Machine learning-based NO2 estimation from seagoing ships using TROPOMI/S5P satellite data
Kurchaba, S.

Machine learning-based NO$_2$ estimation from seagoing ships using TROPOMI/S5P satellite data

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op dinsdag 11 juni 2024
klokke 10:00 uur

door

Solomiia Kurchaba
geboren te Lviv, Oekraïne
in 1995

Promotores:
Prof.dr.ir. F.J. Verbeek
Prof.dr. J.J. Meulman

Co-promotor:
Dr. C.J. Veenman

Promotiecomissie:
Prof.dr. I. Aben (SRON Netherlands Institute for Space Research)
Prof.dr. T.H.W. Bäck
Dr. K.F. Boersma (Wageningen University and Research)
Prof.dr. M.M. Bonsangue
Prof.dr.ir. N. Mentens

Human Environment and Transport
Inspectorate
*Ministry of Infrastructure
and Water Management*

To my grandpa.

"… We must have perseverance and above all confidence in ourselves. We must believe that we are gifted for something and that this thing must be attained …"

Maria Salomea Skłodowska-Curie

Contents

Contents

Chapter 1

Introduction

Air pollution ranks among the most pressing challenges in our society. Decades of research have proven a strong adversarial effect of air pollution on human health, the vitality of ecosystems, the state of the atmosphere, and climate change [16, 54, 83, 84]. At the same time, according to the World Health Organization (WHO), 9 out of 10 people currently breathe polluted air. Moreover, due to the continuous urbanization [98], industrialization, and economic development, the number of potential sources of air pollution is dramatically increasing [3, 87, 83, 111].

One of the most harmful components of air pollution are nitrogen oxides gases ($NO_x \approx NO + NO_2$). These gases play an important role in the destruction of the atmospheric ozone [25]. In addition, anthropogenic NO_x is known to be one of the main precursors of photochemical smog [46, 82], whose harmful effects include aggravation of asthmatic attacks, irritation of eyes and throats of humans and animals, reduction of visibility, damage of the structure of plants and materials [100, 66]. Anthropogenic sources of NO_x include industrial emissions, biomass burning, and emissions from vehicle transport. One of the strongest sources of anthropogenic emission of NO_x is the industry of international shipping. The NO_x is produced in a ship engine through the combustion process, where nitrogen in the air reacts with oxygen, forming nitrogen oxides, primarily in the form of nitric oxide (NO). Subsequently, atmospheric conditions and chemical reactions transform NO into nitrogen dioxide (NO_2), a more reactive and harmful component of NO_x emissions. The global contribution of the shipping industry to the emissions of NO_x is estimated to vary between $15\% - 35\%$ [24, 52], causing approximately 60,000 premature deaths annually [23]. For the Netherlands, the contribution of the shipping industry is estimated to be around 10% [48]. While
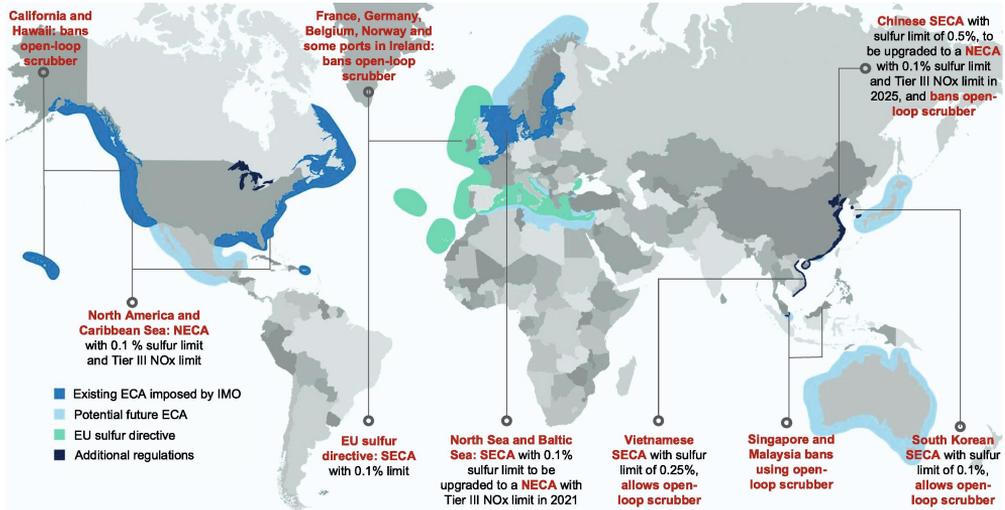
Figure 1.1: Map of ECA's restrictions. Source: [113].

