Universiteit Leiden

Master Computer Science

Detecting ship plumes using TROMOPI NO_2, SO_2 and HCHO observations.

Name:	Ruben Ahrens
Student ID:	s3677532
Date:	27/01/2025
Specialisation:	Artificial Intelligence
1st supervisor:	Prof. F. J. Verbeek
2nd supervisor:	Dr. C. Veenman

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract

Earth observation is becoming important in gaining insight into shipping emissions which produce a large percentage of global emissions. Maritime pollution contributes to global warming, causes health issues in harbors and coastal regions, and disrupts marine ecosystems. Therefore, the International Maritime Organization (IMO) has imposed strict emission limits, and national governments are restricting emissions in harbors and national waters. Because self-reporting is insufficient for measuring compliance, remote sensing is essential for collecting data on ship emissions in the open sea. In previous works, individual ship plumes have been detected using TROPOMI/Sentinel-5P satellite measurements of NO_2. This study uses machine learning to investigate the impact of incorporating SO_2 and HCHO measurements, alongside NO_2, in detecting ships' exhaust plumes. Showing the potential of using gases other than NO_2 in finding anomalously emitting ships, benefitting compliance monitoring.

The study focuses on a portion of the eastern Mediterranean, covering the period from January 2020 to July 2022. From AIS (Automatic Identification System) data, we derived each ship's path for the past two hours until the satellite overpass. We derived the path of the plume from this path and wind data. We trained an XGBoost classifier on data samples from $80km \times 80km$ areas to predict whether ship plumes are present. Using XGBoost, we assess the impact of adding SO_2 and HCHO on the detectability of plumes. We created subsets by binning the data according to each data sample's estimated total of NO_x emissions using length and speed from ships (NO_x proxy). The bins help us study the relationship between the expected ship emission and the plume detectability. We created learning curves where we logarithmically increase the number of samples in the (sub)dataset. This shows if performance can be increased by obtaining more TROPOMI and AIS data.

We present heatmaps that combine learning curves with proxy bins to compare each possible combination of gases (NO_2, SO_2, and HCHO). Similar to other experiments, the heatmaps showed an increase in detectability for higher NO_x proxy bins for gas combinations involving NO_2.

Comparing the learning curves of different proxy bins shows that for all proxy bins, more data for training the model is beneficial.

We also see the added benefit of SO_2 and HCHO data in learning curves that use the entire tabular database without splitting the data into proxy bins.

Our findings show that incorporating SO_2 and HCHO measurements alongside

NO_2 improved the classifier's performance at higher and lower NO_x proxy values. We also observed SO_2 and HCHO data by themselves to detect ships with a ROC AUC of 0.647 and 0.634 respectively, while showing room for improvement in the learning curves when acquiring more TROPOMI and AIS data. While this performance is worse than NO_2 (ROC AUC 0.684) it shows the value of these gases in ship plume detection.

Contents

Chapter 1

Introduction

While essential for global trade, shipping is a major contributor to air pollution. Traditional compliance monitoring methods, which rely on self-reporting and infrequent inspections, are often insufficient to capture the full extent of ship emissions and to identify vessels exceeding regulatory limits. Remote sensing, particularly through satellite observations, offers a promising solution for monitoring emissions in open waters.

The European Space Agency's Sentinel-5P satellite, launched in October 2017, carries the TROPOspheric Monitoring Instrument (TROPOMI) [2]. TROPOMI measures nitrogen dioxide (NO_2), sulfur dioxide (SO_2), formaldehyde (HCHO), methane (CH_4), carbon monoxide (CO), and ozone (O_3) with daily global coverage. While previous studies have focused primarily on NO_2 measurements, TROPOMI also provides data on SO_2 and HCHO measurements, offering the potential for improved plume detection through multi-gas analysis.

Ships emit substantial amounts of these gases [1, 45, 25], each with different environmental and health impacts. Upon inhalation, NO_2 reaches the lungs, where it is

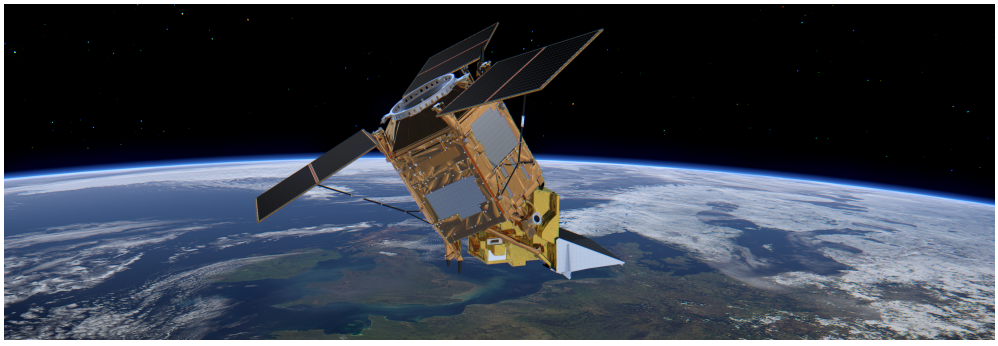

Figure 1.1: Sentinel-5p carrying the TROPOMI instrument measuring air quality. ©ESA/ATG medialab

converted into nitrous and nitric acid. These acids are highly irritating and can cause lung damage [22]. SO_2 contributes to acid rain and forms sulfate aerosols that contribute to climate change, while also causing respiratory and cardiovascular diseases in coastal populations [19]. HCHO is a known carcinogen [44].

The International Maritime Organization (IMO) has implemented increasingly stringent regulations to control these emissions. IMO2020 reduced the sulfur oxides (SO_x) content limit in marine fuels from 3.50% to 0.50% [37], while nitrogen oxides (NO_x) emissions are regulated globally through MARPOL ANNEX VI, with stricter limits in regions like the North and Baltic Seas [38]. Currently, there are no regulations for HCHO emissions, and their impact on air quality and human health in the context of maritime emissions hasn't been as extensively studied as for NO_x and SO_x.

Previous studies have successfully identified plumes from single TROPOMI orbits using NO_2 measurements combined with Automatic Identification System (AIS) [43] ship location data [18, 28]. Individual ship's emissions have been estimated by comparing measured NO_2 concentrations with expected emissions based on ship characteristics [31]. A ship's expected NO_x emissions can be estimated using its length and speed [18], creating a proxy that helps identify potentially anomalous emissions.

This thesis explores whether incorporating SO_2 and HCHO measurements alongside NO_2 data improves our ability to detect ship plumes. Using machine learning techniques and validation through AIS data, we investigate how different combinations of these gases affect plume detection accuracy across varying levels of expected emissions. The insights from this multi-gas analysis could inform future developments in satellite-based compliance monitoring approaches like those developed by Kurchaba et al. (2023) [31].

1.1 Research Questions

- (RQ 1) How does the inclusion of SO_2 and HCHO data, alongside NO_2 data, improve the detection accuracy of ship plumes using data from a single TROPOMI orbit?
- (RQ 2) Can ship plumes be detected based on SO_2 or HCHO data from a single TROPOMI orbit?

To answer these questions, We will use TROPOMI and AIS data from the Mediterranean for 2020-2022 to create a database of gas values and support features like wind and satellite measurement angles. These features will be used to classify ship plume presence in $80km \times 80km$ sample patches. The plume locations are calculated using wind data from the European Centre for Medium-Range Weather Forecast (ECMWF) and AIS ship location data. To study the relation between the expected NO_x emission and the detectability and compare detection performance between sample patches containing different levels of proxy values, we use the NO_x proxy introduced by Georgoulias et al. (2020) [18].

Because shipping lanes are visible in three monthly-averaged HCHO data [6], we hypothesize that including HCHO and SO_2 TROPOMI data will improve ship plume

detection performance. This also makes us think that HCHO data will result in a better than random performance. As ship plumes have yet to be detected in single TROPOMI overpasses based on only SO_2 or HCHO, we do not expect detection performance based on data from these gases alone to be better than NO_2.

This chapter has briefly introduced why this study is important and delimits the scope of this research. In the next chapter, background information about the satellite Sentinel-5P, TROPOMI, AIS, and the IMO will be provided. Additionally, the importance of reducing gas pollution will be described in more detail. Chapter 3 will cover previous studies in this field, while Chapter 4 will detail the design of this study. The experiments conducted and their results are presented in chapter 5. The overall results are discussed in chapter 6, chapter 7 draws a conclusion from the results, and the future work is discussed in chapter 8.

Chapter 2

Background

This chapter describes the background information on the thesis. First, the impact and magnitude of NO_2, SO_2, and HCHO emissions from ships are described. Then, we describe the Sentinel-5P satellite carrying the TROPOMI sensor and the TROPOMI data. After, we describe the data from which we derive ship locations. Finally, we describe the background of the machine learning techniques used in the study.

2.1 Marine Pollution

Marine pollution is a significant issue that has far-reaching consequences for both human health and marine ecosystems. Polluting gases such as NO_2, SO_2, and HCHO can cause severe health issues and disrupt marine life.

2.1.1 Emissions

The shipping industry is responsible for roughly 2.89% of GHG emissions in 2018, up from 2.76% in 2012 [23]. The emissions are assumed to increase by 150-300% by 2050 [3]. Emissions from shipping have grown more than 90 percent since 1990. This increase in emissions is alarming, necessitating stricter regulations. International shipping has a significant effect on air quality, with a contribution to global emissions of more than 20% for NO_x, around 10% for SO_x [1, 45]

The health implications of the pollution of NO_2, SO_2, and HCHO are serious. Formaldehyde is known to be carcinogenic in humans and likely poses similar risks to animals, affecting both the marine food chain and wildlife. NO_2 can lead to higher mortality rates [57]. When inhaled, NO_2 can turn into nitrous and nitric acids, which are highly irritating and can cause lung damage. This is one reason why NO_2 is correlated with asthma attacks. The World Health Organization (WHO) has established limits on NO_2 concentration for indoor air quality [26]. The effects of sulfur dioxide are similar, with significant impacts on respiratory health. Most SO_x is emitted in the form of sulfur dioxide (SO_2) by ships [25]

2.1.2 Regulations

To address these issues caused by shipping emissions, the IMO introduced the *IMO2020* [37] sulfur regulations. These regulations limit the sulfur content in marine fuel oil to 0.50%, down from the previous limit of 3.5%. This regulation aims to cut sulfur oxide emissions from ships by about 70%, improving air quality and protecting human health.

The European Union's *FuelEU Maritime* regulation, part of the "Fit for 55" package, aims to reduce GHG emissions from the maritime sector by promoting the use of renewable and low-carbon fuels. Targets are in place to make sure that the greenhouse gas intensity of fuels used in this sector will gradually decrease over time. The targets start with a 2% decrease by 2025 and reach up to an 80% reduction by 2050. The targets cover not only CO_2 but also methane and NO_x emissions over the full lifecycle of the fuels used on board. [12].

MARPOL Annex VI [38], adopted by the IMO, sets global limits on sulfur oxide and nitrogen oxide emissions from ship exhausts and prohibits deliberate emissions of ozone-depleting substances. It also establishes Emission Control Areas (ECAs) with stricter limits depending on a ship's age and engine type and includes measures like the Energy Efficiency Design Index to improve ship energy efficiency.

Unintended consequences

Regulations have been proven hard to design and enforce, causing friction between regulators and shipping companies; when sulfur regulations sharpened with *IMO2020*, big ships turned to use "scrubbers" [46, 45], a technique that involves spraying water over the exhaust, "cleaning" the exhaust plume from harmful gases by making the exhaust gases dissolve in water. Scrubber systems are mainly used to comply with *IMO2020*. Scrubbers will also decrease HCHO concentrations in the air, it is highly soluble in water, so a scrubber can remove it effectively. 3% of the total number of vessels and 16% of the gross tonnage as of November 2019 with this percentage likely increasing after 2020. The downside of scrubber systems is that it creates acidic water that is then dumped into the ocean or sea. This shows the unintended side effects of creating new regulations based solely on gas emissions when it does not result in less overall pollution. Additionally, regulations limiting SO_2 emission have also resulted in more NO_2 emission [54].

Compliance monitoring

Enforcing regulations requires sensitive measurement devices. A sniffer is a device that can locally measure the concentration of gases. Remote measuring techniques can be used when it is not possible or viable to measure near the ships. TROPOMI is a sensor on the satellite Sentinel-5P that remotely measures gas concentrations.

TROPOMI's predecessor, OMI, provided some insight into ship pollution [55]. However, the resolution of $13km \times 24km$ did not allow the detection of individual ship plumes. Thanks to TROPOMI's resolution improvement, much more is possible. The next section will discuss the Sentinel-5P satellite and TROPOMI sensor in detail.

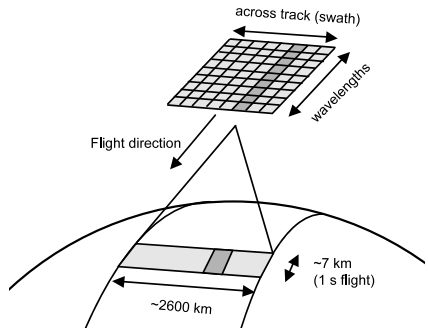

Figure 2.1: How Sentinel-5P flies and captures data using the TROPOMI sensor [11]

2.2 Sentinel-5P

The Sentinel-5 Precursor (Sentinel-5P or S5P) satellite, launched in October 2017 by the European Space Agency (ESA), stands at the forefront of Earth's atmospheric observation as part of the broader Copernicus program's [48] fleet of Sentinel satellites. Each satellite in this program plays a unique role in monitoring various aspects of Earth's system. S5P hosts the TROPOspheric Monitoring Instrument (TROPOMI), which, with its superior spectral and spatial resolution capabilities, significantly advances our ability to monitor air quality and atmospheric composition. The Royal Netherlands Meteorological Institute (KNMI) leads the TROPOMI Mission Performance Center consortium, which is responsible for deriving gas concentrations from S5P satellite measurements. TROPOMI data is collected through about 14 daily orbits [17], providing daily global coverage.

2.2.1 TROPOMI

The TROPOMI instrument distinguishes itself by offering high-resolution measurements of atmospheric gases and pollutants down to a spatial resolution of $7km \times 3.5km$ ($5.5km \times 3.5km$ since 6 August 2019 [10] or orbit number 9388) at near-nadir (the point directly beneath the satellite on the Earth's surface) allowing for monitoring of pollution sources and air quality on both local and global scales. The satellite's ability to cover a swath width of approximately 2600 kilometers guarantees daily global observations, enabling continuous monitoring of the atmosphere. TROPOMI is a sensor featuring four spectrometers that measure ultraviolet and visible light (270–495 nm), near-infrared light (675–775 nm), and shortwave infrared light (2305–2385 nm) [2]. TROPOMI flies in polar orbit and gets daily global coverage with an overpass time of 13:30 local time.

Parameter	Description
NO_2 SCD	Slant column density measurements of nitrogen dioxide
SO_2 SCD	Slant column density measurements of sulfur dioxide
HCHO SCD	Slant column density measurements of formaldehyde
Quality assurance value	Measurement quality indicator (filtered at 0.5)
Wind measurements	Meridional and zonal wind speed components
Sensor angles	Sensor zenith and azimuth angles
Solar angles	Solar zenith and azimuth angles
Cloud fraction	Fraction of cloud cover (0-1 scale)

Table 2.1: TROPOMI Data Products Used in the Study

2.2.2 Other Satellites

The Sentinel satellites, integral to the Copernicus program [48], provide earth observation data across a variety of domains:

Sentinel-1 offers radar imaging capabilities for all-weather, day-and-night surveillance, crucial for monitoring changes in land use, detecting oil spills, and supporting disaster management efforts. Sentinel-2 provides high-resolution (10-60m) optical imaging, Which is used for tracking vegetation health, observing land and water use changes, and supporting agricultural and forestry management. Sentinel-3 delivers global measurements of the ocean and land, including sea surface topography, temperature, and color, and is instrumental in oceanography, climate monitoring, and land surface analysis. From a geostationary orbit, Sentinel-4 tracks air quality and atmospheric composition over Europe and North Africa, offering data used for creating and monitoring environmental health and policy. Sentinel-5 is to be the next instrument after TROPOMI. It will be carried on the MetOp-SG weather satellite. While the launch was planned for 2021, it has yet to be launched in 2024. Its resolution will be similar to TROPOMI. It is meant to supersede TROPOMI as a more permanent solution. The Sentinel-6 mission, including Sentinel-6 Michael Freilich (Sentinel-6A), is tasked with the precise monitoring of global sea levels, contributing to climate change research and oceanography. The TANGO project is developing two satellites that will measure NO_2, methane, and CO_2 with a resolution of $300m \times 300m$. TANGO (Twin ANthropogenic Greenhouse Gas Observers) can revolutionize the space of compliance monitoring by being able to truly isolate local emitters.

