations (tests, analysis) to identify and select the most suitable one, and to define the observation selection criteria for monitoring emissions from sea shipping. Subsequently, if relevant, the version of the TROPOMI product used for these inversions will be updated regularly during the project.

The two products for the year 2019 have already been collected, including total columns (XCH_4) and the corresponding retrieval errors, averaging kernels, retrieval pressure levels, prior profiles, and quality flags globally, even though the main analyses focus on the maritime area of Europe. Initial tests and analyses for the selection of the most suitable product and, subsequently, for regional inversions are performed over the year 2019, and then the analysis period will be extended to 2019-2023.

The process of the model NO_2 simulations and of the TROPOMI NO_2 TVCD observations (their aggregation into super-observations at the model spatial resolution) detailed above for the regional analysis of the NO_x emissions is applied to the model (CHIMERE) CH_4 simulations and TROPOMI XCH_4 observations.

Emission products derived from the TROPOMI data.

The overall atmospheric inversion process used to derive emission estimates from the TROPOMI data will be fully documented in Deliverable 3 (Scientific Publication), which will have its final version delivered in the project's final review meeting.

In brief, the regional scale analyses for NO_x and CH_4 emissions consist in variational inversions at 0.5° and 1-day resolution based on the CIF-CHIMERE regional atmospheric chemistry-transport forward and inverse modelling frameworks detailed in Fortems-Cheiney et al. (2025) and in Sicsik-Paré et al. (2025). The current local-scale analyses of the NO_x emissions are based on three types of approach: two of them are based on the divergence method documented in Rey-Pommier et al. (2025) and the third is based on a plume cross sectional flux approach derived from the descriptions in Kuhlmann et al. (2024) and Santaren et al. (2025). The divergence approach is applied in a systematic way on TROPOMI orbits and the corresponding concentration fields to derive NO_x emission maps on the TROPOMI observation grid that are then subsampled and averaged to produce estimates of the average lineic NO_x emissions per major straight segments of the sea shipping lanes in Europe.

The divergence approach and the plume cross sectional flux approach are also used to derive instant estimate of the NO_x emissions from individual large ships. This estimation relies on the detection and extraction of large NO_2 plumes from large ships in the TROPOMI images, based on the crossing of the information on large emitters from the STEAM emission database provided by FMI (see section 3.1.1) with the identification of clusters of large NO_2 observations.

A challenge is the separation of the signal associated to the plumes from the background NO_2 field, which can be highly noisy or variable. All the CO_2 emission estimates are currently based on the NO_x emission estimates at all scale with a conversion based on the corresponding NO_x-to-CO_2 emission ratios from the STEAM emission database provided by FMI

The different types of CH_4, NO_x and CO_2 emission products currently generated and distributed from this overall process consist in:

- CH_4 and NO_x monthly maritime emission maps at 0.5° resolution over 2019 and the European domain 15.25°W to 40°E; 25°N to 74.25°N, and the corresponding monthly emission budget per large sea areas in domain. These maps are based on the regional scale inversions, and the mask of the major sea areas has been based on an adaptation of the International Hydrographic Organization (IHO) Sea Areas (https://www.marineregions.org/sources.php).
- NO_x and CO_2 monthly mean lineic emissions of major shipping lanes in the European domain 15.25°W to 40°E; 25°N to 74.25°N. The shipping lanes' locations and extent are defined based on masks extracted from an analysis of the STEAM emission database provided by FMI.
- Series of instant (~3-hour mean) estimates of NO_x and CO_2 emissions from individual large ships. Each estimate is shared with the corresponding geographical coordinates and time.

Validation of dataset

All the validation of the emission datasets so far is based on comparisons to the emissions estimates from FMI (based on the STEAM model).

If available during the project development, "external" datasets will be used as ground truth for further validation.

Challenges, gaps, and lessons learnt.

The potential biases and the noise in the TROPOMI NO_2 data over the seas, especially outside the sun glint areas of the satellite view (Riess et al., 2022), raise uncertainties for the local to regional scale analysis of the emissions. Furthermore, close to the coasts, the signal from the ships overlaps large signals from land emissions. As a result, at local scale, the detection of plumes from individual ships is often hampered by the noise and other patterns in the background field, and, at the scale of shipping lanes or at regional scale, the estimate of sea shipping emissions can be biased close to the coasts. Several measures are taken to limit the problem, such as the restriction of the analysis of the shipping lanes far from the coasts and from the crossing with other lanes, and the joint inversion and analysis of the land and maritime emissions at regional scale by main sectors of activity. Furthermore, the local scale analyses of NO_x emissions are based on a relatively simple modeling of the local atmospheric chemistry, which can raise significant uncertainties. Some specific investigations are conducted to strengthen the corresponding parameterization.

Regarding the regional scale inversion of the CH_4 emissions, the current analyses tend to confirm the lack of major signal from the sea shipping in the TROPOMI XCH_4 observations, but the attempts at extracting as much information as possible on the maritime emissions and at assessing the potential of the TROPOMI data based on the inversions are maintained.

3.1.3. HYPERSPECTRAL PROXY DATA

Overview of the PRISMA, Sentinel-2 and EnMAP instruments and data.

The PRISMA (PRecursore IperSpettrale della Missione Applicativa) high spatial resolution hyperspectral satellite imager (HCY) was launched into orbit in 2019. PRISMA has a spatial resolution of 30 m and a spectral resolution of 10 nm across a spectral range from 400 to 2500 nm, encompassing 239 spectral bands. The platform's revisit time is 29 days, which can be reduced to less than a week thanks to the PRISMA platform's ability to depoint up to ± 14.7°. For this study, it has been decided to utilise Level 1C products, which are spectrally and radiometrically corrected. The Signal-to-Noise Ratio (SNR) of PRISMA products is on average equal to 200 between 0.4 and 1.0 µm and 100 between 2 and 2.5 µm. The footprint of PRISMA images is 30 by 30 km.

(The Sentinel-2 (S2) platform was launched in June 2015 (first platform), followed by a second launch in March 2017. It is equipped with the MultiSpectral Instrument (MSI) sensor, which acquires a luminance signal across 13 spectral channels ranging from 442 nm to 2202 nm. The spatial resolution of the MSI sensor varies from 10 to 60 m, depending on the spectral channel, with a revisit time of 5 days. The swath width of S2 images is 290 km).

EnMAP (Environmental Mapping and Analysis Program) is a German hyperspectral satellite mission designed to deliver high-quality Earth observation data for environmental monitoring and resource management. Operating in the VNIR (420–1000 nm) and SWIR (900–2450 nm) spectral ranges, EnMAP captures over 230 contiguous spectral bands with high radiometric and spectral fidelity. Its 30 m spatial resolution and 30 km swath width enable detailed mapping of surface materials, including vegetation, soils, minerals, and water bodies.

The data gathered will be in radiance for both PRISMA and EnMap. COCHISE (Miesch et al., 2005) tool based on the Copernicus Atmosphere Monitoring Service (CAMS) atmospheric products will then be used to provide surface reflectance level data associated with bands of interest, e.g., the Short-Wave InfraRed (SWIR) domain for methane. Spatial registration between the different datasets over the same site will be carried out using the Gefolki tool from (Plyer et al., 2015) on a Universal Transverse Mercator (UTM) grid.