over the last 20 years, the pollution produced by power plants, the industry sector, and cars has been constantly decreasing, the impact of maritime transport continues to grow [12]. This causes a big societal pressure, which calls for a collective effort for efficient regulation and monitoring of emissions from ships towards reducing the negative impact of the industry.

1.1 Ship emission monitoring

In 1997, aiming at the reduction of the negative impact on human health, the International Maritime Organisation (IMO) amended Annex VI to the International Convention for the Prevention of Pollution from Ships (MARPOL). This annex sets standards on sulfur dioxide and nitrogen oxides emissions from ship exhausts [50]. The amendments include the installation of emission control areas (ECAs) within which the emission constraints for ships operating in these areas are established and then tightened step-by-step. The map of currently established and considered ECAs is depicted in Figure 1.1. Within ECA regions, we distinguish nitrogen and sulfur emission control areas (NECA and SECA respectively). The latest step that was turned into force as a part of IMO directives is an 80% reduction of NO_x emission for diesel engines of newly-built ships operating in the Baltic and North Sea [51]. Compliance with these regulations requires shipowners and operators to invest in cleaner, more ex-

pensive technologies (e.g. installation of a selective catalytic reduction (SCR) system, which converts harmful gases into inert nitrogen and water vapor). The responsibility for the enforcement of IMO regulations is shared between the country where a given ship is registered and the authorities of the port where the ship operates. In the Netherlands, this is the Human Environment and Transport Inspectorate (ILT). Given the legislation, the responsibilities of the inspectorate are as follows: 1) Monitoring of emissions coming from ships to assess the effects of the legislation; 2) Verification of compliance of individual ships. The performance of neither of the above-mentioned is possible without efficient measurements of real-world emissions. Hence the support of ILT of the research presented here.

However, monitoring of ship emissions on a large scale is a challenging task. For instance, the methods currently used by port state authorities are checks on engine room logs and bunker delivery notes, or chemical analysis of fuel samples. Such practices, however, can be applied to only a limited number of ships. Other applied methods are on-board measurements at exhaust pipes [4], land- or ship-based downwind measurements using sniffer techniques [66, 81], and the DOAS (differential optical absorption spectroscopy) approach [73, 88, 59]. Alternatively, ship plume measurements are performed from airborne platforms like helicopters, small aircraft, and drones [102, 103]. Mobile platforms often measure pollutant ratios during plume transects [7] or use the DOAS technique for remote optical sensing [9]. All these methods require proximity to the ships under surveillance, are applied sporadically, and are too costly for monitoring the entire shipping fleet. Moreover, since such measurement stations are usually located at the entrance of the ports, the data collected with such methods provide limited information on how much the selected ships emit outside ports. As a result, there is currently no effective method for comprehensive and cost-efficient large-scale ship emission monitoring.

1.2   Satellite observations

A potential solution efficient for ship emission monitoring on a global scale is the application of satellite observations [90]. For more than a decade scientists have been using the available satellite data to quantify the NO_x emission produced by the shipping industry. For instance, using the measurements from the Global Ozone Monitoring Experiment (GOME) [17] instrument onboard the second European Remote Sensing satellite (ERS-2), the authors estimated the NO_2 emission levels above the shipping lane between Sri Lanka and Indonesia [8]. With the images from the SCanning Imaging
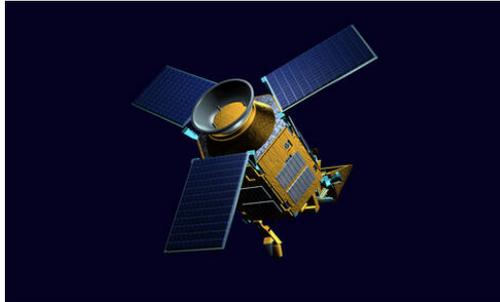
Figure 1.2: Sentinel-5 Precursor satellite. Credit image ESA 2017.