2.2.3 Data Products Derived from Sentinel-5P Observations

Sentinel-5P offers a variety of data products; most of these products are gases. In table 2.1, all main S5P/TROPOMI products have been described. Aside from gases, support products such as wind speed and cloud fraction are also available [11]. The data products are available in various processing streams.

Column densities

Data products include Slant Column Densities (SCD) and vertical column densities (VCD). SCD measures the gas concentration in a column from the measurement point

to the sensor, while VCD measures the column of the measurement point along the zenith. The VCDs are influenced by historical data, as mentioned in the TROPOMI NO_2 product user manual[11]. The VCD is calculated with the TM5 system, in the last step of which the stripe amplitude is determined using a 7-day running mean for data over the Pacific Ocean. This influence could cause data leakage between the train and test set as it could indicate seasonal patterns in shipping data to the machine learning model, which can result in overly optimistic results, which is why we use SCD.

The slant columns we use are derived using the Differential Optical Absorption Spectroscopy (DOAS) technique. DOAS is a widely used method for retrieving trace gas concentrations in the atmosphere from satellite and ground-based measurements. It takes measured Earth radiance to the solar irradiance data and converts it to concentrations of trace gases. It exploits the light absorption characteristics of gases in the ultraviolet, visible, and near-infrared spectral regions. The DOAS algorithm has a different *fit window* and parameters for each derived gas. All parameters can be found in the Algorithm Theoretical Basis Document (ATBD). The ATBD of each gas provides the retrieval algorithms and its parameters in detail for NO_2 [51], SO_2 [47] and HCHO [21].

NO_2 The slant columns for NO_2 [51, 50] are also retrieved using DOAS. For NO_2, the fit window is 405–465 nm. The NO_2 retrieval algorithm features a step where the SCDs are separated into stratospheric and tropospheric values, a feature absent in the other gases' data products. As we are interested in ship plumes, we will use the tropospheric slant column density of NO_2.

HCHO The retrieval of HCHO slant columns [21, 7] is difficult due to its low optical density relative to other atmospheric gases like NO_2 and SO_2, which makes detection more sensitive to noise and spectral interference. A two-step DOAS retrieval process is applied, using wavelength intervals of 328.5-359 nm and 328.5-346 nm, to effectively de-correlate HCHO from BrO and minimize noise. The broader wavelength window reduces slant column noise while maintaining accuracy by addressing potential biases from interfering absorbers like O_4 and O_3.

SO_2 The slant column of SO_2 [47] is derived using multiple fitting windows, such as 312-326 nm for low concentrations and 325-335 nm or 360-390 nm for higher concentrations, to optimize retrieval accuracy. The observed radiance data is fitted to a model that accounts for various absorbing species and their densities, using least-squares fitting techniques to estimate the slant column density. An offset correction is also applied to reduce outliers.

Sensor and Solar angles

The zenith and azimuth angles for both the sensor and Sun influence TROPOMI's measurements of gas concentrations. The sensor (or viewing) zenith angle measures between the local vertical at the observation point and TROPOMI's line of sight.

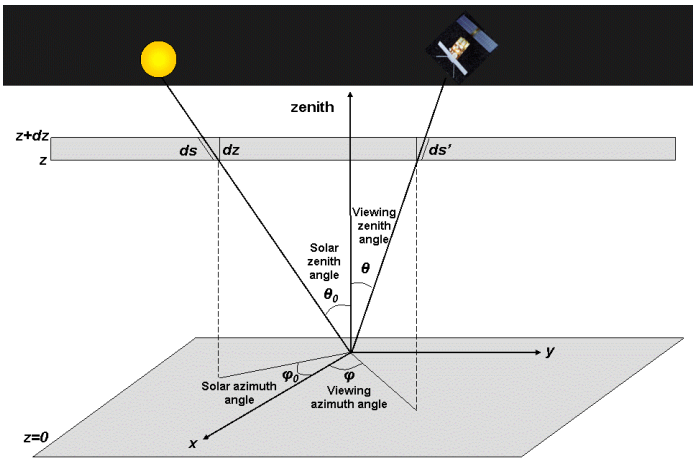

Figure 2.2: Visualisation of zenith angle θ and azimuth ϕ angle. It can be either for the sun or a satellite. Image from: *Algorithm theoretical basis document for the TROPOMI SO_2 data processor* [47]

Larger angles result in more oblique viewing angles and longer atmospheric paths, which affect measurement accuracy.

The solar zenith and azimuth angles define the Sun's position relative to the observation point, impacting the intensity and directionality of incoming solar radiation. These angles particularly affect measurements over water surfaces due to specular reflections. By accounting for these geometric parameters in our analysis, we can better distinguish actual gas concentration variations from measurement artifacts caused by viewing geometry.

Measurement time

For each TROPOMI measurement, the time is given in seconds since 2010-01-01, UTC midnight [11].

Meteorological data: wind

Wind data in TROPOMI data products are supplied by the ECMWF [11]. The wind data is at 10 meters in height, which is lower than the funnel of big seagoing ships and thus the plume. ECMWF wind data is considered to be highly accurate.

The wind data is supplied in a horizontal and vertical component ECMWF GRIB (General Regularly distributed Information in Binary form) parameter ID 165 and 166, respectively). The horizontal component represents the zonal wind velocity in m/s. In contrast, the vertical component represents the meridional wind direction in m/s. They can be positive or negative depending on direction. The zonal wind blows

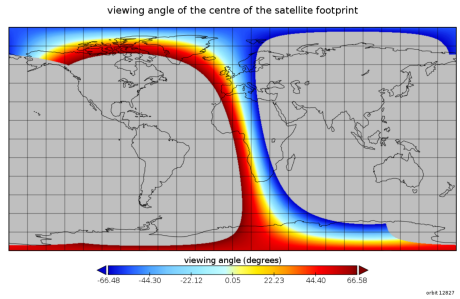

Figure 2.3: Sensor zenith angle variable of TROPOMI data product. [49]

east to west, while the meridional wind blows north-south. For example, a zonal wind of -5 blows in the western direction coming from the east.

Each TROPOMI grid cell is accompanied by a meridional and zonal wind value; for an illustration of the grid cells, see figure 2.1.

QA value

Each TROPOMI measurement is accompanied by a Quality Assurance (QA) value. The QA value indicates the usefulness of the measurement as a fraction from 0 (useless) to 1 (useful) and is influenced by cloud cover, sun glint, and snow coverage. The calculation of the QA value is described in the NO_2 ATBD [51].

Cloud cover

Cloud cover can introduce a lot of noise into TROPOMI trace gas measurements. The cloud cover support product in TROPOMI data ranges from 0 (no clouds) to 1 (completely covered). It is useful for interpreting TROPOMI measurements.

2.2.4 Orbit files

Each TROPOMI orbit has its own file using the extension *.nc*, also known as a *NetCDF* [40] file. This file can be read using the GUI application *PANOPLY* developed by Nasa. The files can also be observed using *EOBROWSER* or *SENTINELHUB BROWSER*. These files consist of all the variables previously mentioned. Each orbit has multiple measurement points, bounded by latitude and longitude. Aside from every orbit, the measurements of every gas TROPOMI measures are also divided into a separate orbit file.

Finding orbit files

Catalog API, part of the SentinelHub API, offers a fast API for retrieving TROPOMI orbit files and metadata about the files.

Downloading orbit files

Up until October 2023, Copernicus Open Access Hub [9] was the place for downloading TROPOMI orbit files. This service has been migrated to Copernicus Data Space Ecosystem [8], where a user can search and download orbit files using a web GUI. For bulk downloading, the current option is the Amazon Simple Storage Service (S3) API. S3 [24] is a cloud storage service.

Processing streams

Near-real-time (NRTI) Data is a processing stream focusing on speed while sacrificing data quality. It is available within 3 hours of measurement [51].

RePROcessing (RPRO) Data: provides historical TROPOMI orbit files updated to the latest algorithm version. RPRO exists to update and improve historical data to match the quality and accuracy of newer data processed with the latest algorithms, where the historical OFFL orbit files can be created using an obsolete algorithm version [27]. Offline (OFFL) Data is a processing stream that is processed at KNMI for more recent orbit files compared to RPRO. OFFL and RPRO are processed using the TM5 (Tracer Model version 5) algorithm described in KNMI's NO_2 Algorithm Theoretical Basis Document (ATBD) [51], although, for each version, the algorithm can change slightly. Meanwhile, the OFFL stream provides recent data. For data after 25 July 2022, no RPRO data is available, whereas OFFL data is available from 17 July 2022 to 12 March 2023 on version 2.4.0 and up until the present on more recent versions.

Data level

The data products are available in levels 1, 2, and 3. The NO_2, SO_2, and HCHO data products are classified as Level-2, which are derived using the Level-1b spectra through various processing steps to generate the final NO_2 data products. Level 2 data is provided in a sparse geolocated matrix representation. Level 3 data means regridded data. Regridded data is similar to an image on a latitude/longitude grid. When you plot level 2 data, you will see dots on the map at the center of the measurement location. When you plot level 3 regridded data, the measurements are projected within the bounds of the measurement, leaving no empty space in between measurements.

Several collections of TROPOMI data products are available. The collection number is currently linked to the version of the level-1b (ir)radiance spectra [27]. As recommended by KNMI, collection number 3 has been used, and version number 2.4.0 [13] (also denoted as processor version 02.04.00) has been used, as well as the processing stream RPRO.

2.3 Ship Location and characteristics

In compliance monitoring, it is common to use ship location data. From the AIS data, we obtained ship locations, while ship characteristics were available through ShipDB

(SDB). Both datasets were supplied by Inspectie Leefomgeving en Transport (ILT) under a non-disclosure agreement.

2.3.1 AIS

AIS provides ship location data. Vessels over 300GT (Gross Tonnage) and all passenger ships have been obliged to be equipped with an AIS transponder since 2005. AIS provides the *Latitude*, *Longitude*, and *Timestamp*: where the date and time are given in the UTZ timezone.

Maritime Mobile Service Identity (*MMSI*) is a 9-digit code that identifies ships and coast radio stations. Aside from AIS, they are also used in Digital Selective Calling, which allows captains to send distress messages. The International Telecommunications Union (ITU) oversees *MMSI* assignment and regulation [36]. The *MMSI* identifiers are not unique to each ship. There is limited availability of *MMSI* numbers for each country. When a ship is scrapped, its *MMSI* can later reappear under a different ship.

CallSign is a code, e.g., A1AA1, that uniquely identifies ships. The code is assigned by the local authorities of the ship's home harbor.

The IMO has a unique identifier for each ship. This identifier is provided as the *IMO* variable.

Tier is the classification level of maximum NO_x emission imposed by the IMO [37]. Newer ships are in tier 3 and have stricter NO_x emission regulations. The variable *Name* is the name of the ship, e.g., "Ever Given", the container ship infamously stuck in the Suez Canal in 2021.

Latitude and *Longitude* give the position in coordinates. The variable *Speed* denotes the speed in knots of the ship at the moment of the AIS sample. This moment is given by variable *Timestamp*, which gives a date and time in Universal Timezone (UTZ) format.

The sampling frequency of ship positions in the AIS database ranges from 5 samples per hour to 0.5 per hour, see figure 2.4.

2.3.2 ShipDB

From SDB, we used the following information: the identifiers: *MMSI*, *CallSign*, *IMO*, *Name*, and the following characteristics date the ship was built, the length of the ship, type of ship (e.g. cargo), GT (gross tonnage) and TIER.

2.3.3 NO_x proxy

To evaluate TROPOMI measurements of ship emissions, Georgoulias et al. (2020) introduced a NO_x emission proxy that could estimate emissions based on only AIS data.

$$E_s = l_i^2 \cdot v_i^3 \tag{2.1}$$

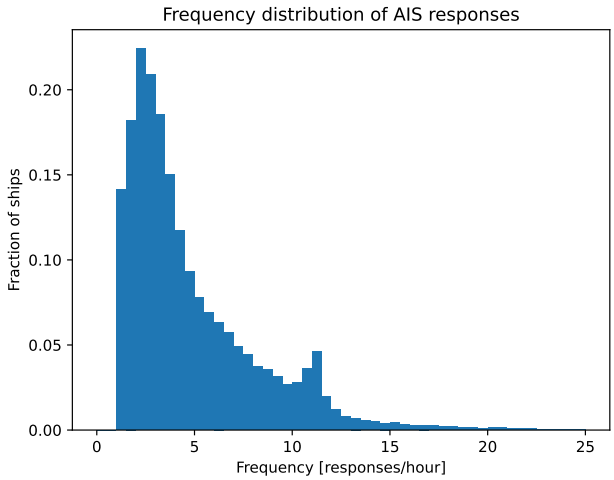

Figure 2.4: Distribution of AIS sample frequency

Where E_s is the NO_x emission of ship s, l_s is the length of ship s in meters and v_s is the velocity of ship s in meters per second. the emission proxy does not have a specific unit, as it is a combination of the ship length and speed.

2.4 Machine Learning

2.4.1 Supervised Machine Learning

In this study, we investigate the detectability of ships using a classification task with supervised machine learning. Supervised learning consists of learning model parameters θ that can predict a class label y using features X. To be able to generalize to new data, the data is typically split into a train and test set. In the training phase, the model parameters θ are fit to predict y_{tr} based on X_{tr} The model θ is then used to predict the class y_{te} of unseen data X_{te}. The difference between the predictions and the actual targets y_{te} can be used to evaluate the performance of the model. The algorithm parameters set by the user are referred to as HyperParameters (HPs): e.g. learning rate.

2.4.2 XGBoost Classifier

XGBoost (XGB), which stands for *eXtreme Gradient BOOSTing* [5], is the machine learning algorithm used in this study. XGB is a supervised machine learning algorithm suitable for both regression and classification tasks. In this study, we will use the XGB classifier to predict the presence of a ship plume. Kurchaba et al. (2024) [28] have

found XGB to be the best performer for the problem of ship plume detection compared to other machine learning models.

XGB is an improvement upon Gradient Boosting [15]. Gradient Boosting uses an ensemble of decision trees to do classification. XGB creates decision trees iteratively and adjusts them based on the errors of the previous trees. In XGB, regularization prevents overfitting and handles missing values. XGB prunes trees by removing splits that do not contribute to a positive gain in model performance

2.4.3 Hyperparameter Optimization

Meta-learning [4] is a subject in machine learning that encompasses the learning of the machine learning itself. Traditionally in Machine learning, HPs have to be chosen by an expert or tuned manually. Meta-learning automates the tuning of HPs. An example of a hyperparameter is the learning rate in a neural network; a parameter can be the weights of a neural network. A subset of Meta-learning, Hyperparameter Optimization (HPO), searches for the best HPs on a validation set, turning HP selection into an optimization problem. Three popular search algorithms for HPO are grid search, randomized search, and Bayesian Search (BS) (also known as Bayesian Optimization [42]). Grid search needs a discrete search space and will try out all possibilities. Randomized search can handle continuous values and search a set number of random configurations. BS uses an evaluation metric to assign probabilities of goodness to each configuration and will try more probable good hyperparameter configurations.