Investigations using PRISMA and EnMap data will provide insights into the potential of the future CHIME observations at a local scale.

PRISMA and S2 data gathering and processing for local atmospheric inversions of CH_4.

PRISMA's limited spatial coverage makes it unsuitable for regional-scale analyses. However, a range of data is available over various large-scale port complexes in Europe (e.g., Antwerp, Rotterdam, Fos/Mer, etc.), as well as some open sea acquisitions. Time series and inter-sites analyses are being carried out to determine to what extent observations of local or temporal anomalies linked to ship exhausts can be observed using these instruments. This analysis is based on Matched Filter tools (Foote et al., 2020; Nesme et al., 2021) and Constrained Nonnegative Matrix Factorisation (CNMF) (Calassou et al., 2024) when using both PRISMA and S2 data. Exhaust plume anomaly products can be delivered at a 30 m spatial resolution or at coarser resolution if needed.

PRISMA and S2 data are being collected over selected areas of interest, such as Liquefied Natural Gas (LNG) terminals, main harbours, and their vicinities, and expected "hot spots" along major maritime routes. While data collection has begun in certain ports, the site selection needs to be refined to ensure spatially and temporally consistent data across the different instruments. Additionally, the areas of interest are consolidated in collaboration with the stakeholder, EMSA.

Validation of dataset

To be updated in the next version of the document

Challenges, gaps, and lessons learnt.

To be updated in the next version of the document

3.1.4. GHGSAT PROXY DATA

Overview of the GHGSat instrument and data.

GHGSat data is collected using a Wide-Angle Fabry-Perot (WAF-P) hyperspectral imaging spectrometer operating in a narrow band of the SWIR region of the electromagnetic spectrum. The sensor measures backscattered solar SWIR at 1635-1670nm (0.1 nm spectral resolution). The precision of the methane column retrieval is 1%-3% (more details about the parameters of the GHG constellation are presented in Table 3.1).

This sensor has both (i) very high spatial resolution, and (ii) very high spectral resolution, to enable high-precision measurement of vertical column densities of methane in several hundred thousand pixels within the instrument field of view. The sensor is designed to provide a measurement of the excess methane concentration due to local sources with respect to the surrounding background within a given scene.

The sensor is ideally suited for detecting large point-source emissions. More information can be found in a paper from (Jervis et al., 2021) and (MacLean et al., 2024). Key parameters of the satellite and constellation can be found in Table 3.1.

Table 3.1: Key parameters of GHGSat constellation

Parameter	Onshore	Offshore
Wavelength sensitivity	~1.6 um (short-wave infrared)	~1.6 um (short-wave infrared)
Spatial Resolution	25 m	30 – 100 m
Detection threshold	100 kg/hr (50% PoD at 3 m/s wind speed)	150 – 1000 kg/hr (50% PoD at 3 m/s wind speed)
Field of view	12 km x 12 km	Variable
Combined revisit time	1-2 days	1-2 days

GHGSat data gathering and processing for local atmospheric inversions of CH_4.

GHGSat will task up to 24 new observations of offshore and coastal sites. The goal is that these observations will be timed with activities at the sites that could lead to emissions, such as the transfer of LNG between ships and onshore facilities. Observations will be coordinated with other researchers and stakeholders to ensure they can be effectively used.

It is possible to task anywhere globally but it is important to consider factors such as cloud cover and sun elevation. GHGSat is continuously working on the optimal imaging mode for the targeted sites, in most cases for this project that will be Glint mode, which is designed for offshore sites. However, for analysis of coastal areas, GHGSat is working to determine whether to use the standard onshore imaging mode or Glint mode. GHGSat is carefully planning and coordinating observations with other parties as two working days' notice is required for new tasking, and there can be fewer opportunities in a month for Glint observations.

For offshore and near-shore areas, satellites will be tasked in Glint mode, better enabling the detection of methane from facilities. Short-wave infrared radiation reflects poorly from water. Tasking the satellites over water the same way it is done over land would return insufficient signal to detect the presence of methane in the atmosphere. To resolve this issue, satellites are tasked in glint mode. By using a large satellite viewing angle at a calculated time to maximise the sun's reflectance on the ocean's surface, GHGSat can increase signal levels in the SWIR spectrum for observations over water where it is usually much weaker than on land. This increase in signal permits for the detection of methane over water which would otherwise be undetectable under normal operating procedures. More information on the methodology can be found in MacLean et al., (2024).

The project's observations are processed with GHGSat's proprietary toolchain software. The images are corrected for sensor response and other instrument-level effects. The methane column densities and surface reflectance (albedo) information are retrieved for each pixel using a measurement model which includes surface, instrument, and atmospheric contributions to the spectral radiance. The resulting arrays are:

- Surface reflectance (albedo): This layer shows the SWIR surface reflectance within every pixel. The values represent the fraction of light received on a surface that is reflected. A value of zero indicates that no light is reflected and a value of one indicates that all the light received on a surface is reflected.
- Methane concentration: The methane concentration values represent how many methane particles are in a given volume of air compared to the local background. The background is set to a value of zero. The concentration of the methane particles is measured in parts per billion (ppb).
- Methane concentration error: The error is given in parts per billion and is expressed as the standard deviation for a single satellite pass. This includes instrument errors, and the quality of the model fitted on the data.
- Quality flag: This is a layer that presents the quality of each pixel.

Finally, the resulting arrays are georeferenced, assigning real-world coordinates to each pixel of an image. Image location is determined through the satellite's GNSS and altitude outputs, time of acquisitions, and the number of frames taken. The final georeferencing is then done by image registration with a reference georeferenced image such as Landsat reflectance products.

Every emission identified is isolated with a plume mask. GHGSat's machine learning algorithms create masks for each emission automatically. Upon inspection by a human operator the operator may elect to modify the mask to better isolate the emission if necessary. Pixels that are flagged as "bad quality" in the flag layer are excluded from the mask. The masked plume can then be used to calculate the source rate.

The source rate (Q) represents the mass of methane released in the atmosphere in a given period of time and is given in kilograms per hour (kg/hr). The emission rate is estimated for each emission by applying a source rate retrieval method that uses the methane concentration, emission morphology and wind data from global meteorological models. GHGSat uses the Integrated Mass Enhancement (IME) method of calculating a source rate. Further details on the IME method can be found in (Varon et al., 2018).

Validation of dataset

There is limited information on the volume of methane emissions from LNG carriers during transport and transfer. As such, it is difficult to validate measurements of ship emissions during this project. Future work will use historical AIS data to confirm the presence of LNG carrier ships either in anchorages or in port to improve the confidence of detections.

Challenges, gaps, and lessons learnt.

During the observation campaign that has already been put in place, it was found that that reflectance behaviour was highly observation and scene dependent. Fixed camera configurations that were not adjusted for each observation and scene led to inconsistent signal quality, including significant signal saturation or under-exposure on land or water surfaces depending on the conditions.