Absorption spectroMeter for Atmospheric CartograpHY (SCIAMACHY) [13] onboard the ENVIronmental SATellite (Envisat) mission, traces from the shipping industry over the Red Sea were quantified [85]. Finally, data from the Ozone Monitoring Instrument (OMI) [69] aboard the NASA Aura spacecraft was used to visualize the NO_x emission inventory of shipping in the Baltic Sea [106]. The obtained results were further associated with the temporal patterns of global economic activity [26, 12]. Nevertheless, all the above-mentioned studies were based on multi-month data averaging, which was necessary to perform in order to reduce the signal-to-noise ratio of the satellite measurements and enable distinguishability of NO_x traces along the shipping lanes. The low spatial resolution of the satellites did not allow for the distinction of ship plumes from individual ships on a daily basis.

The game changer in high-resolution atmospheric measurements is the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Copernicus Sentinel-5 Precursor (TROPOMI/S5P) satellite (illustration: Figure 1.2). Launched in 2017, the TROPOMI/S5P creates daily global maps of atmospheric substances relevant to air quality and climate monitoring [1]. More importantly, the instrument has a significantly higher spatial resolution than all its predecessors (GOME: 40×320 km^2, SCIAMACHY: 30×60 km^2, OMI: 13×25 km^2, TROPOMI: 3.5×5.5 km^2). The TROPOMI instrument measures an extensive list of trace gases, including NO_2. Since the NO_2 gas is the product of photo-chemical reactions of NO_x emitted by ships, it can be utilized for ship emission monitoring. As reported in [41], the spatial resolution of the TROPOMI instrument is high enough to distinguish some of the NO_2 plumes produced by individual ships. This study, however, focused on the largest ships in the area, as the NO_2 traces of most of the ships seemed not sufficiently stronger than the background concentrations. In addition, the presented approach involves multiple manual steps, which prevents its application on a large scale [41].

1.3 Machine learning

In order to increase the sensitivity of ship plume detection, in this thesis, we propose to address the problem with machine learning techniques that have proven very valuable in many domains. Machine learning is a computational paradigm that enables the automatic extraction of complex patterns and relationships in data, not only significantly reducing the human effort required, but also facilitating finding patterns that are otherwise unnoticeable to a human eye. A general definition of machine learning was proposed by Tom Mitchell in 1997 [74]. It goes as follows:

Definition 1.1. Machine learning is the study of computer algorithms that improve automatically through experience. An algorithm is said to learn from experience E with respect to task T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

Today, machine learning algorithms have demonstrated their efficiency in various fields of everyday life and science. The list of application domains that were revolutionized by machine learning includes health care (disease diagnosis [6], drug discovery [99]), finance (algorithmic trading [79], fraud detection [71]), computer vision (facial recognition [108], autonomous vehicles [65]), education (adaptive learning systems [56], automated feedback [28]), space exploration (spacecraft control and navigation [95], data processing for remote sensing missions [112]), and many more. In the domain of Earth observation, the list of tasks to which machine learning algorithms have made ground-breaking contributions includes (non-exhaustively) land cover classification, identification of crop diseases, algorithms for the optimization of the retrieval of satellite measurements, flood prediction, and optimization of computer code performance [92, 97, 68, 15, 91].

Different applications and tasks require different types of machine learning algorithms, such as supervised, semi-supervised, unsupervised, or reinforcement learning. In the domain of Earth observation, one of the most often used types of machine learning algorithm is supervised learning [37]. In supervised learning, we aim to learn a function to predict the output Y for a feature vector X. The learning process uses pairs of feature vectors and the corresponding outputs that are given as a training set (the Experience). Depending on the type of output variable, the supervised learning task can further be split into classification (categorical output variable) and regression (continuous output variable).