When performing HPO, the traditional train-test split in machine learning is not sufficient. A validation set is necessary to evaluate the algorithm hyperparameter configuration while tuning the hyperparameters.

The background chapter has described the problem of maritime pollution. How earth observation data can be used for compliance monitoring. What TROPOMI data entails, as well as information on the ship location and characteristics data. Because we use XGB with HPO, machine learning, and the relevant techniques have been described. Some of these techniques have been used in similar studies, which will be expanded upon in the next chapter: Related Work.

Chapter 3

Related Work

This work is built upon recent advancements in ship plume detection and their NO_2 emissions using satellite data. This chapter describes the works closely related to this study, mainly studies where machine learning has been applied to TROPOMI NO_2 data. We also present related works in the domain of earth observation.

3.1 Ship plume detection

3.1.1 Temporal averaging satellite measurements

Yuan et al. (2023) [58] developed a convolutional neural network that segments shipping lanes based on NO_2 data. It does not, however, segment individual ship plumes. The difficulty of manual labeling for this task was mentioned as the reason. This manual labeling problem could be solved by using AIS data to derive the ship plume locations.

Luo et al. (2024) [32] also made a regression model for estimating regional NO_x emissions from ships in a region. This model can be used to monitor inter-regional differences in emissions.

From HCHO measurements with TROPOMI, shipping lanes have been found [6] when averaging data over 3 months. This detection limit is down from 5 years with OMI measurements. This proves that TROPOMI can measure HCHO emission from ships, which is why we hypothesize that it could be useful as a feature in a prediction model.

3.1.2 Individual satellite measurements

Georgoulias et al. (2020) [18] first showed that individual ship plumes can be observed in TROPOMI NO_2 measurements. They chose a day with sun and no cloud cover. Using AIS and wind data, they estimated the plume locations within three hours prior to the TROPOMI overpass. They projected the estimated Wind-Shifted Ship Tracks (WSSTs) over a TROPOMI image and observed that the WSSTs line up with higher

NO_2 concentrations in the TROPOMI image. From fluid dynamics, they derived an NO_x emission proxy that can be used to estimate the NO_x emission from a ship based on its length and speed. When comparing the NO_2 concentrations at the plume locations, they observed a correlation of $R = 0.88$ between NO_2 and the NO_x proxy.

Building on Georgoulias' findings, Kurchaba et al. (2021) [29] introduced an automated approach to improve the estimation of plume locations, observing a higher correlation between the NO_x proxy and the TROPOMI NO_2 concentrations at the plume locations. Pearson correlation improved from 0.73 to 0.63.

In 2022, Kurchaba et al. [30] presented a supervised machine-learning approach for segmenting plumes emitted by individual ships. Their methodology utilizes the XGBoost algorithm to classify whether TROPOMI pixels within a *ship sector* contained a ship plume. The *ship sector* is an area around the expected plume location based on AIS data and wind conditions. This area accounted for margins of error within calculating the ship plume location. However, the reliance on human-labeled data for training—without actual ground truth—introduces potential label errors and subjectivity, which may affect model robustness.

In 2023, Kurchaba et al. [31] developed a machine-learning method for detecting anomalous NO_2 emissions from ships. Their study proposes an algorithm that uses regression to predict NO_2 emissions from a ship and then calculates the difference between this prediction and the actual emissions from TROPOMI data. This difference generates a metric called "inspection worthiness." The ship is flagged for inspection if the actual emissions significantly exceed the predicted emissions from the regression model. Importantly, this method does not require manual labeling.

The 2024 paper by Kurchaba et al. [28] is closely related to our thesis. It focuses on establishing sensitivity limits for detecting ships at sea in four different areas, among them a region in the Mediterranean Sea, which is the main focus of our thesis. The sensitivity limits are defined in terms of ship length and speed. The authors used *proxy thresholding* where a lower threshold, based on the expected NO_x emissions, filters out samples from the data set from smaller or slower ships. The main goal is to determine the length and speed at which the model can no longer distinguish ships from TROPOMI data. A disadvantage of this approach is that high-proxy measurements are also included at a lower threshold. Creating an overlap of data between higher and lower proxy thresholds.

Riess et al. (2024) [41] used PARANOX, a plume chemistry model, and XGBoost to estimate NO_x emissions from NO_2 TROPOMI measurements. They notice good detection conditions in the Mediterranean in summer.

3.1.3 Other methods for compliance monitoring

Aside from earth observation, TROPOMI-like sensors can also be used from the coast to measure ship compliance remotely [16].

SCIPPER (Shipping Contributions to Inland Pollution Push for the Enforcement of Regulations) [35] is an EU-funded project that encompasses using remote sensing with the aim of compliance monitoring. It finished successfully in 2023. The main goals of SCIPPER were to provide evidence of the performance and capacity of different

techniques for monitoring shipping emissions and enforcing regulations and to assess the impacts of shipping emissions on air quality under different regulatory enforcement scenarios.

Aside from satellite measurements, Belgian authorities fitted SO_2 and NO_x sniffers on coastguard aircraft. With the goal of compliance monitoring for the MARPOL Annex VI regulations, the Coast Guard measures emissions in and around the harbor of Antwerp [53].

This study is based upon previous works, contributing to knowledge in the field on the value of including other gases in addition to NO_2 for ship plume detection. The related works show potential in this multi-gas approach. The respective approach will be detailed in the following method chapter.

Chapter 4

Method

In this chapter, we explain our method to detect the presence of ship plumes in single TROPOMI orbits. We define this problem as a binary classification task, where the task is to predict whether ≥ 1 or 0 ship plumes are present in an $80km \times 80km$ sample patch. We create a tabular aggregate database from the TROPOMI, SDB, and AIS data, where each row in the table represents one of these $80km \times 80km$ areas. From TROPOMI data, we get the features. From the AIS data, we get the ship tracks: the path each ship in the AOI has traveled in the two hours prior to the TROPOMI overpass. Using ECMWF wind data we shift the ship tracks to get the location of the WSSTs.

The method has three main parts: Downloading and merging TROPOMI orbit files. Preprocessing AIS, where for each orbit file, we get the ship tracks, WSSTs, and NO_x proxy value for every ship in the Area Of Interest (AOI). Finally, we create a tabular database by sampling patches from the AOI and assign all data belonging to the sampled patch to a row in the tabular database.

Section 4.1 describes how the data is collected and describes the AOI in more detail. Section 4.2 goes into detail on how the AIS data is aligned with the TROPOMI data and turned into a machine learning database, while section 4.3 describes the machine learning setup.

4.1 Data acquisition

The first step in the study is acquiring the data. TROPOMI data is available through the Sentinelhub, which has several APIs. With the Processing API, one can query an image if more metadata is available, like capture time per pixel. This could have been used in this study. Catalog API offers a method to query an AOI and time and get a list of orbit files.

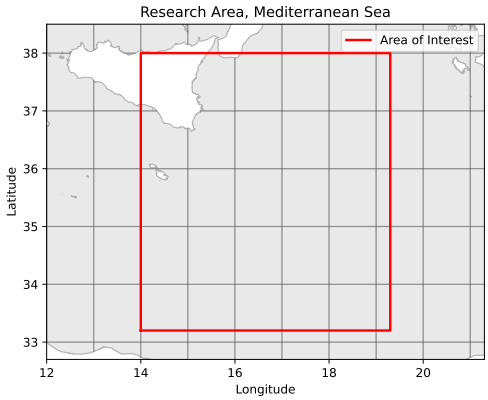

Figure 4.1: Illustration of this study's AOI: The AOI is bound by a minimum latitude of 33.2 and a maximum latitude of 38. The minimum longitude is 14, while the maximum longitude is 19.3. It is an area in the Mediterranean Sea southeast of Sicily, Italy.

4.1.1 Area Of Interest & Time period

The AOI is the Mediterranean east of Sicily. We sample $80km \times 80km$ image patches from this area, which has busy shipping lanes for trade between Europe and Asia. The time period studied is Jan 2020 - Jul 2022. The AOI is bound by a minimum latitude of 33.2° and a maximum latitude of 38°. The minimum longitude is 14°, while the maximum longitude is 19.3°. This is the same area as is used by Kurchaba et al. in their 2024 study investigating TROPOMI sensitivity limits [28]. This study is similar but focuses on the impact of adding SO_2 and HCHO data as a machine learning feature.

4.1.2 AIS & ShipDB

Automatic Identification System (AIS) [43] data in this study is provided by ILT under a non-disclosure agreement. Part of the AIS preprocessing is therefore not explained in detail, additionally to this data not being published.

ILT also kindly provided SDB: a database containing information about ships. We use SDB to get the length of the ships in the AIS database. With the length, we can calculate the NO_x emission proxy that is used in several of our experiments.

4.1.3 TROPOMI

Using the catalog API, we built a database containing 740 (January 2020- July 2022) orbits with full coverage for NO_2, SO_2, and HCHO. We also retrieved the S3 URLs to download the TROPOMI orbit files. To obtain the files we queried based on the following file characteristics:

Time period: Since we have AIS data for 2020-2022, we query for orbit files within that time period.

Processing stream: Recommended by KNMI [27], we only use files from the RPRO processing stream.

Version & Collection number: Another filter in the Catalog query is the version and collection number. We use orbits from version 2.4 and collection 3, as recommended by KNMI [27].

AOI: With the catalog API, we were able to query for all the orbit files that fully covered the AOI, meaning that the intersection of the AOI and the orbit was equal to the AOI.

Gases: We want to create a database containing all orbits covering the AOI and time period. We query for NO_2, SO_2, and HCHO orbit files.

Using the S3 API, we were able to download all orbit files we retrieved using the Catalog API. As the size of an orbit file can range from 300MB to 600MB, we only retained measurements in the AOI to cut file size before moving on to downloading the next file.

4.2 Data preprocessing

4.2.1 Data products used in preprocessing and machine learning

This section describes the features used for the classification problem. In addition to the gas concentrations, support features like wind, sensor, and solar angles have also been derived from the TROPOMI data products.

Gases

The main features used in the study are the gases NO_2, HCHO, and SO_2. For these gases, we use the SCD [52]. SCD is expressed in molec./cm^2.

The reason we don't use VCD is because the algorithm calculating the VCD is influenced by historical data of the respective gas [11]. Using VCD would introduce the risk of data leakage. This is also why other studies [28] used SCD instead of VCD for ship plume detection using machine learning. Using VCD in this study would require a train-test split method that prevents samples that are within 7 days of each other from being in both the train and test set.

4.2.2 AIS & ShipDB

In this section, all the steps performed during the preprocessing of AIS and SDB data will be described. For each TROPOMI orbit, we preprocess the AIS data to get the ship tracks and WSSTs from the two hours leading up to the TROPOMI overpass. We use this lookback of two hours as used by [31]. This lookback is based on NO_2 dispersing after about two hours [56].

As mentioned in section 2.3.1, *MMSI* identifiers can appear for multiple ships, if there are duplicate *MMSI* values, we keep the newest ship based on year of build, while discarding ships built after the TROPOMI overpass date.

Missing values

AIS and SDB can contain missing values like identifiers or length; therefore, we use SDB and AIS to cross-reference or interpolate these missing values. This allows us to merge as many ships in the AIS database as possible with ships in SDB, where the length of ships is registered. This length is needed to calculate the NO_x emission proxy.

Identifiers like *IMO*, *MMSI*, *CallSign*, and Name can contain missing values in either SDB or AIS. For instance, some ships have *MMSI* missing in AIS, while *IMO* is missing in SDB. Using every instance in the AIS database and SDB, we create a mapping of these variables for cross-referencing: *IMO to MMSI* mapping from AIS data. *MMSI to IMO* mapping combining AIS and SDB. *CallSign and Name to MMSI* mappings from combined datasets. Using these mappings, we fill in missing identifiers.

As length was not available for every ship, we interpolated the missing length values based on the relationship between GT and length calculated from the SDB. We found the polynomial that mapped length to GT being:

$$l_s = 8.494e^{-14} \cdot GT^3 - 3.223e^{-08} \cdot GT^2 + 0.004532 \cdot GT + 84.66 \tag{4.1}$$

Where l_s is the length of ship s, and GT is the weight of the ship in metric tonnes.

If the length and GT are both missing, we fill in the length as the mean of the length of other ships with the same ship type and in the same TIER.

Since we filter out ships based on length, filling in the length using these steps prevents missing ships in the AIS data that we use to derive the targets of the sample patches used to create the tabular database. This comes with the downside of reducing the reliability of the NO_x proxy sum since it is calculated using the length and speed of all ships in a sample patch.

Data Filtering and Merging

Preprocessing has been done to remove all vessels not in the categories {Tanker, Bulk, Container, Passenger}. Because a minimum length of 90m is applied, smaller ships will not be considered in this study. This is done to protect the privacy of small, private, or military vessels.

With the aim of removing docked ships, we filter out ships that do not travel with a minimum speed of 6 knots. This speed is determined by taking the average speed of all samples of a ship within two hours before up until the TROPOMI overpass. These filters on length and speed have also been used by Kurchaba et al. (2024) [28].

We finally merged AIS and SDB based on *MMSI*, *IMO*, and *CallSign*.

Ship track

With the aim of eventually estimating the location of a ship plume, we reconstruct the path of every ship in the AOI that the ships have traveled before the TROPOMI overpass. Based on the dispersion of NO_2 [56], we reconstruct the tracks from the two hours leading up to the TROPOMI overpass.

First, we need to estimate, the time of TROPOMI overpass t_{S5P}. We calculate t_{S5P} for every orbit file by taking the mean timestamp of all TROPOMI measurements within the AOI. Due to S5P's speed of 7km/s (see figure 2.1, the margin of error for this simplification is in the order of 1-2 seconds.

To reconstruct the path of a ship during the two hours preceding the S5P overpass (t_{S5P}) and at the time of the overpass itself ($t - 2 \cdot 60^2$), we use all available AIS measurements within this period. Since the sample rate of each ship in AIS is variable (Figure 2.4), we do not have access to the precise location at the time of the overpass $\{x_{\text{ship}}(t_{S5P}), y_{\text{ship}}(t_{S5P})\}$ and $\{x_{\text{ship}}(t - 2 \cdot 60^2), y_{\text{ship}}(t - 2 \cdot 60^2)\}$.

To estimate the ship's location at t_{S5P} and $t - 2 \cdot 60^2$, we interpolate between the points recorded just before and after both points as such: given timepoint t_i, we identify the nearest measurements before and after the time t_i as t_{i-1} and t_{i+1} respectively. The ship's position at t_i is then estimated by assuming a constant speed between these two points. The position at t_i is calculated by scaling the difference between the coordinates at t_{i-1} and t_{i+1} based on the fraction of time elapsed between these two points using the following equations:

$$x(t_i) = x(t_{i-1}) + \frac{(t_i - t_{i-1})}{(t_{i+1} - t_{i-1})} \cdot (x(t_{i+1}) - x(t_{i-1})) \tag{4.2}$$

$$y(t_i) = y(t_{i-1}) + \frac{(t_i - t_{i-1})}{(t_{i+1} - t_{i-1})} \cdot (y(t_{i+1}) - y(t_{i-1})) \tag{4.3}$$

An assumption made, however, is that the speed is constant between these two points. This interpolation is not applied in previous work [28] and is thus expected to improve performance slightly, as it will make the target data more accurate.

Deriving the Wind-Shifted Ship Track

The previous section described how we extracted the ship tracks that every ship in the AOI has traveled from two hours before to the time of the TROPOMI measurement. However, this ship track only aligns with the exhaust plume at the point $\{x_{\text{ship}}(t), y_{\text{ship}}(t)\}$ where t is the time of the TROPOMI measurement. To accurately estimate the dispersion of ship exhaust plumes, we apply a wind shift model that calculates the displacement of the plume based on wind data, ship location, and the time difference between the ship's position and the satellite overpass.