This insight has informed the development of a predictive, observation-specific configuration strategy: optimal camera settings must be adjusted for each satellite pass based on the reflectance properties of both land and water, the latter which will vary with each observation based on the solar and satellite viewing angles and the water surface roughness. For each observation, predicted radiometric signal levels are calculated in advance over water and land, and this information is used to optimize the camera exposure and gain settings. We are now in the process of implementing this observation strategy and plan to reevaluate its effectiveness in the coming months.

A significant challenge encountered is the validation of the emission and linking the emission to a specific ship. To date, GHGSat has validated the presence of LNG ships using in-house satellite data and open-source satellite imagery. However, this imagery is not always available or of sufficient resolution to identify the vessel type. Future work will use historical AIS data to validate the presence of a ship and aid our linking of plumes to specific ships.

4. APPLICATION DEVELOPMENT PRINCIPLES

The EO4SEM representative dataset also supports the development of an application that aims at demonstrating the value of the Copernicus Expansion multi-mission and its synergy with the existing Copernicus Sentinel missions.

This section provides an overview of the ESA APEx platform (https://APEx.esa.int/) components which are being used as the technological framework that allows the development of the project application.

4.1. ESA APEX PLATFORM

The overall configuration of the EO4SEM APEx environment has been agreed with VITO (leaders of APEx) in a dedicated onboarding meeting and is illustrated in Figure 4.1.

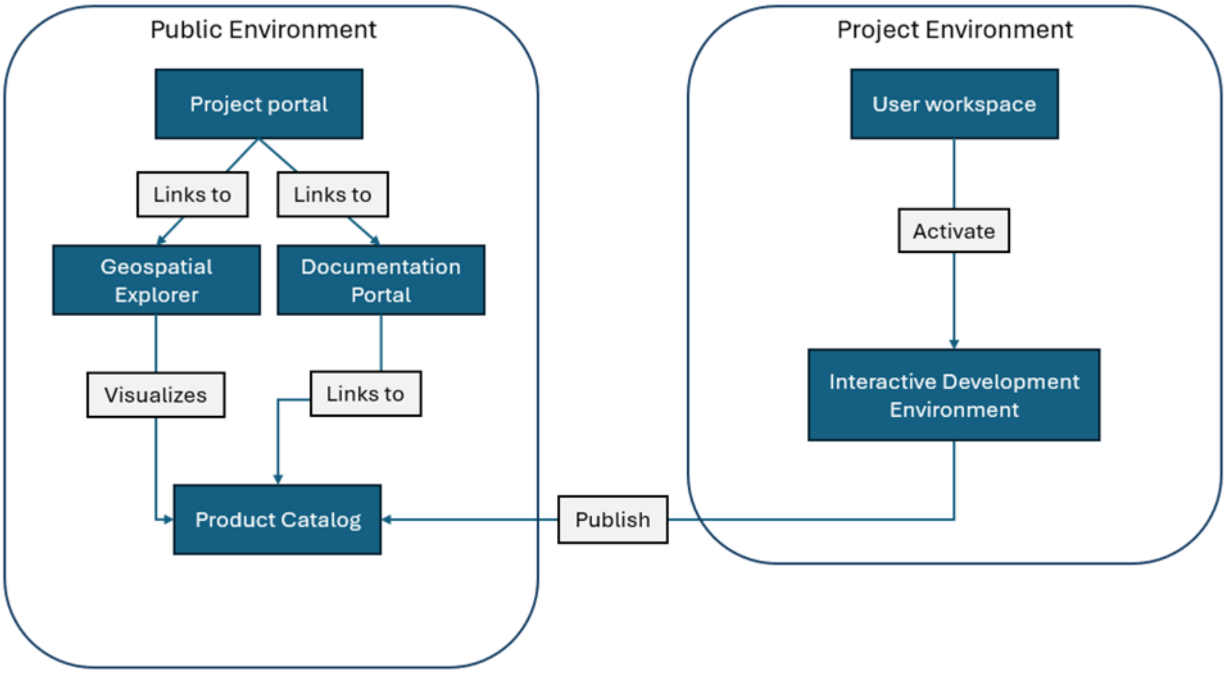

Figure 4.1: EO4SEM APEx cloud-based development platform diagram requested by GMV

As illustrated, there are four main APEx components that have been requested for EO4SEM:

- Geospatial explorer
- Product catalogue
- Project portal
- Documentation portal

These components are described in the following sections of this document, with the exception of the documentation portal that is still to be setup.

4.1.1. GEOSPATIAL EXPLORER

The EO4SEM Geospatial Explorer (https://explorer.eo4sem.apex.esa.int/) is a configurable web dashboard to display and visualize geospatial data. It is an interactive, cloud-based platform that enables users to visualize and analyse EO data alongside related geospatial information. It supports viewing EO imagery, derived products, and vector layers with control over transparency, ordering, and split-screen comparisons.

Users can inspect data with tools for spatial queries, distance measurement, and interactive charts, while metadata access ensures transparency. Flexible and user-friendly, it integrates EO and non-EO data into custom dashboards for diverse audiences.

At the moment, several datasets are already being provided in the Geospatial Explorer, as illustrated in Figure 4.2.

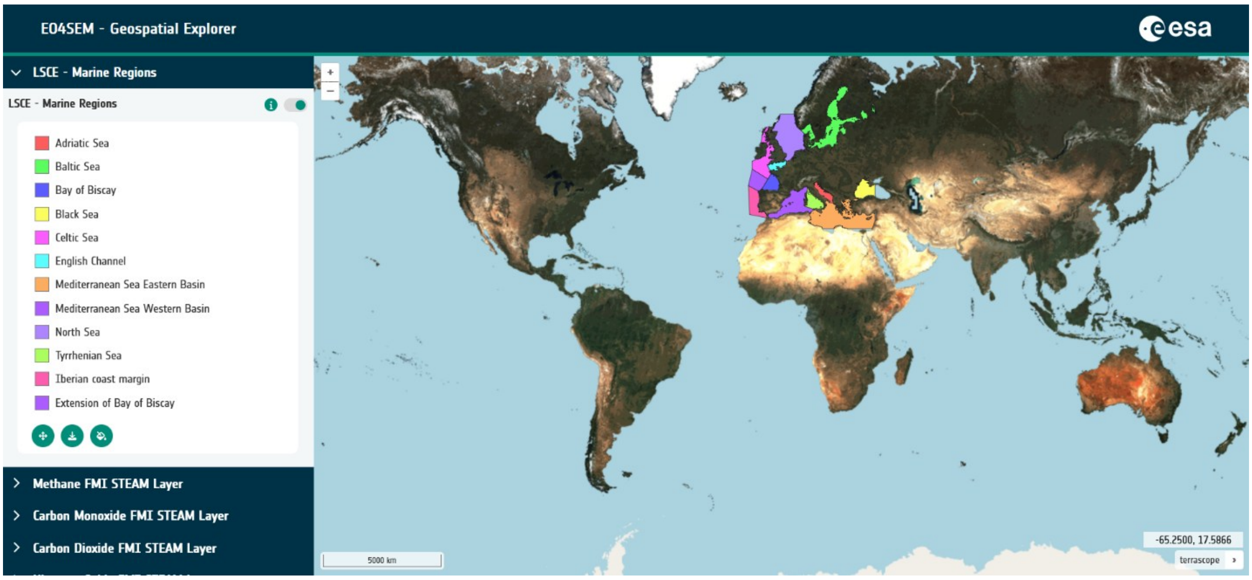

Figure 4.2: EO4SEM Geospatial explorer

4.1.2. APEX PRODUCT CATALOGUE

The EO4SEM Product Catalogue (https://browser.eo4sem.apex.esa.int/?.language=en) is a cloud-based service designed to support efficient geospatial data management through the creation of customisable SpatioTemporal Asset Catalogues (STAC).