Emerging studies show the potential of supervised learning techniques for the analysis and information extraction from TROPOMI data. For instance, researchers ap-

plied a multivariate regression model to estimate the NO_2 emission rate over Germany [20], and O_3 concentrations in California [110]. Furthermore, classification models were used to automatically detect images containing NO_2 [38] or CH_4 [89] plumes from super-emitters, scanning the TROPOMI data around the globe. Analyzing the above-mentioned studies, we see two ways of representing TROPOMI measurements to a machine-learning algorithm, depending on the problem addressed. That is, in terms of a two-dimensional grid (an image), or a set of one-dimensional data features, calculated based on the measurement values for a specific area of interest. The former enables the application of techniques originating from the fields of computer vision or image processing (i.e. kernel-based filters, convolution neural networks, etc.), while the latter is more suited for the usage of multivariate techniques, combining the TROPOMI measurements with other data sources.

In this thesis, we explore the possibilities of estimation of the NO_2 emissions from individual ships using TROPOMI data. The emissions produced by a ship, if strong enough, will be registered by a TROPOMI sensor as an image of a plume. However, to estimate emissions produced by a certain individual ship, the information contained in the TROPOMI measurement is not sufficient. Other pieces of necessary information are the position of the ship, the speed, and dimensions of the ship, and the direction and speed of the wind. To efficiently exploit all the necessary sources of data, we will mostly focus on the application of multivariate supervised machine learning, while the spatial characteristics of the data will be utilized for image enhancement.

1.4 Research questions

The objective of this thesis is to pave the way toward the application of the TROPOMI instrument data for the monitoring of ship compliance with the regulations of IMO. The overarching research question addressed in this thesis can be formulated as follows:

Is it possible to use TROPOMI/S5P instrument data to monitor NO_2 emissions from individual seagoing ships?

We address this overall question step by step by answering the following list of intermediate research questions:

RQ1: *What is the minimum speed and length of a seagoing ship so that the NO_2 plume from it can be detected with a detection system using TROPOMI data?*
To understand the potential of the TROPOMI instrument for ship emission monitoring, it is crucial to estimate the required strength of the emitter (in this case, a

ship) for the detection of its NO_2 plumes. With the detection system, we refer to a sequence of steps needed for the automatic identification of NO_2 plumes from a ship on a TROPOMI image patch. The first step of this sequence is a measurement performed by the TROPOMI sensor. The last step is an automated detection of a plume on an image patch using machine learning models. We propose to estimate the detection capabilities of the TROPOMI data-based detection system using parameters such as speed and length of the ship, known to be reliable indicators of ship emission potential.

RQ2: *To what extent can the detectability of NO_2 plumes be improved if only the biggest emitters are taken into account?*
It is not possible to monitor all ships with a detection system based on TROPOMI data – there is a system sensitivity limit. There also will be a set of ships for which detection is possible, although difficult. Finally, there will be a set of the biggest emitters, from which the plumes are clearly the easiest to detect. Another example of the biggest emitters is when several ships are sailing in proximity to each other. To establish the baseline for the current possibilities of the application of a detection system using the data from the TROPOMI instrument for ship emission monitoring, the potential quality of detection of plumes produced by those biggest emitters should be evaluated separately.

RQ3: *Is there a potential for improvement of detectability of NO_2 plumes from the slow/small ships if more data were used to train the used classification model?*
Since the application of machine learning is an important part of the studied detection system of ship NO_2 plumes, the factor of data availability plays an important role in establishing the sensitivity limits of this system. The noisier the pattern we would like to detect, the more data are required for the training of a machine learning model. This will be especially relevant for ships that are just above the sensitivity limit of the detection system. Therefore, we would like to understand to which extent the addition of training data can help with the detection of the noisiest patterns.

RQ4: *How to assign a TROPOMI signal associated with a certain plume to a potential emitting ship?*
A characteristic feature of a ship as an emitter is the fact that it moves continuously. In addition to the movement of a ship, the plume emitted by it at a certain moment will gradually move in accordance with the direction and speed of the wind. These factors make the process of association of the detected plume with a ship emitter a non-trivial task. The task, however, is a necessary step in order to be able to use the TROPOMI data for the performance of the monitoring of emissions from individual ships.

RQ5: *To what extent can the NO_2 plumes be segmented in the TROPOMI data using a simple thresholding method?*

Segmentation of NO_2 plumes from individual ships using data from satellite-based sensors has not been performed before (because of the too-low spatial resolution of the previous satellite-based instruments). Therefore, to set up a proper baseline for a given task, it is reasonable to start with the application of the simplest potentially suitable approach. The thresholding approach can be considered as a good starting point due to the following reasons: ship plumes in a simplified setting can be considered as a blob of pixels with a concentration higher than the surrounding environment; the thresholding method does not require human labeling (unsupervised learning) and could be directly applied on the data.