The time difference Δt between the ship's timestamp t_{ship} and the satellite overpass time t_{S5P} is calculated to determine how far the plume has traveled during this period. We assume the gases in the plume have no resistance and travel with wind speed.

Assuming the Earth is a perfect sphere, we approximate the distance per degree of latitude or longitude as $D = 111139$ meters per degree.

This wind shift method has been used in previous studies [28, 18]. The plume's shifted longitude x_{plume} and latitude y_{plume} are calculated using the following equations:

$$\Delta t = t_{\text{S5P}} - t_{\text{ship}} \tag{4.4}$$

$$x_{\text{plume}} = x_{\text{ship}} + \frac{V_{\text{zonal}} \cdot \Delta t}{D} \tag{4.5}$$

$$y_{\text{plume}} = y_{\text{ship}} + \frac{V_{\text{meridional}} \cdot \Delta t}{D} \tag{4.6}$$

Where x_{ship} and y_{ship} are the longitude and latitude of the ship's original position (in degrees), V_{zonal} and $V_{\text{meridional}}$ are the zonal and meridional wind velocities at the ship's location (in meters per second) obtained from the TROPOMI data, Δt is the time difference in seconds, and $D = 111139$ meters per degree is the approximate conversion factor from meters to degrees. This conversion factor is a simplification specifically for the longitude degrees, as mentioned in the discussion: section 6.4.

An example of applying this method is illustrated in figure 4.2, where the original ship and the calculated WSSTs are shown.

NO_x proxy

Each WSST has its accompanying NO_x proxy value based on the length of each ship and the speed of the last AIS sample of the ship before the TROPOMI overpass.

4.2.3 Building the tabular database

The tabular classification consists of drawing samples from satellite data and aggregating the data from those samples into a tabular database. From AIS, the presence of a ship plume in the sample is deduced from the intersection with the sample patch inner area and the WSSTs and added as a label for training a classifier.

Sampling pictures / TROPOMI: regridding

The TROPOMI data is available in a NetCDF-5 file that is indexed by timestamp. Using *S[&]T* 's *HARP* toolkit, an image can be regridded to a level 3 image. This means that a 2d image is constructed on a flat map representation. The coordinates used are WGS84 (World Geodetic System 1984) coordinates [34], which is a global reference system for geographic coordinates. It was developed by the United States Department of Defense and has become the standard for GPS (Global Positioning System).

For each orbit, we sample 50 random images (sample patches) of about $80km \times 80km$. If the ship plume intersects with the center $60km \times 60km$ area, we consider a ship present in that area. These sample patches can overlap, which is why we ensure samples from the same orbit do not end up in both the train and test set, as will be explained in section 4.3.1. In previous work [28], a sliding window resulting in a fixed grid of non-overlapping patches was used. Both with the aim of generating more data and reducing the risk of overfitting on specific grid cells, we chose to sample patches at random locations. The only condition is, that the sample patch must not overlap

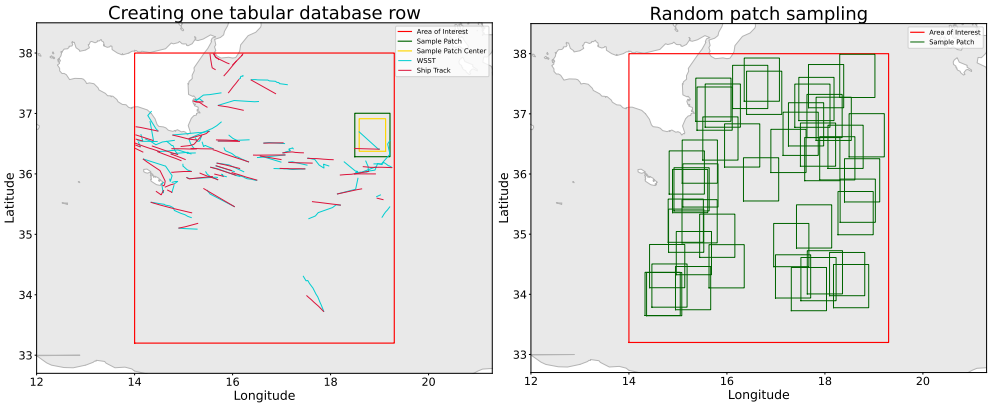

Figure 4.2: In the left plot, one random sample patch in the AOI is plotted along with the tile center. Along with the ship tracks, WSSTs, and the AOI. On the right, 50 random sample patches are plotted to visualize how the random samples are drawn to create the tabular database. This sampling is done randomly for every orbit.

with any land area, as the aim of the study is to investigate the detectability of ships based on emission gases.

Filters

After sampling 50 random images of $1^\circ \times 1^\circ$, we assign NaN values to TROPOMI measurements that were not measured in good conditions. For example, cloud cover decreases the quality of TROPOMI measurements, as well as high wind.

Measurement quality While KNMI recommends a filter of 0.75, in this study, we use a filter of 0.5 to prevent disposing of too much data as done by Kurchaba et al. (2024) [28]. This gives us more data to work with but decreases the quality of the data. The TROPOMI measurements that do not meet this condition are set to NaN values.

Cloud cover TROPOMI measurements are accompanied by a cloud cover variable. This assigns a 0 to 1 value of the fraction of cloud cover. When the cloud cover exceeds 0.5 in a TROPOMI measurement, we assign a NaN value to the respected TROPOMI measurements.

Wind If wind speed exceeds 10m/s in any direction, we assign a NaN value to all TROPMOMI measurements in this TROPOMI pixel.

If more than 50% of the sample patch is made up of NaN values, we do not add data from that sample patch to the tabular database. In Figure 4.3, the TROPMI NO_2 data can be observed, regridded to level 3, on a clear day with little cloud cover.

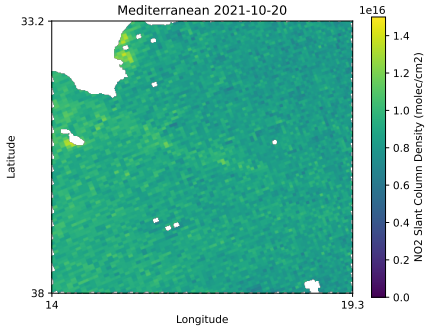

Figure 4.3: NO_2 image with little cloud cover

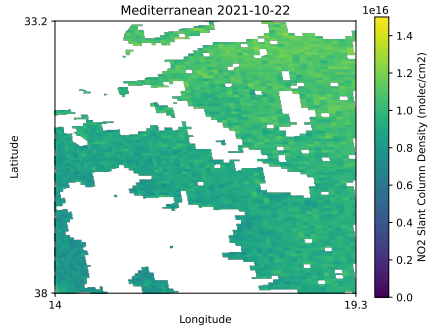

Figure 4.4: Image with cloud cover, NaN values are illustrated by white area.

Features

Gases For NO_2, SO_2, and HCHO, we have SCD values. We aggregate all these values from each sample into multiple features: min, mean, median, max, and standard deviation. Each gas thus comes with five features; when using all gases, there are 15 features along with six support features.

Support features For support features (wind meridional, wind zonal, sensor zenith, sensor azimuth, solar zenith, and solar azimuth angles), we use only the mean value of all measurements within a sample patch. These measurements are spatially consistent at the $80km \times 80km$ scale, with low variance compared to gas concentrations, which can vary due to local emissions.

Feature Category	Details
Gases	NO_2: min, mean, median, max, and standard deviation
	SO_2: min, mean, median, max, and standard deviation
	HCHO: min, mean, median, max, and standard deviation
Wind	Wind meridional: mean
	Wind zonal: mean
Angles	Sensor zenith: mean
	Sensor azimuth: mean
	Solar zenith: mean
	Solar azimuth: mean

Table 4.1: Feature vector components

Target

Every row in the tabular database represents a $80km \times 80km$ sample patch. If a WSST lies (partially) within the inner $60km \times 60km$, the target is 1. Otherwise, it is 0. Table 4.2 shows the target balance.

Count	0 (No WSST)	1 (One or more WSST(s))
Frequency	11,374	13,858
Percentage	45.07%	54.93%

Table 4.2: Label balance for tabular dataset

The majority of random samples have ship presence. The AOI in the Mediterranean has been chosen because it contains a busy shipping lane. From the TROPOMI data, we derive multiple gas features as well as support features. This process is described in the next section.

Other variables

Not used as a feature for the XGB algorithm, but important for performing the experiments are the NO_x proxy and the orbit date belonging to each sample patch.

NO_x proxy From the ship specifications in the SDB and an approximation of the speed of the ship, the NO_x proxy is calculated. The NO_x proxy values belonging to all WSSTs in the sample patch were summed up. The proxy values were summed because gases are additive, the more and the faster ships are in an area, the higher the concentration of NO_x. This serves as an approximation of the total amount of NO_x emitted by the ships in the samples.

Orbit date Not used as a learning feature, but used to create train-test-validation splits, we record the orbit date of every sample patch. This will be expanded upon in the next section: section 4.3.1.

4.3 Tabular classification

The previous section described the process of making the tabular database usable for classification with a machine-learning model. This section describes the process of performing classification and how the performance is measured. We use the `xgboost-python` classification algorithm, with optimized hyperparameters. Along with SciKit-Learn (sklearn) [39] for the cross-validation.

4.3.1 Train-Test split: Cross Validation

An important part of machine learning is the train and test split. We want to ensure there is no data leakage between the two. Because we are drawing many random

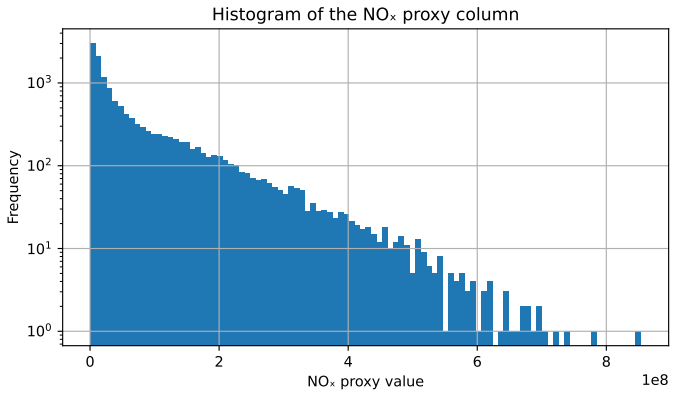

Figure 4.5: Distribution of the NO_x proxy values in the tabular database. For samples where a ship is present. For samples without ship presence, the proxy value is 0

sample pictures from a grid for each orbit, we want to ensure that pictures from the same day do not end up in both the training and test sets. This could cause the algorithm to train on data that is so similar to test data that however high the accuracy is, this ability to predict ship presence will not apply to real-world scenarios where unseen data is used.

In this study, we use a StratifiedGroupKfold (SGKF) split and GroupShuffleSplit (GSS). A grouped-based train-test split mitigates data leakage by using the "group," in this case, the date, and ensuring that there are no data points in the train set from the same orbit as any data points in the test set. SGKF goes one step further and creates train-test-splits that retain the class balance of the dataset in the train and test splits. SGKF also allows shuffling. If groups (dates) are ordered in a particular way (e.g., chronological order), splitting without shuffling can introduce bias. For instance, the training set might contain only earlier dates, and the test set might contain only later dates, effectively turning the cross-validation into a time series split. GSS is similar to SGKF but gives control over the training and test set size in their sklearn implementations.

4.3.2 Proxy binning

Proxy binning is the approach of dividing the dataset into bins based on the NO_x proxy [18] sum of each sample patch respectively. We perform proxy binning to analyze the sensitivity of ship plume detection to the speed and or size of ships.

Each row in the tabular database has an accompanying proxy value. For example, when we create a bin with a window center of 1.5e8 and a window size of 1e8. We remove all samples in the database that have a non-zero proxy value lower than 1e8 and higher than 2e8. We end up with about 5000 samples within this proxy bin. Since a non-zero proxy value requires ship presence, we can sample 5000 image patches where no ship is present to end up with a class-balanced database.

Why proxy bins instead of proxy thresholds In our study, we adopted a different approach than the proxy thresholds described by Kurchaba et al. (2024) [28], which involves overlapping data between thresholds. Instead, we utilized the NO_x proxy to create distinct bins. This methodology aims to provide a clearer understanding of the impact of the NO_x proxy value, particularly at lower threshold values while excluding higher NO_x proxy values. By doing this, we can better analyze how estimated NO_2 emissions from shipping affect ship plume detection accuracy.

Kuchaba found a linear relationship between proxy threshold and ship plume detection performance. Throughout the experiments, we test whether this relationship remains with binning on proxy values. We will also see how this relationship is when we include HCHO and SO_2 data alongside NO_2 data.

4.3.3 Hyperparameter Tuning

To ensure that experiments are not conducted on suboptimal models, it is good to tune the hyperparameters of a model before performing experiments on it. We use BS Cross Validation implemented by the Scikit-Optimize project [20] to find the best HPs. The HP search space used in this study is presented in Table 4.3. It is based on the search space used by Kurchaba et al. (2024) [28] for their XGB model, with some adjustments.

For `n_estimators`, `max_depth`, `subsample`, and `learning_rate`, We changed the search space from categorical to continuous ranges to enable the optimizer to explore intermediate values not explicitly listed in the original space.

For `gamma`, the lower bound was extended to include near-zero values because small values can be effective in some cases, and

For `min_child_weight`, we changed the lower limit from 2 to 1, because the default value in `xgboost-python` is 1, so including it might be beneficial. We explicitly specified a uniform prior, which asserts equal likelihood of sampling values within the range.

For `reg_alpha`, we changed the lower limit from 0 to 1e-8, as it is a near-zero value that is compatible with the log-uniform prior.

The `learning_rate`, `gamma` and `reg_alpha` were both modified to continuous ranges with log-uniform priors because a log-uniform prior ensures better exploration of these smaller ranges.

Hyperparameter	Type	Range	Prior
`n_estimators`	Integer	[10, 500]	-
`gamma`	Real	[1e-8, 0.5]	Log-uniform
`max_depth`	Integer	[2, 10]	-
`min_child_weight`	Integer	[1, 12]	-
`subsample`	Real	[0.6, 1.0]	Uniform
`learning_rate`	Real	[1e-3, 1.0]	Log-uniform
`reg_alpha`	Real	[1e-8, 1.0]	Log-uniform

Table 4.3: Hyperparameter Search Space for XGBoost Classifier

Figure 4.6: The shuffled group split that was used in experiments. Visualized on data from the first 20 dates.

We used log-uniform priors for the parameters that span across multiple orders of magnitude. In HPO, it is recommended to explore HP in different orders of magnitude, and this prior forces the BS to explore on an order of magnitude level.

Nested cross validation

We use nested cross-validation to get the best hyperparameters without leaking data. In nested cross-validation, the dataset is split into K folds. Out of each of those K folds, N nested folds are done to tune the hyperparameters of the algorithm. Then, the best hyperparameters out of those N folds are used for the outer loop, where the model is trained and tested on the train and test set of that fold.

Even though available, combining stratification with grouping can be complex due to conflicting constraints. Fortunately, we have a reasonably balanced dataset. When we use proxy binning, we only sample ship-free samples in the same amount of samples available in each proxy bin. Resulting in a data subset more balanced than the original tabular database as described in table 4.2.

In the inner loop, it is more important that we use all available data than in the outer loop. The large number of iterations/splits of the outer loop guarantees that we use next to all data, but in the inner loop, we only have five folds. We want to ensure that all data is being used. Of course, preventing leakage of same-day data points is still important. This is why we use the GSS cross-validator for the outer fold and SGKF for the inner folds. As described in section 4.3.1, we split based on the orbit date.

In our experiments using HPO, we create 15 outer folds for the train+validation/test splits. Then, we create 5 inner folds for the train/validation splits. In the inner folds, we tune the HPs using BS, based on the validation set performance. We perform 30 iterations of BS in every inner fold.