It enables seamless integration of Earth Observation datasets into workflows via a RESTful API and user-friendly JSON metadata, supporting both public and private collections. This service enhances FAIR (Findable, Accessible, Interoperable, Reusable) data compliance, interoperability with tools like openEO, and facilitates rapid data discovery and sharing across EO projects.

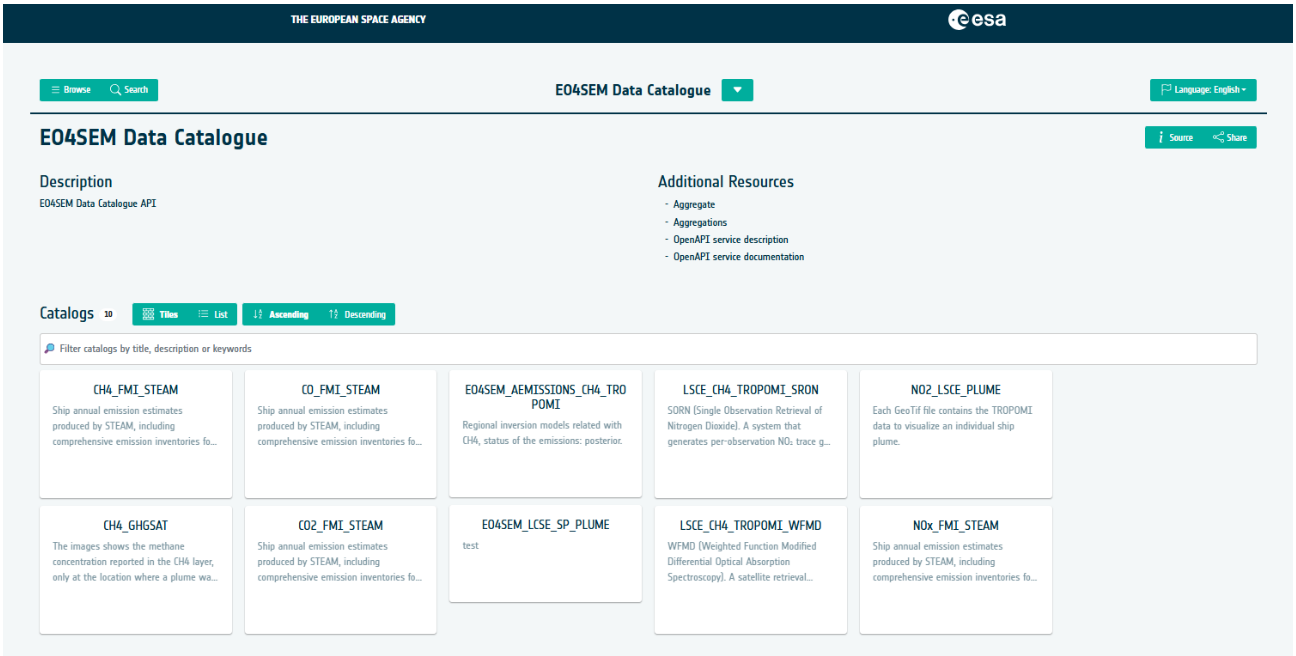

Figure 4.3 Screenshot of EO4SEM Data Catalogue

4.1.3. APEX WEB PORTAL

The EO4SEM Web Portal (https://eo4sem.apex.esa.int/en) provides Earth Observation (EO) projects with a customisable, secure, and content-rich website platform built on a flexible Drupal CMS. It enables advanced content management, seamless integration with other APEx services (like dashboards and forums), and automatic security patching. Designed to support EO project dissemination and community engagement, the portal allows teams to share updates, results, and documentation efficiently, while maintaining a cohesive digital ecosystem throughout the project lifecycle.