RQ6: *Can we improve the segmentation quality of NO_2 plumes from individual ships using supervised machine learning?*

Once the simplest baseline is established, we would like to understand how the quality of ship plume segmentation can be improved once a more complex methodology is applied. With supervised machine learning, we provide the model with the human labels of the position of the NO_2 plume of interest. With this, the model could pick up the nonlinear dependencies that differentiate a pixel that belongs to a plume from a pixel that is part of the background.

RQ7: *Does the machine learning-based segmentation allow for the detection of NO_2 plumes that cannot be recognized visually?*

The fact that some of the NO_2 plumes cannot be recognized when visually studying the data, does not mean that the signal has not been registered by TROPOMI. Among other reasons, there can be an insufficiently detailed color scheme selected when visualizing the data, or insufficient capabilities of the human eye. Such a signal could still potentially be recognized by a machine-learning model.

RQ8: *How to identify ships that are potential anomalous emitters using TROPOMI data?*

Another characteristic of the problem of ship emission monitoring is the fact that the ground truth data are not available. Potential bias of the TROPOMI measurements above the open sea on a global scale is unknown (there are no stationary in-situ measurement points). Moreover, due to the nonrigid structure of the ship plume, and the fact that some of the signal related to the plume can be below the detection capabilities of the human eye, the human-made labels used for the training of the ship-plume segmentation model may contain errors.

1.5 Outline

This thesis is based on a series of publications. The chapters present articles that have been peer-reviewed and published. The exceptions are this chapter and Chapter 2, which serve as background. Each following chapter of the thesis builds upon the findings of its predecessor, as a whole representing state-of-art knowledge in the application of TROPOMI satellite data for the monitoring of NO_2 emission from individual seagoing ships. The structure of the thesis is as follows:

In **Chapter 2**, we explain the general workflow that will be used in the thesis and introduce data sources that are necessary to combine in order to perform a ship NO_2 emission monitoring using TROPOMI data.

In **Chapter 3**, using the developed machine learning-based methodology, we examine the sensitivity limits of the detection system using TROPOMI data with respect to the detection of NO_2 plumes from individual seagoing ships. With this, we set up the research scope for further study. The chapter is based on the paper:

- Kurchaba, S., Sokolovsky, A., van Vliet, J., Verbeek, F.J., Veenman, C.J., 2024. Sensitivity analysis for the detection of NO_2 plumes from seagoing ships using TROPOMI data. Remote Sensing of Environment 304, 114041. doi:10.1016/j.rse.2024.114041.

After the limits of the satellite capabilities are established, we focus our attention on the evaluation of ship NO_2 emission. To focus the area of analysis on the region where the ship plume is expected to be located, in **Chapter 4**, we present a method that enables the automated assignation of a region of interest (RoI) to a studied ship. The RoI of a ship is established based on information about the position of the ship as well as the speed and the direction of the prevailing winds so that the plume of the studied ship is located within the designated area. Using the RoI of the ship, we can show the first attempts of automatic segmentation of a ship's plume. The chapter is based on the conference paper:

- Kurchaba, S., van Vliet, J., Meulman, J.J., Verbeek, F.J., Veenman, C.J., 2021. Improving evaluation of NO_2 emission from ships using spatial association on TROPOMI satellite data, in: 29th International Conference on Advances in Geographic Information Systems, pp. 454–457. doi:10.1145/3474717.3484213.

In **Chapter 5**, we study the possibilities of improving the quality of ship plume segmentation. To address the problem, we use supervised machine learning. Based on the previously defined RoI of a ship, we construct a set of features for training a

classification model to distinguish pixels that are part of a plume of a studied ship from those that are not. The chapter is based on the paper:

- Kurchaba, S., van Vliet, J., Verbeek, F.J., Meulman, J.J., Veenman, C.J., 2022. Supervised segmentation of NO_2 plumes from individual ships using TROPOMI satellite data. Remote Sensing 14. doi:10.3390/rs14225809.