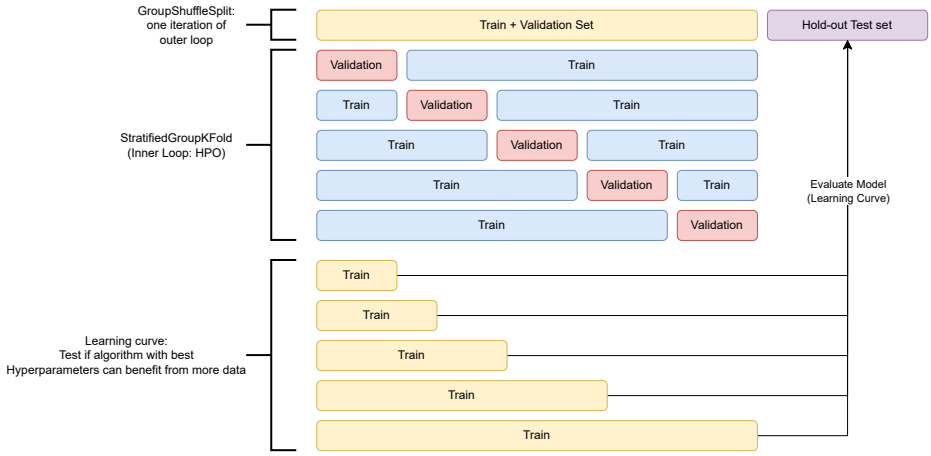

Figure 4.7: An example of how one outer fold of the GSS is used to optimize HPs and how, afterward, those HPs are used on all data except the hold-out test set to test the benefit of more data with a learning curve. In our learning curve experiments, we increase the training set size exponentially. In our experiments, we repeat this experiment for multiple outer folds.

4.3.4 Learning curves

In this study, we use learning curves to assess the effect of acquiring more TROPOMI and AIS data. We create learning curves by iteratively increasing the training set size of the machine learning algorithm. Showing how performance increases when more data becomes available to the algorithm. The algorithm is retrained for every iteration of the learning curve.

4.3.5 Perfomance metrics

ROC AUC The Receiver Operating Characteristic (ROC) [14] curve plots the True Positive Rate (TPR or sensitivity) against the False Positive Rate (FPR or 1-specificity). The ROC AUC is the Area Under the Curve (AUC). The AUC determines the classifier's ability to discriminate between positive (one or more WSSTs in the sample patch) and negative (no WSST) classes. An AUC of 1.0 represents perfect classification, while an AUC of 0.5 indicates performance no better than random chance. ROC AUC evaluates classifier performance across all possible classification thresholds, where metrics like accuracy only consider the decision threshold of 0.5. ROC AUC represents the probability that a randomly chosen positive instance will be ranked higher than a randomly chosen negative instance. While Precision-Recall curves are often preferred for highly imbalanced datasets, ROC AUC is more appropriate for this moderately balanced classification task.

$$\text{ROC AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR}) \tag{4.7}$$

Confidence Interval When we plot learning curves, we use a confidence interval of 95% as bounds around the mean values of the repeated experiments of the learning curves. This shows the variance in model predictions.

$$\text{CI}_{95\%} = \bar{X} \pm t_{\alpha/2,\, df} \times \frac{S}{\sqrt{n}} \tag{4.8}$$

Where $\bar{X}$ is the sample mean, $t_{\alpha/2,\, df}$ is the critical value from the t-distribution corresponding to a confidence level of 95%, and df is the degrees of freedom. n is the sample size and S is the sample standard deviation. The number of degrees of freedom is sample size minus one $(n - 1)$.

This chapter shows all the steps needed to reproduce the ship plume detection, mainly the data preprocessing and the creation of a tabular database. Although we only use XGB in this study, this tabular database can be used for many machine-learning algorithms. The next chapter will use the methods in three experiments, showing the impact of proxy binning, restricting dataset size (learning curves), and HPO.

Chapter 5

Experiments & Results

In this chapter, the experiments are described. We present learning curves binned by proxy values with HPO. We also present learning curves using all data with and without HPO, as well as a table overview of the learning curves when using all data. We also present learning curves of many proxy bins with an equal number of samples as a heatmap. We plot the ROC AUC as a function of the window center of the proxy bins.

The experiments compare different combinations of gases, always with support features: wind and angles. Additionally, the performances on different proxy bins are compared.

The experiments show that WSST detection performance increases when adding SO_2 and HCHO data alongside NO_2 and that performance increases for higher proxies. For SO_2 and HCHO by themselves, the performance is worse than for NO_2, but there is a moderate increase in performance for higher proxies.

All experiments were repeated with a different seed in each repetition to produce reproducible results.

5.1 Heatmaps

In this experiment, we use the NO_x proxy to study the effect of different levels of expected NO_x emissions on the WSST detection performance in a sample patch. Using the NO_x proxy, we implement a sliding window to generate a learning curve in two dimensions. Each aggregated data point with one or more ships present has a summed NO_x proxy value. By moving a sliding window over the dataset as a filter, we can create multiple smaller datasets. Each of these datasets has a different number of data points. To make the comparison fair, we only consider the amount of data points that are available in the smallest subset. Additionally, we sample this same amount of data points out of the no-ship subset to create a dataset with a 50-50 label balance. This results in an image with the window center on the y-axis and a traditional learning curve on every row, with the number of data points on the x-axis. This is done for

each possible combination of NO_2, SO_2, and HCHO. Using the features described in table 4.1.

A linear spacing with 20 steps between 1e8 and 4e8 proxy sum has been used for the proxy window centers, with a window size of 2e8, creating overlapping bins. The reason for choosing 4e8 as the maximum proxy sum and the window size of 2e8 is to perform experiments with a reasonable amount of data. There are fewer very large ships. The maximum training set size in the heatmap is now 947. If we increased the maximum window center or decreased the window size, we would constrain the training set size in all proxy bins to less than 500 samples. This would both decrease performance and make the results less meaningful as the amount of data would not be sufficient to make any concluding remarks. The downside is that a window size of 1e8 would give us more precise insight into the difference in performance for different proxies, which is the reason we used this window size in section 5.2.1.

With the aim of negating the effects of randomness, each experiment is repeated 10 times, improving the robustness of the experiment. Additionally, in each of these 10 repeats, a 5-fold cross-validation was performed using SGKF, splitting based on orbit date. This results in 50 experiment runs from which we plot the mean ROC AUC.

Due to the large number of combinations this experiment has, we do not perform HPO within the experiment. Instead, we use HPs found by performing HPO in a different study using similar data from the Biscay Bay [28]. This ensures that no data leakage occurs while ensuring the experiment is performed using HPs suited for the task. The HPs are as follows: {'subsample': 0.8999999999999999, 'reg_alpha': 0.01, 'n_estimators': 150, 'min_child_weight': 8, 'max_depth': 6, 'learning_rate': 0.1, 'gamma': 0.05}. This is the only experiment where we do not perform the HPO described in section 4.3.3. The 7 heatmaps are spread over two figures, figure 5.1 displays heatmaps for NO_2, SO_2 and HCHO, figure 5.2 displays heatmaps for {NO_2, HCHO}, {NO_2, SO_2 }, {HCHO, SO_2 }, and {NO_2, SO_2, HCHO}.

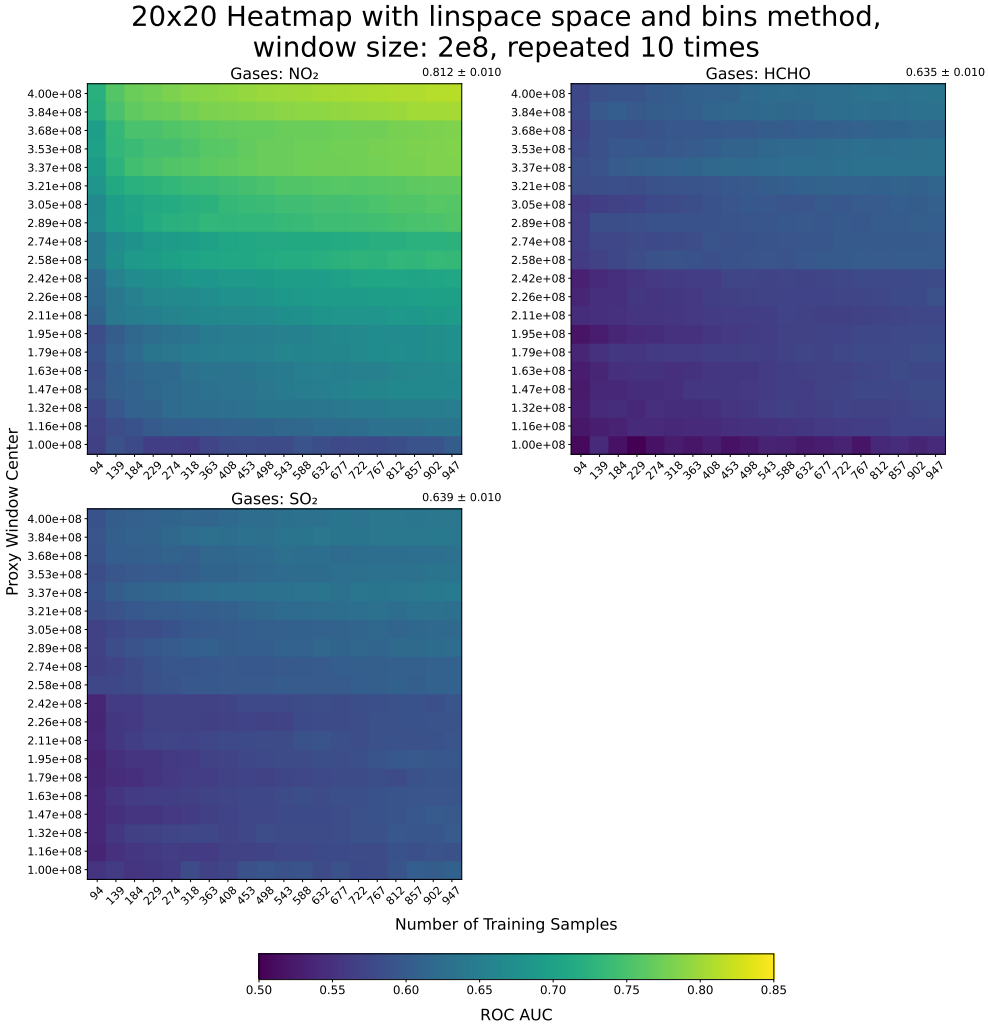

Figure 5.1: Heatmaps for NO_2, SO_2 and HCHO. The heatmaps plot NO_x proxy versus data sets size, where the cell with displayed NOx proxy (p) contains all samples with proxy $p - 1e8$ to $p + 1e8$. The dataset size is increased linearly. The final performance scores for the highest proxies are annotated as $ROCAUC \pm CI_{95\%}$.

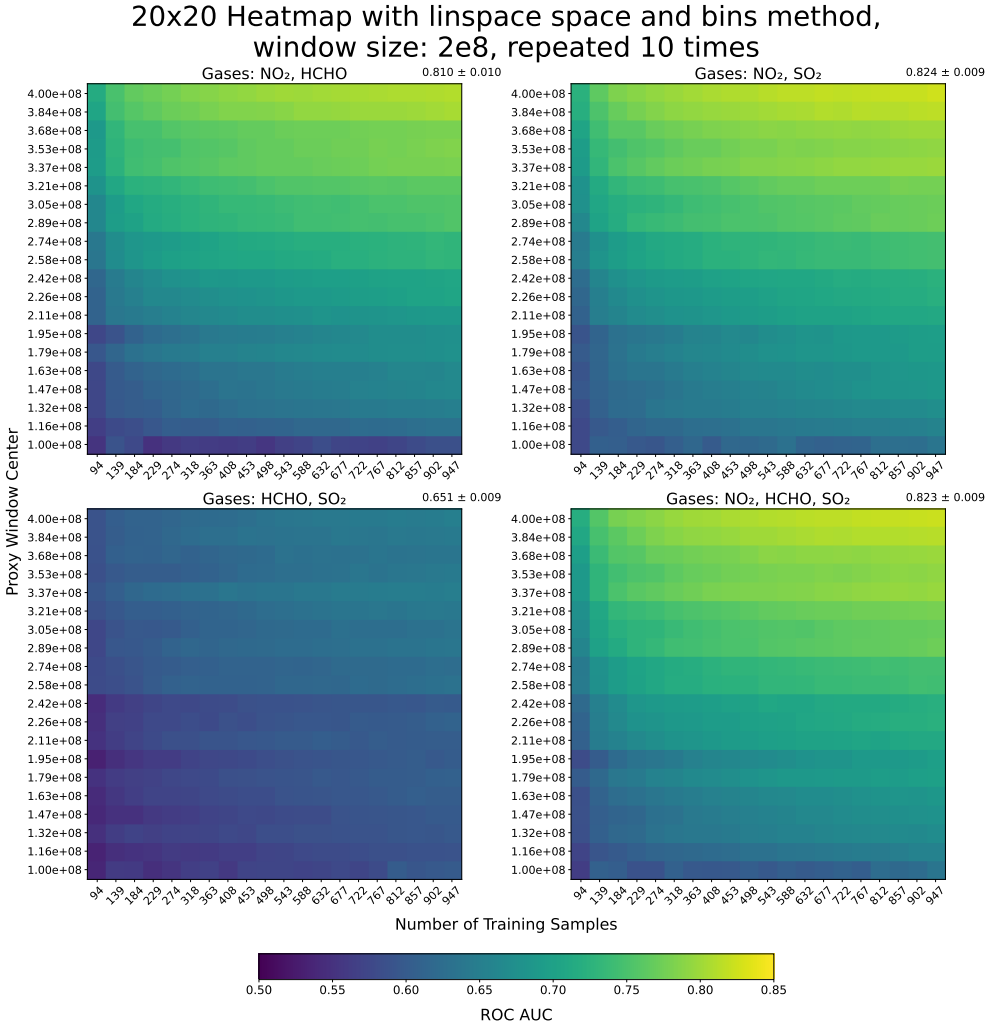

Figure 5.2: Heatmaps for {NO_2, HCHO}, {NO_2, SO_2 }, {HCHO, SO_2 }, and {NO_2, SO_2, HCHO}. The heatmaps plot NO_x proxy versus data sets size, where the cell with displayed NOx proxy (p) contains all samples with proxy $p - 1e8$ to $p + 1e8$. The dataset size is increased linearly. The final performance scores for the highest proxies are annotated as $ROCAUC \pm CI_{95\%}$.

The results show each combination of gases. It can be observed that the higher proxy bins achieve higher performance and learn slightly faster in terms of training samples. NO_2 is clearly the most important gas, as all experiments without NO_2 achieve significantly lower ROC AUC scores. Combining HCHO and SO_2 does improve performance, but detecting ships based on these two gases is still lower than with only NO_2. Using additional support features might improve performance, but NO_2 remains

the key feature.

If we want to judge the impact of adding HCHO and SO_2 alongside NO_2 to a machine learning model, the color heatmap is insufficient in providing a clear comparison. The underlying values of the rightmost column of each gas combination can be used in a two-dimensional plot. The number of training samples will be constant, but the proxy window center can be used on the x-axis. A 95% confidence interval describes the distribution of the experiment results.