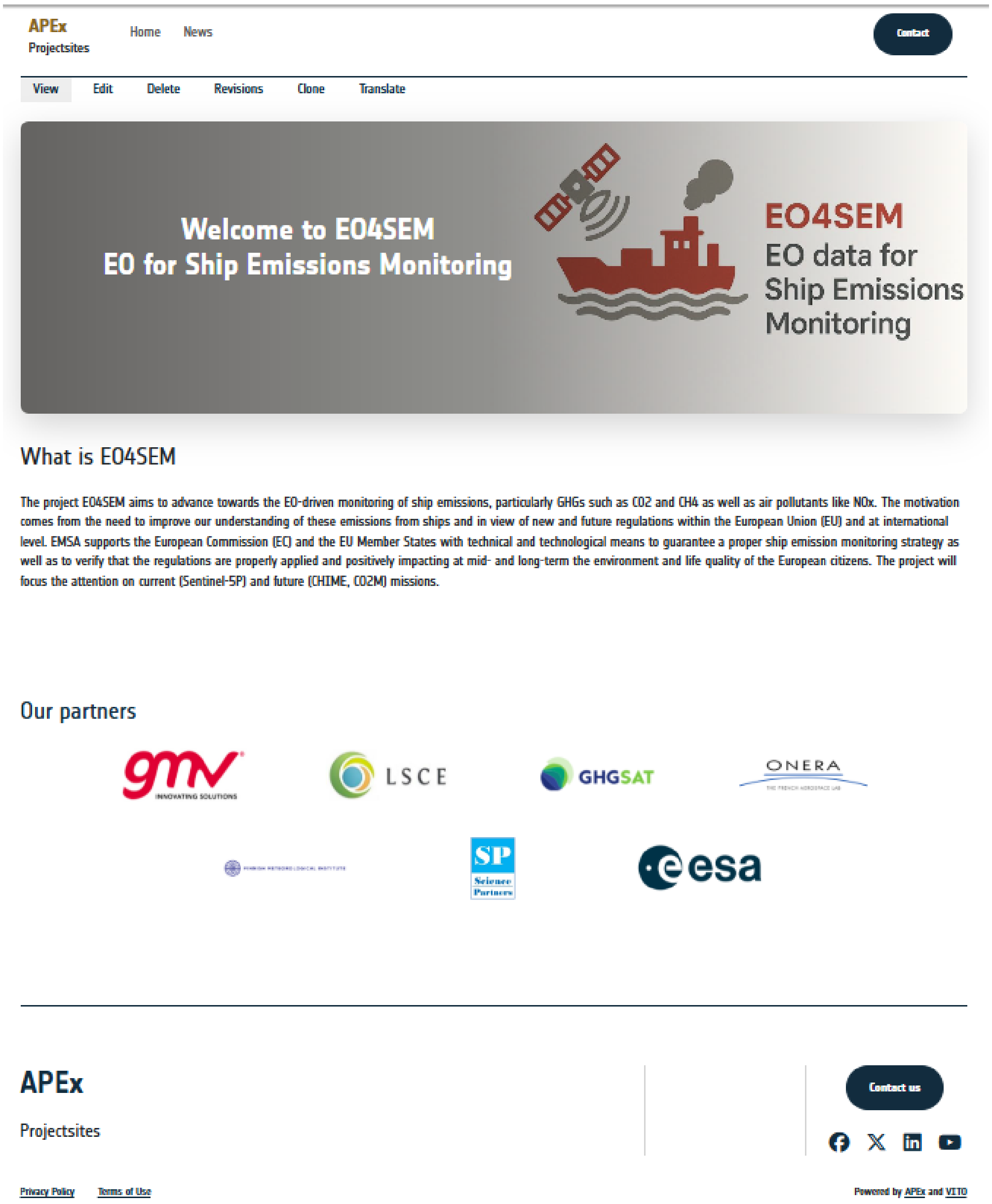

Figure 4.4: Screenshot of EO4SEM project portal (more content coming soon)

5. REPRESENTATIVE DATASET PACKAGING

For its inclusion in APEx, the data that is being generated in the scope of EO4SEM is being systematically organised based on file format and data type. This involves reviewing the different datasets to understand their structure and preparing them for further processing.

As part of this workflow, data files are converted to Cloud Optimised GeoTIFFs (COGs) to enable efficient storage access and integration with the APEX services.

5.1. DATA EXPLORATION AND ANALYSIS

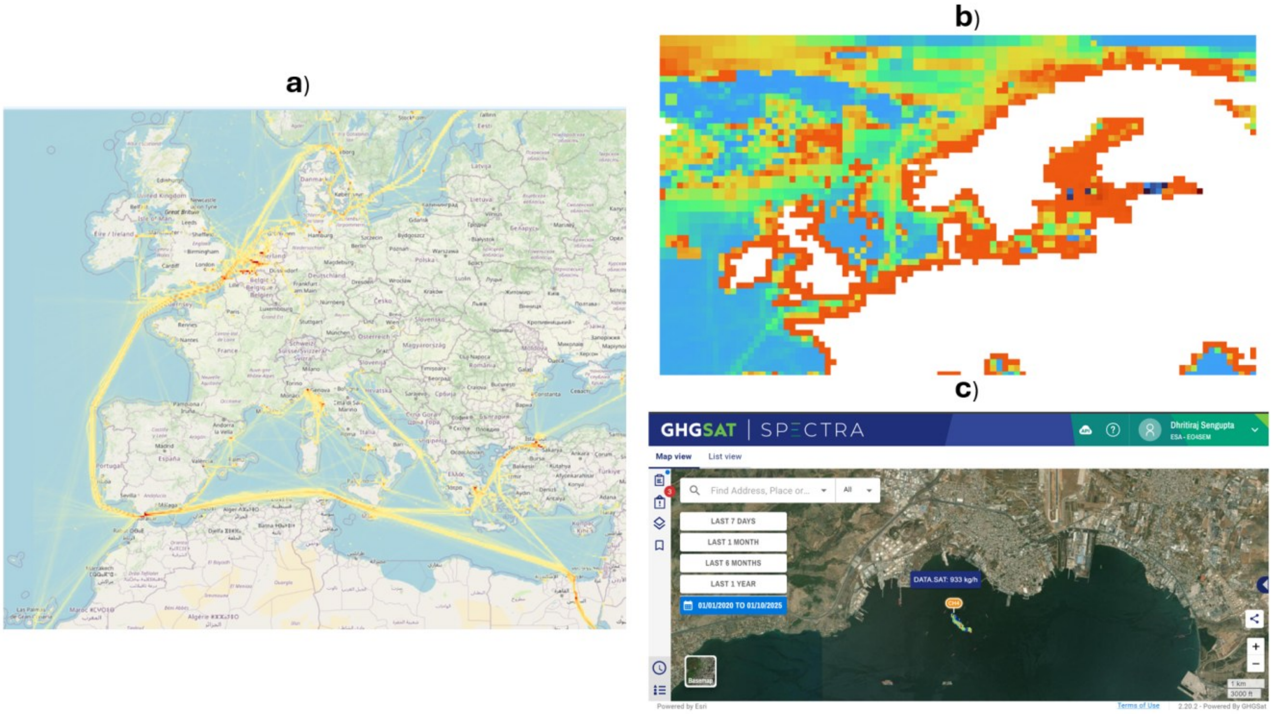

Figure 5.1 a) Nox STEAM 2019 from FMI b) EO4SEM_AEMISSIONS_CH4_TROPOMI_WFMD_CAMSREGv8_201911 — ch4_posterior. The principle of the traditional atmospheric inversion approach is to correct a priori estimates of the emission maps, also denoted "prior emissions", to reduce differences between atmospheric observations and their simulations with a CTM fed with the estimates of emission maps. The prior estimates of NOx and CH4 anthropogenic emissions in this study are based on emission estimates from the CAMSREG-v8 gridded inventory for the year 2019. We assimilate observations from the TROPOMI RPRO-v020400 for NO2 and from TROPOMI-SRON or TROPOMI-WFMD for CH4. Two types of files will be available from LSCE/Science Partners for the regional inversions of CH4 and NOx maritime emissions. **c)** Plume detected by GHGSAT near the port of Athens.

Also, a specific function to generate JSON objects for adding layers/items to the APEx Catalogue Viewer (Geospatial Explorer) has been developed. This function:

- Outputs (and prints) the JSON dictionary with necessary parameters.
- Integrates seamlessly into the complete STAC schema for visualization.

This function accepts the necessary parameters to populate the JSON and returns (while also printing) a Python dictionary containing the data for the layer to be added to the viewer. This output can then be integrated into the complete schema.

```
configuration_schema = {
    "name": name,
    "isActive": is_active,
    "meta": {
        "description": description,
        "attribution": {
            "text": atribution_text,
            "url": source if atribution_text else None
        },
    },
    "categories": categories,
    "layout": {
        "layerCard": {
            "toggleable": toggleable,
            "legend": {
                "type": legend_type
            },
            "controls": {
                "opacitySlider": opacity_slider,
                "zoomToCenter": zoom_to_center,
                "download": source
            }
        },
        "interfaceGroup": interface_group
    },
    "data": [
        {
            "url": source,
            "zIndex": z_index,
            "format": format
        }
    ]
}
```

Figure 5.2: Function to to generate JSON objects for adding layers/items to the Catalogue Viewer (Geospatial Explorer)

5.2. DATA PROCESSING

In terms of processing of data for its integration in the project's representative dataset, there are also steps that are relevant to refer to, namely:

1. Adding coordinate reference system (CRS): to ensure spatial consistency and interoperability across datasets from different partners, it is essential to add a common Coordinate Reference System (CRS) to all .nc (NetCDF) files. This process involves:
    - Verifying the presence of geolocation metadata (latitude, longitude, or projection parameters) in each file.
    - Standardizing the CRS, typically to a widely accepted system like EPSG:4326 (WGS 84)

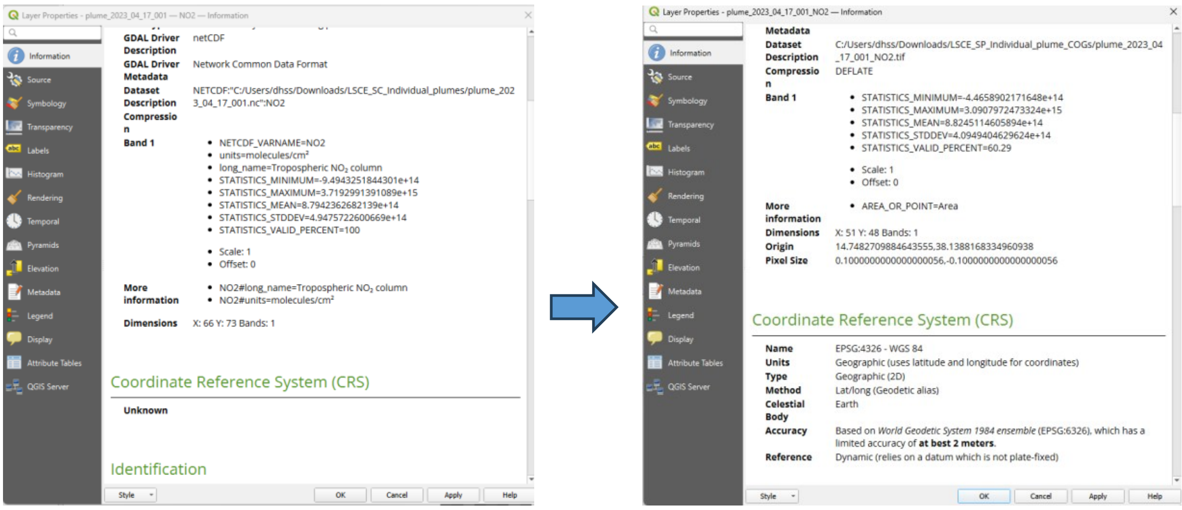

Figure 5.3 Adding coordinate reference system to input data

This step ensures that all datasets can be accurately aligned and analyzed together, especially when integrating EO data for visualizing in the geospatial explorer.

2. Conversion to COGs for all files in a folder: A COG is a geospatial raster format designed for efficient access in cloud environments. Unlike traditional GeoTIFFs, COGs allow partial data reads via HTTP range requests, enabling users to stream only the needed portions of a file. This is achieved through internal tiling and overviews, which support fast visualization and analysis. COGs are ideal for web mapping, remote sensing, and large-scale platforms like Google Earth Engine or AWS. They reduce bandwidth and storage costs, improve scalability, and are supported by tools like GDAL and QGIS. As an open format, COGs enhance accessibility and performance in cloud-based workflows.

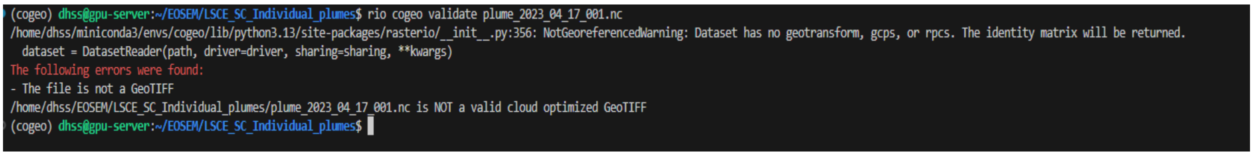

Figure 5.4 Data analysis showing not valid COGs

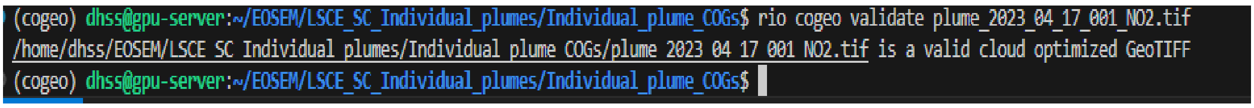

Figure 5.5 Conversion of data do valid COGs

5.3. METADATA FILE

Another key aspect that has been accounted for in the metadata file, which contains a comprehensive data description and characteristics of all datasets. This task was carried out in close collaboration with project partners to ensure accuracy and consistency and materialized in the form of an Excel file (Figure 5.6) that serves as a central reference for the project team, providing essential information about the datasets and facilitating better coordination and data management across the consortium.

A version of this file will be publicly available to download from the web portal.

Data name	Dataset name (as Shared with GMV)	Description	Unit	File format	Parameters/Bands/ Variables	CRS	Time interval	License	Color Palette	Min and Max values
Ship Emission Nitrogen Oxides Monthly	EO4SEM_NOx_Emissions_from_TROPOMI_orbits_2023061	The principle of our traditional atmospheric inversion approach is to correct a priori estimates of the emission maps, also denoted "prior emissions", to reduce differences between atmospheric observations and their simulations with a CTM fed with the estimates of emission maps. The prior estimates of NOx and CH4 anthropogenic emissions in this study are based on emission estimates from the CAMSREG-v8 gridded inventory for the year 2019. We assimilate observations from the TROPOMI RPRO-v020400 for NO2 and from TROPOMI-SRON or TROPOMI-WFMD for CH4.	kteqNO2/month	.nc	Nox_post, Nox Prior	WGS84	Monthly	Open Data		
Shipping Lanes Binary Masks	DEFINE_SHIPPING_LANES_MASK_V1	This dataset provides binary masks for all defined shipping lanes derived from STEAM 2023 dataset through geospatial analysis of maritime traffic patterns. Each lane is represented by a variable 'shipping_lane_X_mask' (e.g., 'shipping_lane_17_mask', 'shipping_lane_18_mask'), where 1 indicates pixels within the lane and 0 elsewhere. These masks support spatial analysis and emission calculations for maritime routes.	Unitless (boolean type)	.nc	shipping_lane_X_mask (X = 1 to 59, e.g., shipping_lane_17_mask, shipping_lane_18_mask)	WGS84	Static			0, 1
NOx and CO2 Linear Emissions for shipping lanes	NOX_CO2_EMISSIONS_OF_SHIPPING_LANES_17_AND_18_SUMMER2023	Linear emissions of NOx (kgeqNO_2/km/month) for shipping lanes 17 and 18, calculated from TROPOMI satellite data using a divergence-based methodology and STEAM for June to August 2023. Data includes route length (km) and is sorted by shipping lane and month. These files are intended to provide an initial overview of the shipping lanes and their emissions. We expect to refine the results and potentially revise the file structure as we expand the scope and improve our extraction and calculation methods.	kgeqNO_2/km/month	.csv	Route, Month, Length_km, NOx_STEAM, NOx_TROPOMI, CO2_STEAM, CO2_TROPOMI	WGS84	Summer 2023			

Figure 5.6 Data description Excel file

5.4. TECHNICAL MONITORING USING GITLAB

Technical Project Management is being done using GitLab's integrated DevOps platform to manage code, track issues, and coordinate development workflows (Figure 5.7). Code is stored in GitLab repositories, enabling version control and collaboration.