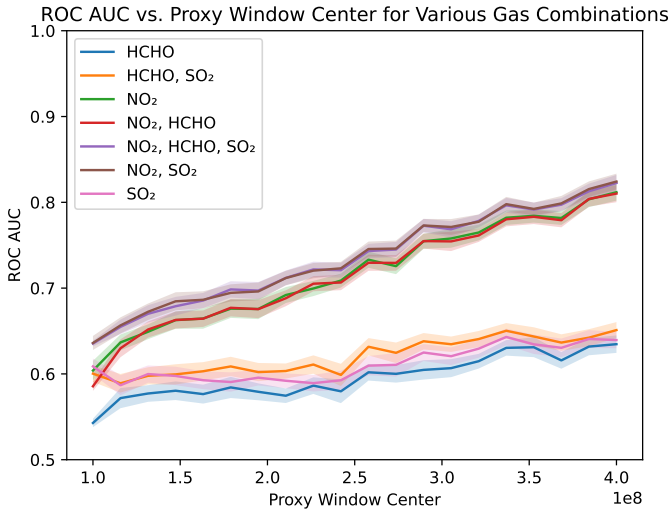

Figure 5.3: This figure shows the relationship between NO_x proxy and the detectability of ships. A window size of 2e8 is used.

A clear difference can be seen between the performance on data including NO_2 and the data consisting solely of HCHO and SO_2. SO_2, HCHO, and their combination do not increase at the same rate for higher NO_x proxies as data including NO_2 does. We again see improvement when adding SO_2 and HCHO. However, the overlapping confidence bounds indicate no significant gain. Likely caused by the limited amount of data (947 training samples, $\pm$473 per class), the results are not very stable.

5.2 Learning curves

In this section, we present learning curves as defined in section 5.2, computing performance for an increasing training set size.

5.2.1 With proxy binning

In this experiment, we show learning curves for five non-overlapping proxy bins. Each proxy bin can be seen as a data subset. Higher proxy bins have more data available (see figure 4.5), but we don't constrain the size of the subsets to the size of the smallest subset. This way, we can see for each proxy window what the benefit of collecting more data can be.

For this experiment, we create four datasets: one with NO_2 data plus all support features, one with SO_2 data plus all support features, one with HCHO data plus all support features, and one with NO_2, HCHO, and SO_2 data plus all support features; for all datasets, we create five subsets of different proxy bins, with a window size of 1e8. We have a bin from 0 to 1e8 (center 0.5e8), one from 1.12e8 to 2.12e8 (center 1.62e8), one from 2.25e8 to 3.25e8 (center 2.75), 3.38e8 to 4.38e8 (center 3.88e8) and one from 4.5e8 to 5.5e8.

As described in section 4.3.3, we create 15 outer folds. In the process of creating one fold, we start with all the 1 (WSST(s)) data available in the respective proxy bin, these are n samples. We then sample n 0 (no WSST) rows, resulting in a balanced dataset. We then use GSS to divide this subset into a train and test set.

Without HPO

We created the learning curves for 15 folds of GSS with the default HPs of the `xgboost-python` package. We present the results of this experiment in figure 5.4 where we plot the learning curves, normalized over 15 runs, with 95% confidence bounds.

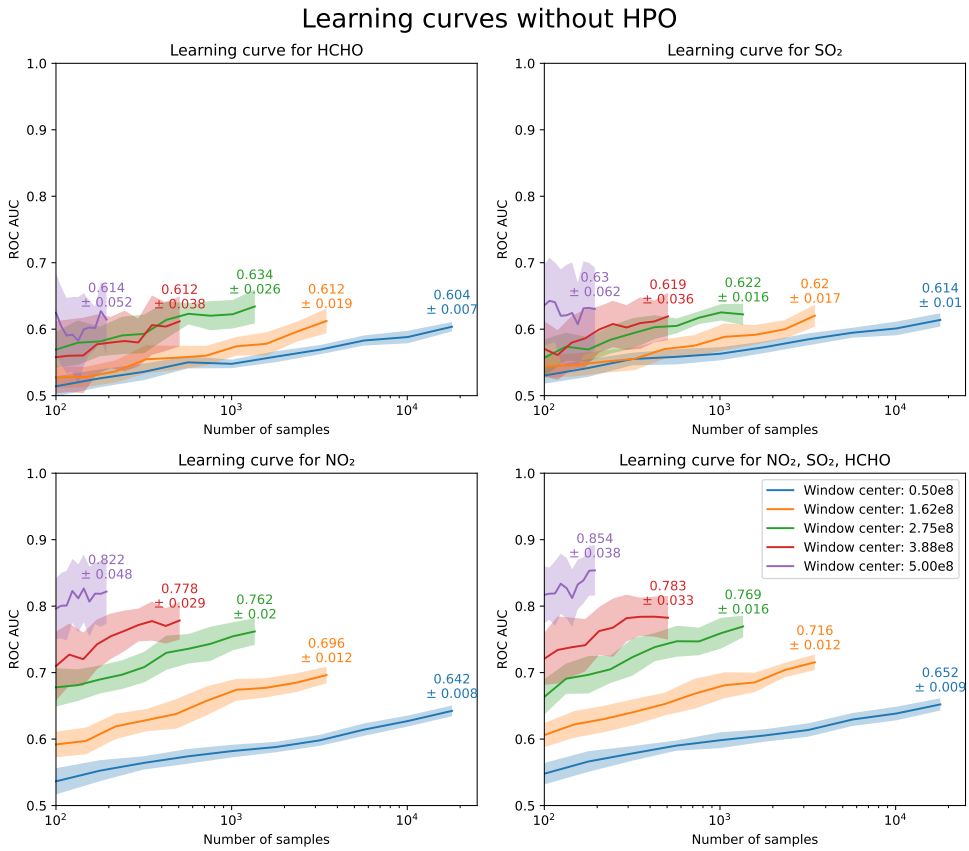

Figure 5.4: Learning curves of ship plume detection with XGBoost with default HPs. Different proxy windows are compared at the availability of NO_2 or NO_2, SO_2, and HCHO data. A 95% confidence interval is displayed around the lines. The window size is 1e8. The final performance scores are annotated as $ROCAUC \pm CI_{95\%}$.

Looking at the results without HPO, we observe a steadily rising learning process. For the experiments done with NO_2 data, we see an increase in performance at the higher proxy bins.

With HPO

We repeat this experiment with HPO to see the effects of HPO. After finding the best HP configuration for every outer fold, we use those parameters to make learning curves using the (train+validation)/test split from the outer fold. For the learning curves, we consider the train+validation set the training set.

To create the learning curves, we train a model on an exponentially increasing number of train samples. Starting at 100 and ending at the maximum number of

available data points. Since with GSS, the training set size can vary across folds, we constricted the maximum number of samples to the smallest outer fold (note this is only about discarding 1-10 samples). This gives us learning curves for each of the five proxy bins, where the higher proxy bins have early-stopping learning curves as they have less data available.

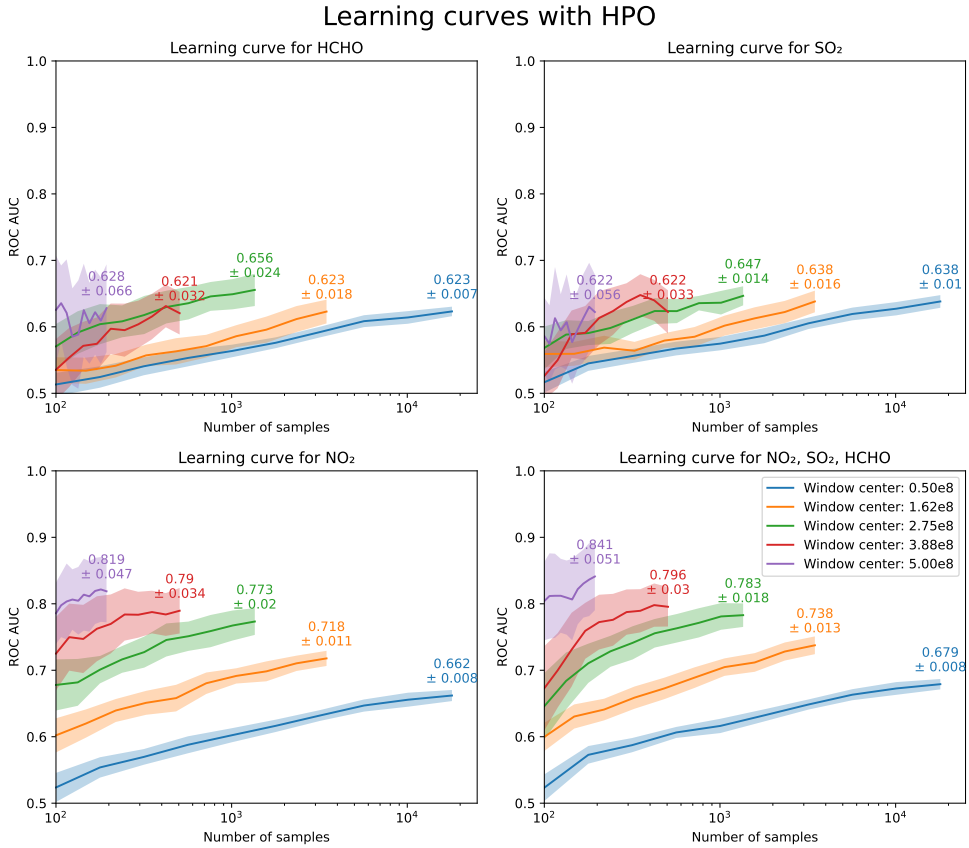

Figure 5.5: Learning curves of ship plume detection with XGBoost with optimized HPs using BS with nested cross-validation. Different proxy windows are compared at the availability of NO_2 or NO_2, SO_2, and HCHO data. A 95% confidence interval is displayed around the lines. The window size is 1e8. The final performance scores are annotated as $ROC\,AUC \pm CI_{95\%}$.

From Figure 5.5, it can be observed that the higher proxy windows have higher performance scores. Since the proxy value is an estimation of the amount of NO_x emitted by ships in an area, it makes sense that the performance does not increase in higher proxy bins for SO_2 and HCHO with the same magnitude that NO_2 performance does. This could indicate that they are correlated, but that there is simply less SO_2

and HCHO emission from ships. For the lower proxy bins, we see improvement in performance by adding HCHO and SO_2 to NO_2 data; for the higher proxy bins, we also see an increase. For the higher bins, there is less data available and the confidence interval is wider. Resulting in a less clear improvement.

The stagnation in some learning curves could be due to using aggregate features. Using aggregate features in a tabular database allows us to use the XGB algorithm. However, it does discard spatial relations within the data. Due to oversampling in the creation process of the tabular database, we also have many similar data points.

Complex data is harder for a machine learning algorithm to learn. Assuming the inclusion of SO_2 and HCHO data makes the training data more complex, this phenomenon can be observed when comparing the learning curves of NO_2 with those using {NO_2, SO_2, HCHO}. The curves of most proxy bins for {NO_2, SO_2, HCHO} start lower and climb faster in between 100 and 1000 samples than the curves for NO_2.

Influence of HPO

For most proxy windows, we see an improvement when using HPO. We see performance start out lower for the HPO learning curves. While the HPO curves do reach a higher ROC AUC score for most proxy windows, for the highest window with center 5e8, the ROC AUC without HPO is higher than with HPO. We do again see an increase in performance when adding SO_2 and HCHO data to the NO_2 data. With HPO, not only does the performance improve, but also the performance gain increases when adding HCHO and SO_2 data. With and without HPO, the performance increases more slowly for higher proxy bins with only HCHO and SO_2 data.

5.2.2 With all ships

In this experiment, we compare the performance of all combinations of gases. We do not use the NO_x proxy value in this experiment. So, for each gas combination, we will use all available data to create learning curves with nested cross-validation. We performed nested cross-validation as described in section 4.3.3. We plot the learning curve of all combinations of gases. The performance is plotted in ROC AUC. Additionally, the 95% confidence bounds are overlaid on the results.

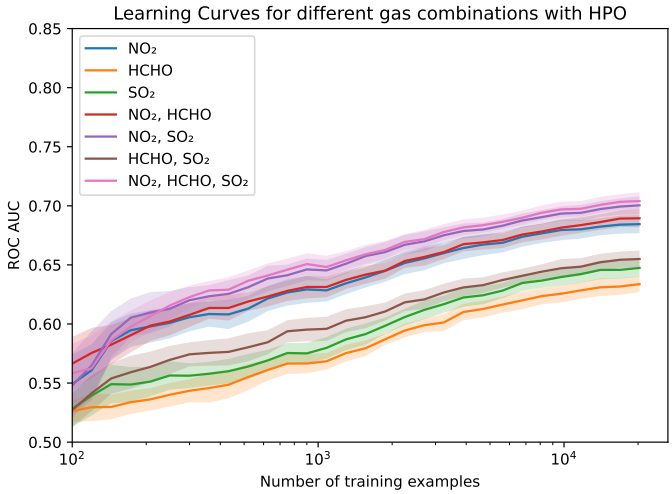

Figure 5.6: Comparison between different gases for predicting ship presence without for all ships (without proxy binning). A 95% confidence bound is plotted around the lines.

We can observe all models have learning abilities. HCHO has the lowest scores. SO_2 is slightly better, like the combination of HCHO and SO_2. It is apparent that NO_2 is a good indicator of ship presence. All combinations involving NO_2 perform reasonably. While we do see an improvement when SO_2 or HCHO are included, the best score is with all gases. This tells us the value of the SO_2 and HCHO data. We see a slight overlap in the confidence bounds of NO_2 and $\{NO_2, SO_2, HCHO\}$ in the figure; in table 5.1, we present a tabular view of the final values of the learning curve along with the 95% confidence interval. Using the table, we can compare the final scores more precisely.

Influence of HPO

In figure 5.6, we compared the learning curves of different gas combinations with an algorithm with optimized HPs. To see the influence of HPO in this experiment, as well as a more precise overview of the results, we present an overview of the performance on all available data with and without HPO for every gas combination in ROC AUC. Because confidence intervals were already shown in Figure 5.6, we present the 95% confidence interval between experiments in addition to the mean ROC AUC.

HPO Combination	No	Yes
NO_2	0.661 ± 0.009	0.684 ± 0.011
HCHO	0.613 ± 0.011	0.634 ± 0.015
SO_2	0.626 ± 0.013	0.647 ± 0.015
HCHO, SO_2	0.630 ± 0.012	0.655 ± 0.017
NO_2, HCHO	0.664 ± 0.011	0.689 ± 0.014
NO_2, SO_2	0.674 ± 0.011	0.700 ± 0.012
NO_2, HCHO, SO_2	0.677 ± 0.011	**0.704 ± 0.013**

Table 5.1: Overview of the final values of the learning curve. An ablation study of HPO. Highest performance is highlighted. Values shown as $ROCAUC \pm CI_{95\%}$.

Table 5.1 shows a ROC AUC improvement from 0.684 to 0.704 when comparing NO_2 to NO_2, SO_2, HCHO. The 95% confidence intervals of these measurements (0.684 ± 0.011 and 0.704 ± 0.013 respectively) have a small overlap: the upper bound of NO_2 (0.695) is higher than the lower bound of NO_2, SO_2, HCHO (0.691). This overlap means that while the combination of gases shows a higher mean performance, we cannot be 95% confident that this improvement would consistently occur, as there is a small possibility that NO_2 alone could perform similarly to the combination of gases in some cases.

This chapter showed the results of the study, in the next chapter, we will analyze the results w.r.t. the research questions.

Chapter 6

Discussion

In this chapter, we answer the research question and analyze whether we could answer the research question sufficiently based on the results. Additionally, we compare the results with findings from other studies.

6.1 Research questions

Recall the research questions defined in section 1.1. To answer the research questions, we used WSSTs to estimate the location of the ship plumes. We then predicted whether a sample patch had a WSST present in the inner area to perform ship plume detection.

RQ 1: How does the inclusion of SO_2 and HCHO data, alongside NO_2 data, improve the detection accuracy of ship plumes using data from a single TROPOMI orbit?

We observe a performance gain w.r.t. ship plume detectability when adding HCHO and SO_2 to NO_2 data across all experiments. Given that HCHO and SO_2 are gases that are emitted by ships, a slight gain was expected.