Key features include issue tracking for tasks and bugs, CI/CD pipelines for automated testing and deployment, and merge requests for streamlined code reviews. Milestones and boards help visualize progress and timelines, while built-in wikis support documentation and knowledge sharing.

This unified approach improves traceability, accountability, and team alignment, making GitLab a powerful tool for managing technical projects efficiently from planning to delivery.

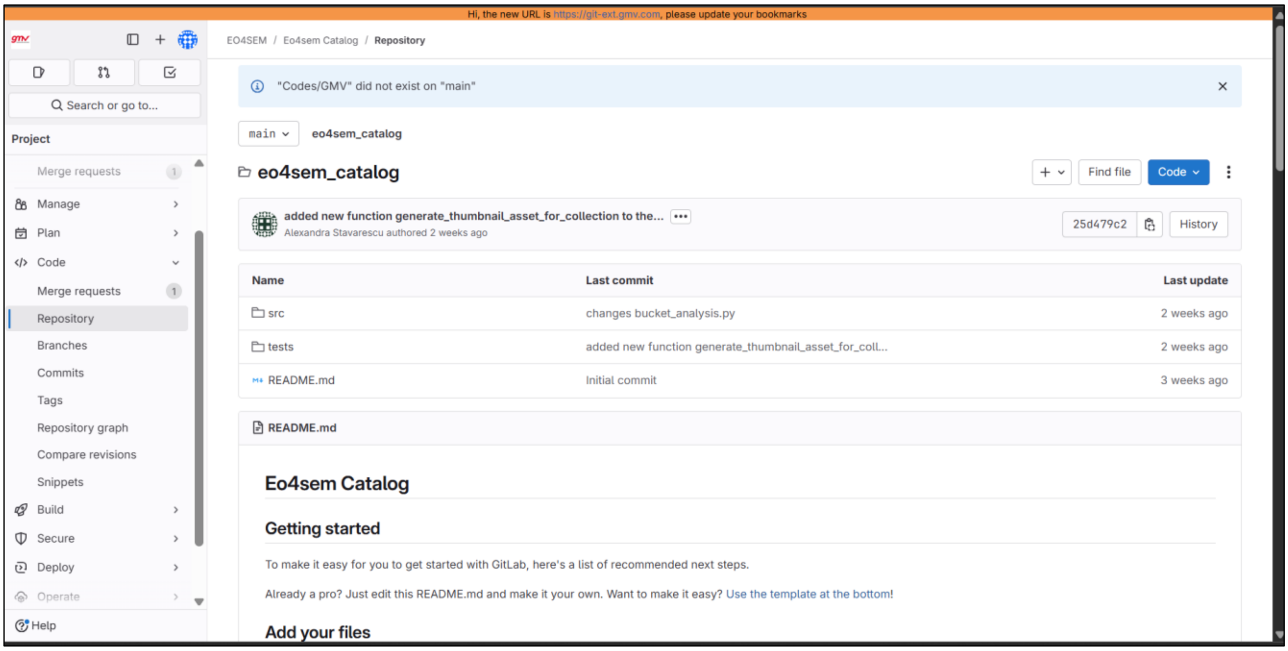

Figure 5.7: EO4SEM GitLab repository

6. ANNEX A: BIBLIOGRAPHIC REFERENCES

Calassou, G., Foucher, P.-Y., & Léon, J.-F. (2024). Quantifying particulate matter optical properties and flow rate in industrial stack plumes from the PRISMA hyperspectral imager. *Atmospheric Measurement Techniques*, *17*(1), 57–71. https://doi.org/10.5194/amt-17-57-2024

Foote, M. D., Dennison, P. E., Thorpe, A. K., Thompson, D. R., Jongaramrungruang, S., Frankenberg, C., & Joshi, S. C. (2020). Fast and Accurate Retrieval of Methane Concentration From Imaging Spectrometer Data Using Sparsity Prior. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(9), 6480–6492. IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2020.2976888

Fortems-Cheiney, A., Broquet, G., Potier, E., Berchet, A., Pison, I., Martinez, A., Plauchu, R., Abeed, R., Sicsik-Paré, A., Dufour, G., Coman, A., Savas, D., Siour, G., Eskes, H., Denier van der Gon, H. A. C., & Dellaert, S. N. C. (2025). Assessing the ability to quantify the decrease in NO_x anthropogenic emissions in 2019 compared to 2005 using OMI and TROPOMI satellite observations. *Atmospheric Chemistry and Physics*, *25*(12), 6047–6068. https://doi.org/10.5194/acp-25-6047-2025

Grigoriadis, A., Mamarikas, S., Ioannidis, I., Majamäki, E., Jalkanen, J.-P., & Ntziachristos, L. (2021). Development of exhaust emission factors for vessels: A review and meta-analysis of available data. *Atmospheric Environment: X*, *12*, 100142. https://doi.org/10.1016/j.aeaoa.2021.100142

Jalkanen, J.-P., Brink, A., Kalli, J., Pettersson, H., Kukkonen, J., & Stipa, T. (2009). A modelling system for the exhaust emissions of marine traffic and its application in the Baltic Sea area. *Atmospheric Chemistry and Physics*, *9*(23), 9209–9223. https://doi.org/10.5194/acp-9-9209-2009

Jalkanen, J.-P., Johansson, L., & Kukkonen, J. (2014). A Comprehensive Inventory of the Ship Traffic Exhaust Emissions in the Baltic Sea from 2006 to 2009. *AMBIO*, *43*(3), 311–324. https://doi.org/10.1007/s13280-013-0389-3

Jalkanen, J.-P., Johansson, L., & Kukkonen, J. (2016). A comprehensive inventory of ship traffic exhaust emissions in the European sea areas in 2011. *Atmospheric Chemistry and Physics*, *16*(1), 71–84. https://doi.org/10.5194/acp-16-71-2016

Jalkanen, J.-P., Johansson, L., Kukkonen, J., Brink, A., Kalli, J., & Stipa, T. (2012). Extension of an assessment model of ship traffic exhaust emissions for particulate matter and carbon monoxide.

Atmospheric Chemistry and Physics, *12*(5), 2641–2659. https://doi.org/10.5194/acp-12-2641-2012

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. *Atmospheric Measurement Techniques*, *14*(3), 2127–2140. https://doi.org/10.5194/amt-14-2127-2021

Johansson, L. (2011). Emission estimation of marine traffic using vessel characteristics and AIS-data. *Master's Thesis, Aalto University, Espoo, Finland*. https://www.semanticscholar.org/paper/Emission-estimation-of-marine-traffic-using-vessel-Johansson/f6da9dd902c6827431528b9675ed479dcc6e6753

Johansson, L., Jalkanen, J.-P., Kalli, J., & Kukkonen, J. (2013). The evolution of shipping emissions and the costs of regulation changes in the northern EU area. *Atmospheric Chemistry and Physics*, *13*(22), 11375–11389. https://doi.org/10.5194/acp-13-11375-2013

Johansson, L., Jalkanen, J.-P., & Kukkonen, J. (2017). Global assessment of shipping emissions in 2015 on a high spatial and temporal resolution. *Atmospheric Environment*, *167*, 403–415. https://doi.org/10.1016/j.atmosenv.2017.08.042

Kuhlmann, G., Koene, E., Meier, S., Santaren, D., Broquet, G., Chevallier, F., Hakkarainen, J., Nurmela, J., Amorós, L., Tamminen, J., & Brunner, D. (2024). The *ddeq* Python library for point source quantification from remote sensing images (version 1.0). *Geoscientific Model Development*, *17*(12), 4773–4789. https://doi.org/10.5194/gmd-17-4773-2024

Lorente, A., Borsdorff, T., Butz, A., Hasekamp, O., aan de Brugh, J., Schneider, A., Wu, L., Hase, F., Kivi, R., Wunch, D., Pollard, D. F., Shiomi, K., Deutscher, N. M., Velazco, V. A., Roehl, C. M., Wennberg, P. O., Warneke, T., & Landgraf, J. (2021). Methane retrieved from TROPOMI: Improvement of the data product and validation of the first 2 years of measurements. *Atmospheric Measurement Techniques*, *14*(1), 665–684. https://doi.org/10.5194/amt-14-665-2021

Lorente, A., Borsdorff, T., Martinez-Velarte, M. C., Butz, A., Hasekamp, O. P., Wu, L., & Landgraf, J. (2022). Evaluation of the methane full-physics retrieval applied to TROPOMI ocean sun glint measurements. *Atmospheric Measurement Techniques*, *15*(22), 6585–6603. https://doi.org/10.5194/amt-15-6585-2022

Lorente, A., Borsdorff, T., Martinez-Velarte, M. C., & Landgraf, J. (2023). Accounting for surface reflectance spectral features in TROPOMI methane retrievals. *Atmospheric Measurement Techniques*, *16*(6), 1597–1608. https://doi.org/10.5194/amt-16-1597-2023

MacLean, J.-P. W., Girard, M., Jervis, D., Marshall, D., McKeever, J., Ramier, A., Strupler, M., Tarrant, E., & Young, D. (2024). Offshore methane detection and quantification from space using sun glint measurements with the GHGSat constellation. *Atmospheric Measurement Techniques*, *17*(2), 863–874. https://doi.org/10.5194/amt-17-863-2024

Majamäki, E., & Jalkanen, J.-P. (2022). Refinement of the STEAM model. *Deliverable D(1.3) of the H2020 EMERGE Project*. https://emerge-h2020.eu/publications/

Miesch, C., Poutier, L., Achard, V., Briottet, X., Lenot, X., & Boucher, Y. (2005). Direct and inverse radiative transfer solutions for visible and near-infrared hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(7), 1552–1562. IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2005.847793

Nesme, N., Marion, R., Lezeaux, O., Doz, S., Camy-Peyret, C., & Foucher, P.-Y. (2021). Joint Use of in-Scene Background Radiance Estimation and Optimal Estimation Methods for Quantifying Methane Emissions Using PRISMA Hyperspectral Satellite Data: Application to the Korpezhe Industrial Site. *Remote Sensing*, *13*(24), Article 24. https://doi.org/10.3390/rs13244992

Plyer, A., Colin-Koeniguer, E., & Weissgerber, F. (2015). A New Coregistration Algorithm for Recent Applications on Urban SAR Images. *IEEE Geoscience and Remote Sensing Letters*, *12*(11), 2198–2202. IEEE Geoscience and Remote Sensing Letters. https://doi.org/10.1109/LGRS.2015.2455071

Rey-Pommier, A., Héraud, A., Chevallier, F., Ciais, P., Christoudias, T., Kushta, J., & Sciare, J. (2025). Global gridded NO_x emissions using TROPOMI observations. *Earth System Science Data*, *17*(7), 3329–3351. https://doi.org/10.5194/essd-17-3329-2025

Riess, T. C. V. W., Boersma, K. F., van Vliet, J., Peters, W., Sneep, M., Eskes, H., & van Geffen, J. (2022). Improved monitoring of shipping NO_2 with TROPOMI: Decreasing NO_x emissions in European seas during the COVID-19 pandemic. *Atmospheric Measurement Techniques*, *15*(5), 1415–1438. https://doi.org/10.5194/amt-15-1415-2022

Rijsdijk, P., Eskes, H., Dingemans, A., Boersma, K. F., Sekiya, T., Miyazaki, K., & Houweling, S. (2025). Quantifying uncertainties in satellite NO_2 superobservations for data assimilation and model evaluation. *Geoscientific Model Development*, *18*(2), 483–509. https://doi.org/10.5194/gmd-18-483-2025

Santaren, D., Hakkarainen, J., Kuhlmann, G., Koene, E., Chevallier, F., Ialongo, I., Lindqvist, H., Nurmela, J., Tamminen, J., Amorós, L., Brunner, D., & Broquet, G. (2025). Benchmarking data-driven inversion methods for the estimation of local CO_2 emissions from synthetic satellite images of XCO_2 and NO_2. *Atmospheric Measurement Techniques*, *18*(1), 211–239. https://doi.org/10.5194/amt-18-211-2025

Schneising, O. (2023). Algorithm Theoretical Basis Document (ATBD)—TROPOMI WFM-DOAS XCH4. *Technical Report*. https://climate.esa.int/media/documents/ATBDv3_GHG-CCI_CH4_S5P_WFMD.pdf

Schneising, O. (2024). Product User Guide (PUG)—TROPOMI WFM-DOAS (TROPOMI/WFMD) XCH4. *Technical Report*. https://www.iup.uni-bremen.de/carbon_ghg/products/tropomi_wfmd/data/v18/pug_wfmd.pdf

Sicsik-Paré, A., Fortems-Cheiney, A., Pison, I., Broquet, G., Opler, A., Potier, E., Martinez, A., Schneising, O., Buchwitz, M., Maasakkers, J. D., Borsdorff, T., & Berchet, A. (2025). Can we obtain consistent estimates of the emissions in Europe from three different CH_4 TROPOMI products? *EGUsphere*, 1–48. https://doi.org/10.5194/egusphere-2025-2622

van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J.-C., Sneep, M., ter Linden, M., Ludewig, A., Boersma, K. F., & Veefkind, J. P. (2022). Sentinel-5P TROPOMI NO_2 retrieval: Impact of version v2.2 improvements and comparisons with OMI and ground-based data. *Atmospheric Measurement Techniques*, *15*(7), 2037–2060. https://doi.org/10.5194/amt-15-2037-2022

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. *Atmospheric Measurement Techniques*, *11*(10), 5673–5686. https://doi.org/10.5194/amt-11-5673-2018

END OF DOCUMENT