With HPO, the gain for adding HCHO and SO_2 to NO_2 data increases. Adding these features makes the data more complex for the machine learning algorithm, causing a slower-rising learning curve. But because it holds valuable information, the final performance is higher.

RQ 2: Can ship plumes be detected based on SO_2 or HCHO data from a single TROPOMI orbit?

For HCHO and SO_2 by themselves, the performance is worse than NO_2 but better than random. There is a moderate rise in performance for higher proxy bins.

6.2 Comparing with other studies

Results are similar to Kurchaba et al. (2024) [28] but are difficult to compare due to differences in methodology, where we made use of proxy binning. Kurchaba et al. (2024) do not mention segregating training/testing samples based on date but use Stratified K-Fold. This segregation is more important in this study as we deal with overlapping sample patches.

The added value of SO_2 and HCHO alongside NO_2 could be due to most ships not directly emitting NO_2. Most of a ship's NO_x emissions are in the form of NO [41], which is then converted in NO_2, peaking 15-30km downwind of the ship. This weakens the correlation between NO_2 data and the WSST. TROPOMI does not measure NO or NO_x, if available in the future, we expect NO observations to be a great addition to NO_2 data for ship plume detection. HCHO and SO_2 could expose different parts of the ship plume, either more at the head or tail of the plume. This could be due to a different half-life of the gases. In previous work, shipping lanes have been observed in three-month averages of TROPOMI data[6]. Although individual ship plumes have not been observed, it proves HCHO to be a ship emission gas that TROPOMI can measure.

6.3 Analysis

NO_2 and Ozone are the only TROPOMI data products that are available in tropospheric and stratospheric column densities. HCHO and SO_2 only have total column densities, as there is no algorithm developed for differentiating between these concentrations. This makes the HCHO and SO_2 data more noisy than the NO_2 data. We would expect an increase in detectability if tropospheric column densities were available for SO_2 and HCHO. Additionally, the TROPOMI sensor could have a lower sensitivity to SO_2 and HCHO.

Since it is assumed individual ships are not visible on SO_2 or HCHO data, the better than random performance based on only those gases and support features is novel. There has not been an experiment with the performance on samples containing one ship, and the support features might play a big role, so we cannot conclude that we can detect individual ships using only SO_2 and HCHO data. We also observe that the length and speed of a ship (NO_x proxy) does have an impact on the detectability, but this relation is not as strong as for NO_2 data. This could indicate that they are correlated, but that there is simply less SO_2 and HCHO emission from ships. The retrieval of HCHO slant columns [21, 7] is difficult due to its low optical density relative to other atmospheric gases like NO_2 and SO_2, which makes detection more sensitive to noise and spectral interference from absorbers like O_4 and O_3.

Because the classification ROC AUC improves for higher NO_x proxies, the NO_x proxy formula seems to give a good approximation of NO_2 emission of a ship. Where the increase is less for SO_2 and HCHO, the NO_x is either inaccurate or does not capture the reason for the increase in these gas emissions.

The higher proxy bins have more unstable learning curves. The 95% confidence

interval is larger than at lower proxy bins. The higher bins might need more repetitions to get more robust results. This instability could also result from the inaccuracy of the proxy calculation. Although previous studies showed that the NO_x proxy is quite an accurate estimator for NO_2 emissions [18], missing values of ship length were interpolated, see section 4.2.2. This interpolation could have decreased the accuracy of the NO_x proxy. Additionally, being based on only length and speed, some accuracy has been sacrificed for versatility. Finally, the impact of a ship being loaded or not has not been considered. This has a great influence on the water resistance and mass of the ship and thus will increase emissions. A more accurate NO_x proxy could be made using ship TIER, GT, year of build alongside length and speed to improve the NO_x proxy, while still being able to calculate it based on AIS data and ship characteristics.

6.4 Limitations

In the calculation of the WSSTs, we assumed the length of one degree of latitude and longitude equals 111.1km. This simplification holds at the equator. At higher latitude values, however, the distance of a longitude degree becomes smaller, approaching zero at the north pole. A simple fix for this issue could be to use a constant longitude degree length based on the center latitude and longitude of the AOI. This approach reduces the accuracy of the WSST locations.

Additionally, using the wind data to both calculate the WSSTs (section 4.2.2), filter out unreliable TROPOMI data (section 4.2.3), and as a model feature (table 4.1) could be considered to introduce slight data leakage, causing optimistic performance scores.

Reliance on AIS data for validation increases noise in the ground truth. Because the AIS data has gaps, variable sample rate, and missing values when aligning the ship length. The wind data from the TROPOMI data products are also insufficient for this study. Ideally, one would utilize wind data from a higher height, as well as wind data from multiple time points starting from two hours before the TROPOMI overpass.

A limitation of this experiment is that we have not seen the effect of using only support features. We can not be certain that this score is mostly due to gas concentration or mostly the support features. The support features could indicate the location of the sample patch, allowing the model to learn plume presence based on location. Kurchaba et al. (2024) [28] saw support features having a strong influence on model decisions. An ablation study or SHAP values [33] could give more insight into the feature importance of SO_2 and HCHO and the support features.

The search space of HPO is quite large despite the logarithmic probability space for some hyperparameters and 30 iterations of tuning. It is possible that the optimal hyperparameter configuration has not been found. A smaller or categorical search space could be beneficial computation-wise, and the optimal configuration could be found in a smaller number of iterations.

When creating the tabular database, we randomly sampled patches. The random sampling makes part of the data go unused. A sliding window with a stride that is low enough to sample 50 patches or more per orbit file would satisfy all three repro-

ducibility requirements, generating many samples and using all available data.

Where the current chapter focuses on discussing the results based on the research questions, comparing results with other studies, and describing the limitations, in the next chapter we will give answers to the research questions and conclude how the findings advance the field.

Chapter 7

Conclusions

This thesis explored the impact of integrating SO_2 and HCHO data with NO_2 measurements from the TROPOMI instrument to improve ship plume detection accuracy.

We conducted several experiments that compared ship plume detection performance. Learning curves show the promise of acquiring more data and show the differences between the detection performance based on gas type and NO_x proxy bin. Below, we elaborate on the research questions and discuss how the results advance the field of compliance monitoring.

RQ1: How does the inclusion of SO_2 and HCHO data, alongside NO_2 data, improve the detection accuracy of ship plumes using data from a single TROPOMI orbit?

The study found that the inclusion of SO_2 and HCHO data alongside NO_2 improved detection performance. Specifically, with hyperparameter optimization, the ROC AUC increased from 0.684 ± 0.011 using NO_2 alone to 0.704 ± 0.013 when combining NO_2, SO_2, and HCHO data. We observed an improvement across all of our experiments. Also, across all NO_x proxy bins, we see this performance improvement.

RQ2: Can ship plumes be detected based on SO_2 or HCHO data from a single TROPOMI orbit?

SO_2 and HCHO by themselves perform better than random chance (with an ROC AUC of 0.626 and 0.613 respectively) and slowly improve in performance when testing on bigger and faster ships, e.g., a higher NO_x proxy bin. We also observed that this performance can be improved when using more TROPOMI and AIS data.

How it advances the field

This study explores what influences ship plume detectability. The knowledge gained can be used to improve models used in compliance monitoring, where regression estimates of emissions are compared with actual emissions to find anomalously emitting

ships. SO_2, HCHO, and potentially other gases could be useful features in those regression models.

The conclusions chapter has answered the research questions, showing performance improvement for interesting HCHO and SO_2 alongside NO_2 for ship plume detection. The next chapter will describe how this knowledge gained can be used and other directions for future research.

Chapter 8

Future Work

This chapter outlines several directions for future research that could build upon the findings of this study.

More repetitions

We need to run more repetitions of the experiments to get more robust results. Particularly at higher NO_x proxy bins, the results showed greater variability. By running more repetitions, we can smooth out potential anomalies.

Ablation studies

Support features like solar angles and weather data could reveal seasonality, resulting in slight data leakage. An ablation study could be performed to test the performance of each gas using only gas concentrations measured by TROPOMI.

IMO2020 and MARPOL Annex VI Regulation 14 can be used further to investigate the impact of SO_2 on ship plume detection. IMO2020 limits the sulfur content in fuels to 0.5%, while MARPOL Annex VI Regulation 14 will impose a limit of 0.1% going into effect in 2025. The ship plume detection performance in the periods pre-2020, 2020-2025, and post-2025 could give more insight into the importance of SO_2 in ship plume detection.

Aquiring or retaining more data

A simple improvement to this study could be to request more AIS data. Having more ground-truth data will lead to a much larger database, and from the learning curve results, we have seen that having an order of magnitude more data is guaranteed to get a higher performance.

When aggregating the data from each sampled patch image, much information is lost. Although previous works have aimed to create a computer vision model that predicts ship segmentation on these images, there is still room for improvement in

these models. This is a promising research direction, especially when doing a binary classification.

We only included orbit files that fully covered the AOI. This means that the intersection of the AOI and the orbit are equal to the AOI. We could have generated more data by being more lenient in this decision, but priority was given to creating a robust dataset and minimizing missing values. If the study aimed to maximize classification accuracy, lifting this constraint would be recommended. Additionally, getting the AOI from the AIS dataset limits Would ensure that one can create a more extensive dataset while retaining the ground truth data.

As mentioned in the discussion, using a sliding window with a stride of, for example, 1/8 would improve the tabular database by allowing for more samples and using all available data.

When TROPOMI data is insufficiently accurate, a land-based TROPOMI-like sensor could also be used to measure HCHO and SO_2 of ship plumes to collect more high-resolution data that can be annotated to make ground-truth labels.

Ship & Wind-Shifted ship tracks

The speed is not used to estimate the distance traveled between A and B, but it could be used to make this interpolation (section 4.2) more accurate in future work. This would increase the accuracy of the ship tracks.

For the WSST, we use wind speed at the time of the TROPOMI overpass, and we have yet to acquire a larger temporal window of wind speed data. With access to this data, a more accurate WSST could be calculated compared to the current method described in section 4.2.2.

Additionally, the wind data is from 10m high, way lower than the top of the funnel of most ships. If higher wind data becomes available, that would help a lot, especially when NO_2 measurements' resolution increases with TANGO. Accurate estimation of the plume location will be most important in automatic monitoring of emissions.

Final remarks

While this study focuses on gas emissions, it remains important to consider all types of pollution when designing new regulations. The sentinel satellites are used for many other environmental studies. Sentinel-2 can be used to monitor plastic soup, and Sentinel-6, designed to monitor the rising sea level, could also be used in many studies. The sentinel satellites have opened up a world of possibilities for earth observation. Having open-access data gives people all over the world the option of helping not only in the fight against climate change but also showing the world the seriousness of its effects in the present.

The results found in this study have shown the value of a multi-gas approach when it comes to ship plume detection. Using gases beyond NO_2 can improve compliance monitoring algorithms for seagoing ships. Improving air quality at open sea and in coastal regions.

Bibliography

[1] Karin Andersson, Francesco Baldi, Selma Brynolf, J. Fredrik Lindgren, Lena Granhag, and Erik Svensson. *Shipping and the Environment*, pages 3–27. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. doi:10.1007/978-3-662-49045-7_1.

[2] Ljubiša Babic, Remco Braak, Werner Dierssen, ´ Jonathan Kissi-Ameyaw, Quintus Kleipool, Jonatan Leloux, Erwin Loots, Antje Ludewig, Nico Rozemeijer, Joost Smeets, and Giuseppe Vacanti. Algorithm theoretical basis document for the tropomi l01b data processor, report s5p-knmi-l01b-0009-sd. Technical report, KNMI, De Bilt, The Netherlands, 2022.

[3] Evert A Bouman, Elizabeth Lindstad, Agathe I Rialland, and Anders H Strømman. State-of-the-art technologies, measures, and potential for reducing ghg emissions from shipping–a review. *Transportation Research Part D: Transport and Environment*, 52:408–421, 2017.

[4] Pavel Brazdil, Jan N Van Rijn, Carlos Soares, and Joaquin Vanschoren. *Metalearning: applications to automated machine learning and data mining*. Springer Nature, 2022.

[5] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. ACM, August 2016. URL: http://dx.doi.org/10.1145/2939672.2939785, doi:10.1145/2939672.2939785.

[6] I. De Smedt, G. Pinardi, C. Vigouroux, S. Compernolle, A. Bais, N. Benavent, F. Boersma, K.-L. Chan, S. Donner, K.-U. Eichmann, P. Hedelt, F. Hendrick, H. Irie, V. Kumar, J.-C. Lambert, B. Langerock, C. Lerot, C. Liu, D. Loyola, A. Piters, A. Richter, C. Rivera Cárdenas, F. Romahn, R. G. Ryan, V. Sinha, N. Theys, J. Vlietinck, T. Wagner, T. Wang, H. Yu, and M. Van Roozendael. Comparative assessment of tropomi and omi formaldehyde observations and validation against max-doas network column measurements. *Atmospheric Chemistry and Physics*, 21(16):12561–12593, 2021. URL: https://acp.copernicus.org/articles/21/12561/2021/, doi:10.5194/acp-21-12561-2021.

[7] I. De Smedt, N. Theys, H. Yu, T. Danckaert, C. Lerot, S. Compernolle, M. Van Roozendael, A. Richter, A. Hilboll, E. Peters, M. Pedergnana, D. Loyola, S. Beirle, T. Wagner, H. Eskes, J. van Geffen, K. F. Boersma, and P. Veefkind. Algorithm theoretical baseline for formaldehyde retrievals from s5p tropomi and from the qa4ecv project. *Atmospheric Measurement Techniques*, 11(4):2395–2426, 2018. URL: https://amt.copernicus.org/articles/11/2395/2018/, doi:10.5194/amt-11-2395-2018.

[8] European Space Agency (ESA). Copernicus dataspace ecosystem. URL: https://dataspace.copernicus.eu/.

[9] European Space Agency (ESA). Copernicus open access hub. URL: https://scihub.copernicus.eu/.

[10] European Space Agency (ESA), Aug 2019. URL: https://sentinels.copernicus.eu/web/sentinel/-/copernicus-sentinel-5p-increased-spatial-resolution-change-now-operational?redirect=%2Fweb%2Fsentinel%2Fmissions%2Fsentinel-5p%2Fnews%3Fp_p_id%3Dcom_liferay_asset_publisher_web_portlet_AssetPublisherPortlet_INSTANCE_y55Z%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26_com_liferay_asset_publisher_web_portlet_AssetPublisherPortlet_INSTANCE_y55Z_delta%3D20%26p_r_p_resetCur%3Dfalse%26_com_liferay_asset_publisher_web_portlet_AssetPublisherPortlet_INSTANCE_y55Z_cur%3D6.

[11] Henk Eskes, Jos van Geffen, Folkert Boersma, Kai-Uwe Eichmann, Arnoud Apituley, Mattia Pedergnana, Maarten Sneep, J. Pepijn Veefkind, and Diego Loyola. Sentinel-5 precursor/tropomi level 2 product user manual nitrogen dioxide, report s5p-knmi-l2-0021-ma. Technical report, KNMI, De Bilt, The Netherlands, September 2023.

[12] European Parliament and Council of the European Union. Regulation (eu) 2023/1805 of the european parliament and of the council of 13 september 2023 on the use of renewable and low-carbon fuels in maritime transport, and amending directive 2009/16/ec. http://data.europa.eu/eli/reg/2023/1805/oj, 2023. OJ L 234, 22.9.2023, p. 48–100, Document PE/26/2023/INIT.

[13] European Space Agency. Copernicus sentinel-5p (processed by esa), tropomi level 2 nitrogen dioxide total column products. version 02. european space agency, 2021. Title of the publication associated with this dataset: TROPOMI Level 2 Nitrogen Dioxide.

[14] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognit. Lett.*, 27(8):861–874, 2006. URL: https://doi.org/10.1016/j.patrec.2005.10.010, doi:10.1016/J.PATREC.2005.10.010.

[15] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189 – 1232, 2001. doi:10.1214/aos/1013203451.

[16] Jean-Philippe Gagnon, Martin Larivière-Bastien, Jacob Thibodeau, and Stephane Boubanga Tombet. Remote estimation of sulfur content in fuel from so 2 and co 2 quantification of ship exhaust plumes. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 7287–7290. IEEE, 2021.

[17] K. Garane, M.-E. Koukouli, T. Verhoelst, C. Lerot, K.-P. Heue, V. Fioletov, D. Balis, A. Bais, A. Bazureau, A. Dehn, F. Goutail, J. Granville, D. Griffin, D. Hubert, A. Keppens, J.-C. Lambert, D. Loyola, C. McLinden, A. Pazmino, J.-P. Pommereau, A. Redondas, F. Romahn, P. Valks, M. Van Roozendael, J. Xu, C. Zehner, C. Zerefos, and W. Zimmer. Tropomi/s5p total ozone column data: global ground-based validation and consistency with other satellite missions. *Atmospheric Measurement Techniques*, 12(10):5263–5287, 2019. URL: https://amt.copernicus.org/articles/12/5263/2019/, doi:10.5194/amt-12-5263-2019.

[18] Aristeidis K Georgoulias, K Folkert Boersma, Jasper van Vliet, Xiumei Zhang, Ronald van der A, Prodromos Zanis, and Jos de Laat. Detection of no2 pollution plumes from individual ships with the tropomi/s5p satellite sensor. *Environmental Research Letters*, 15(12):124037, December 2020. URL: https://dx.doi.org/10.1088/1748-9326/abc445, doi:10.1088/1748-9326/abc445.

[19] Peringe Grennfelt, Anna Engleryd, Martin Forsius, Øystein Hov, Henning Rodhe, and Ellis Cowling. Acid rain and air pollution: 50 years of progress in environmental science and policy. *Ambio*, 49(4):849–864, Apr 2020. doi:10.1007/s13280-019-01244-4.

[20] Tim Head, Manoj Kumar, Holger Nahrstaedt, Gilles Louppe, and Iaroslav Shcherbatyi. scikit-optimize/scikit-optimize, October 2021. doi:10.5281/zenodo.5565057.

[21] Andreas Hilboll, Andreas Richter, Folkard Wittrock, Randall Martin, Michael Barkley, Steven Compernolle, and Kai-Uwe Eichmann. S5p/tropomi hcho atbd. Technical Report S5P- BIRA-L2- ATBD-400F, BIRA-IASB, KNMI & DLR, June 2022. URL: https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-ATBD-HCHO-TROPOMI.pdf/db71e36a-8507-46b5-a7cc-9d67e7c53f70?t=1658313806426.

[22] Dongni Hou, Yihui Ge, Cuicui Chen, Qiang Tan, Renjie Chen, Yanjie Yang, Li Li, Jian Wang, Maosong Ye, Chun Li, Xia Meng, Haidong Kan, Jing Cai, and Yuanlin Song. Associations of long-term exposure to ambient fine particulate matter and nitrogen dioxide with lung function: A cross-sectional study in china. *Environment International*, 144:105977, 2020. URL: https://www.sciencedirect.com/science/article/pii/S0160412020319322, doi:10.1016/j.envint.2020.105977.

[23] International Maritime Organization (IMO). Fourth greenhouse gas study 2020. Technical report, International Maritime Organization (IMO), 2021.

[24] Amazon.com Inc. Amazon s3 - cloud object storage - aws. URL: https://aws.amazon.com/s3/.

[25] International Transport Forum. Reducing sulphur emissions from ships. Technical report, Organisation for Economic Co-Operation and Development (OECD), May 2016. doi:10.1787/5jlwvz8mqq9s-en.

[26] Debbie J Jarvis, Gary Adamkiewicz, Marie-Eve Heroux, Regula Rapp, and Frank J Kelly. Nitrogen dioxide. In *WHO guidelines for indoor air quality: Selected pollutants*. World Health Organization, 2010. URL: https://www.ncbi.nlm.nih.gov/books/NBK138707/.

[27] KNMI and TEMIS. Tropomi no2 data version overview, Nov 2024. URL: https://www.temis.nl/airpollution/no2col/tropomi_no2_data_versions.php.

[28] Solomiia Kurchaba, Artur Sokolovsky, Jasper van Vliet, Fons J. Verbeek, and Cor J. Veenman. Sensitivity analysis for the detection of no2 plumes from seagoing ships using tropomi data. *Remote Sensing of Environment*, 304:114041, 2024. URL: https://www.sciencedirect.com/science/article/pii/S003442572400052X, doi:10.1016/j.rse.2024.114041.

[29] Solomiia Kurchaba, Jasper van Vliet, Jacqueline J. Meulman, Fons J. Verbeek, and Cor J. Veenman. Improving evaluation of no2 emission from ships using spatial association on tropomi satellite data. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '21, page 454–457, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3474717.3484213.

[30] Solomiia Kurchaba, Jasper van Vliet, Fons J. Verbeek, Jacqueline J. Meulman, and Cor J. Veenman. Supervised segmentation of no2 plumes from individual ships using tropomi satellite data. *Remote Sensing*, 14(22), 2022. URL: https://www.mdpi.com/2072-4292/14/22/5809, doi:10.3390/rs14225809.

[31] Solomiia Kurchaba, Jasper van Vliet, Fons J. Verbeek, and Cor J. Veenman. Anomalous no2 emitting ship detection with tropomi satellite data and machine learning. *Remote Sensing of Environment*, 297:113761, 2023. URL: https://www.sciencedirect.com/science/article/pii/S0034425723003127, doi:10.1016/j.rse.2023.113761.

[32] Zhenyu Luo, Tingkun He, Wen Yi, Junchao Zhao, Zhining Zhang, Yongyue Wang, Huan Liu, and Kebin He. Advancing shipping NOx pollution estimation through a satellite-based approach. *PNAS Nexus*, 3(1):pgad430, 12 2023. arXiv:https://academic.oup.com/pnasnexus/article-pdf/3/1/pgad430/55381010/pgad430.pdf, doi:10.1093/pnasnexus/pgad430.

[33] Wilson Estécio Marcílio and Danilo Medeiros Eler. From explanations to feature selection: assessing SHAP values as feature selection mechanism. In *33rd*

SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2020, Recife/Porto de Galinhas, Brazil, November 7-10, 2020, pages 340–347. IEEE, 2020. doi:10.1109/SIBGRAPI51738.2020.00053.

[34] National Imagery and Mapping Agency. Department of defense world geodetic system 1984: its definition and relationships with local geodetic systems. Technical Report TR8350.2, National Imagery and Mapping Agency, St. Louis, MO, USA, January 2000. URL: http://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350_2.html.

[35] Leonidas Ntziachristos. Scipper, Jun 2023. URL: https://www.scipper-project.eu/.

[36] ITU-R (Radiocommunication Sector of ITU). Assignment and use of identities in the maritime mobile service. Technical Report ITU-R M.585-8, International Telecommunication Union (ITU), 10 2019. URL: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.585-8-201910-I!!PDF-E.pdf.

[37] International Maritime Organization. *IMO 2020*. International Maritime Organization, 2019. URL: https://imo-epublications.org/content/books/9789280117189, doi:10.62454/K666E.

[38] International Maritime Organization. International Convention for the Prevention of Pollution from Ships (MARPOL). https://www.imo.org/en/About/Conventions/Pages/International-Convention-for-the-Prevention-of-Pollution-from-Ships-(MARPOL).aspx, 2022.

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

[40] Russ Rew and Glenn Davis. Netcdf: an interface for scientific data access. *IEEE computer graphics and applications*, 10(4):76–82, 1990.

[41] Tobias Christoph Valentin Wern Riess, K Folkert Boersma, Aude Prummel, Bart JH van Stratum, Jos de Laat, and Jasper van Vliet. Estimating nox emission of individual ships from tropomi no2 plumes. *Available at SSRN 4858709*, 2024.

[42] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proc. IEEE*, 104(1):148–175, 2016. doi:10.1109/JPROC.2015.2494218.

[43] Martin Svanberg, Vendela Santén, Axel Hörteborn, Henrik Holm, and Christian Finnsgård. Ais in maritime research. *Marine Policy*, 106:103520, 2019. URL: https://www.sciencedirect.com/science/article/pii/S0308597X18309667, doi:10.1016/j.marpol.2019.103520.

[44] James A Swenberg, Benjamin C Moeller, Kun Lu, Julia E Rager, Rebecca C Fry, and Thomas B Starr. Formaldehyde carcinogenicity research: 30 years and counting for mode of action, epidemiology, and cancer risk assessment. *Toxicologic pathology*, 41(2):181–189, 2013.

[45] Johannes Teuchies, Tom J. S. Cox, Katrien Van Itterbeeck, Filip J. R. Meysman, and Ronny Blust. The impact of scrubber discharge on the water quality in estuaries and ports. *Environmental Sciences Europe*, 32(1):103, Jul 2020. doi:10.1186/s12302-020-00380-z.

[46] Johannes Teuchies, Tom JS Cox, Katrien Van Itterbeeck, Filip JR Meysman, and Ronny Blust. The impact of scrubber discharge on the water quality in estuaries and ports. *Environmental Sciences Europe*, 32:1–11, 2020.

[47] Nicolas Theys, Isabelle De Smedt, Christophe Lerot, Huan Yu, Michel Van Roozendael, and Pascal Hedelt. S5p/tropomi so_2 atbd. Technical Report S5P-BIRA-L2-ATBD-400E, BIRA-IASB & DLR, June 2022. URL: https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-ATBD-SO2-TROPOMI.pdf/c3930dc9-16f6-4403-9049-db768eca6ff9?t=1689935449763.

[48] Jean-Noël Thépaut, Dick Dee, Richard Engelen, and Bernard Pinty. The copernicus programme and its climate change service. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1591–1593, 2018. doi:10.1109/IGARSS.2018.8518067.

[49] L. G. Tilstra. Tropomi l3 product user manual directionally dependent surface lambertian-equivalent reflectivity. Technical report, KNMI, 2023.

[50] J. van Geffen, K. F. Boersma, H. Eskes, M. Sneep, M. ter Linden, M. Zara, and J. P. Veefkind. S5p tropomi no_2 slant column retrieval: method, stability, uncertainties and comparisons with omi. *Atmospheric Measurement Techniques*, 13(3):1315–1335, 2020. URL: https://amt.copernicus.org/articles/13/1315/2020/, doi:10.5194/amt-13-1315-2020.

[51] JHGM van Geffen, HJ Eskes, KF Boersma, JD Maasakkers, and JP Veefkind. Tropomi atbd of the total and tropospheric no2 data products. Technical Report S5P-KNMI-L2-0005-RP, KNMI, July 2019. URL: https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-NO2-data-products.pdf/7a4fdde7-516e-48e7-bf44-da60c62b1e4d?t=1658312240128.

[52] Jos Van Geffen, K Folkert Boersma, Henk Eskes, Maarten Sneep, Mark Ter Linden, Marina Zara, and J Pepijn Veefkind. S5p tropomi no 2 slant column retrieval: Method, stability, uncertainties and comparisons with omi. *Atmospheric Measurement Techniques*, 13(3):1315–1335, 2020.

[53] Ward Van Roy, Jean-Baptiste Merveille, Kobe Scheldeman, Annelore Van Nieuwenhove, Benjamin Van Roozendael, Ronny Schallier, and Frank Maes. The role of belgian airborne sniffer measurements in the marpol annex vi enforcement chain. *Atmosphere*, 14(4), 2023. URL: https://www.mdpi.com/2073-4433/14/4/623, doi:10.3390/atmos14040623.

[54] Ward Van Roy, Benjamin Van Roozendael, Laurence Vigin, Annelore Van Nieuwenhove, Kobe Scheldeman, Jean-Baptiste Merveille, Andreas Weigelt, Johan Mellqvist, Jasper Van Vliet, Danielle van Dinther, et al. International maritime regulation decreases sulfur dioxide but increases nitrogen oxide emissions in the north and baltic sea. *Communications Earth & Environment*, 4(1):391, 2023.

[55] G. C. M. Vinken, K. F. Boersma, A. van Donkelaar, and L. Zhang. Constraints on ship no_x emissions in europe using geos-chem and omi satellite no_2 observations. *Atmospheric Chemistry and Physics*, 14(3):1353–1369, 2014. URL: https://acp.copernicus.org/articles/14/1353/2014/, doi:10.5194/acp-14-1353-2014.

[56] Geert CM Vinken, K Folkert Boersma, Daniel J Jacob, and Ernst W Meijer. Accounting for non-linear chemistry of ship plumes in the geos-chem global chemistry transport model. *Atmospheric Chemistry and Physics*, 11(22):11707–11722, 2011.

[57] Tao Xue, Mingkun Tong, Meng Wang, Xinyue Yang, Yanying Wang, Huan Lin, Hengyi Liu, Jiajianghui Li, Conghong Huang, Xia Meng, Yixuan Zheng, Dan Tong, Jicheng Gong, Shiqiu Zhang, and Tong Zhu. Health impacts of long-term no2 exposure and inequalities among the chinese population from 2013 to 2020. *Environmental Science & Technology*, 57(13):5349–5357, 2023. PMID: 36959739. arXiv:https://doi.org/10.1021/acs.est.2c08022, doi:10.1021/acs.est.2c08022.

[58] Tianle Yuan, Fei Liu, Lok N Lamsal, and Hua Song. Detecting ship-produced no2 plumes and shipping routes in tropomi data with a deep learning model. *Authorea Preprints*, 2023. doi:10.22541/essoar.168771101.14987378/v1.

Acknowledgements

I would like to thank my supervisor Cor Veenman for being very dedicated and taking the time to discuss the project almost weekly. Jasper van Vliet for sharing and explaining the AIS data under NDA. I'd like to thank Fons Verbeek for grading my thesis.

I want to thank my family and my girlfriend for their support in completing the thesis.

Appendix A

Experimental Setup

A.1 Computer Specifications

Component	Specification
Processor	AMD Ryzen 5 5600X 6-Core Processor, 3701 MHz, 6 Cores, 12 Logical Processors
Installed Physical Memory (RAM)	64.0 GB
Graphics Card	NVIDIA GeForce RTX 4070

Table A.1: Computer Specifications

A.2 Python Packages

Name	Version
geopandas	1.0.1
harp	1.23
numpy	1.26.4
pandas	2.2.3
scikit-learn	1.5.2
scikit-optimize	0.10.2
scipy	1.14.1
shapely	2.0.6
xgboost	2.1.1

Table A.2: Python 3.11.10 Packages

Acronyms

AOI
Area Of Interest

ATBD
Algorithm Theoretical Basis Document

BS
Bayesian Search

CH_4
Methane

CO
Carbon Monoxide

DOAS
Differential Optical Absorption Spectroscopy

ECA
Emission Control Area

ECMWF
European Centre for Medium-Range Weather Forecast

FPR
False Positive Rate

GPS
Global Positioning System

GRIB
General Regularly distributed Information in Binary form

GSS

GroupShuffleSplit

GT

Gross Tonnage

ILT

Inspectie Leefomgeving en Transport

IMO

International Maritime Organization

ITU

International Telecommunication Union

MMSI

Maritime Mobile Service Identity

NRTI

Near-Real-TIme

O_3

Ozone

OFFL

OFFLine

QA

Quality Assurance

S3

Simple Storage Service

SCD

Slant Column Density

SCIPPER

Shipping Contributions to Inland Pollution Push for the Enforcement of Regulations

SDB

Shipping DataBase

SGKF

Stratified Group K-Fold

sklearn
Scikit-learn

TANGO
Twin ANthropogenic Greenhouse gas Observers

TM5
Tracer Model version 5

TPR
True Positive Rate

UTZ
Universal Time Zone

VCD
Vertical Column Density

WGS84
World Geodetic System 1984

WHO
World Health Organization

WSST
Wind-Shifted Ship Track