In **Chapter 6**, we focus on developing a methodology for the automated detection of anomalously emitting ships. We leverage the methodology presented in the previous chapter, combining the RoI of a ship and a supervised method of ship plume segmentation, with a proposed machine-learning-based regression model for estimating NO_2 from ships. The chapter is based on the paper:

- Kurchaba, S., van Vliet, J., Verbeek, F.J., Veenman, C.J., 2023. Anomalous NO_2 emitting ship detection with TROPOMI satellite data and machine learning. Remote Sensing of Environment 297, 113761. doi:10.1016/j.rse.2023.113761.

Lastly, in **Chapter 7**, we present the main conclusions of the dissertation and possible directions for future work.

Chapter 2

General workflow

The aim of this study is to develop a methodology enabling the analysis of TROPOMI satellite data for the task of ship emission monitoring. In each chapter of the thesis, we address different aspects of this task. However, there is a sequence of steps that will be performed repeatedly in each chapter. Those steps are the integration of several data sources, feature engineering, selection of a machine learning model and optimization of its hyperparameters, the application to a given problem, and comparison of the estimated values of NO_2 with the theoretical measure of ship emission potential. We call it the general workflow (c.f. Figure 2.1). Each step of this general workflow will be introduced to the reader in this Chapter. The order, technical details, or methodology applied in each step will depend on the problem at hand and will be described in each chapter separately.
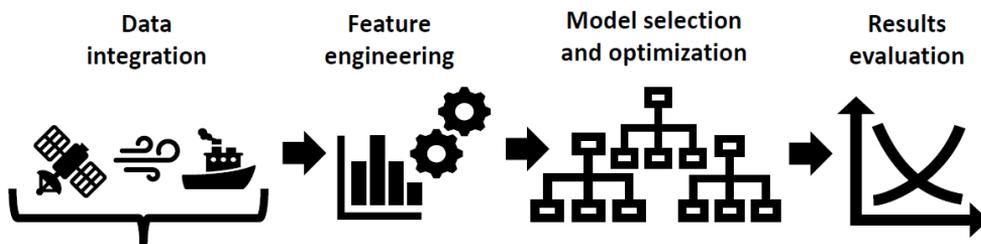

Figure 2.1: Visualization of general workflow.

2.1 Data integration

The first step of the general workflow is data integration. Within this thesis, we understand the process of data integration as follows:

Definition 2.1. Data integration is the process of combining data from several sources into one unified dataset in a way that enables the solution of a particular task at hand.

The main data source of this study is the set of NO_2 measurements coming from the TROPOMI instrument. These data contain among others information of our interest - the amount of NO_2 produced as a result of NO_x emission of individual seagoing ships. However, to enrich the TROPOMI data for our analyses, additional data sources are needed. In this study, we integrate the following data sources: 1) the TROPOMI NO_2 product, 2) data on ship positions and some properties of the ship, and 3) wind information. In the following subsections, we introduce the sources of data used in the study.

2.1.1 TROPOMI data

The Sentinel-5 Precursor (Sentinel-5P) satellite was launched in October 2017 and started its operational phase in April 2018. TROPOMI is a spectrometer on board the Sentinel-5P satellite mission – a sun-synchronous satellite with a local equatorial overpass time at 13:30. The instrument measures the Top Of the Atmosphere (TOA) solar radiation reflected by and radiated from the Earth covering ultra-violet up to the part of the visible spectrum (270-500 nm), near-infrared (675-775 nm), shortwave infrared (2305–2385 nm) spectral bands. The maximal ground pixel resolution of the instrument reaches 3.5×5.5 km^2 at the nadir, while the actual size of the pixel will vary depending on the true distance between the satellite and the captured part of the Earth's surface. We use Level 2 tropospheric NO_2 column data, publicly available via https://dataspace.copernicus.eu/ (previously https://s5phub.copernicus.eu/). In Chapter 3, the analysis is based on the TROPOMI data version 2.4.0. In Chapters 4, 5, the used data version is 1.3.0, and in Chapter 6 the study is conducted using TROPOMI data version 2.3.1.

The retrieval of NO_2 columns is performed using a 3-step procedure described in the Algorithm Theoretical Baseline Document [101]. A visual description of the process is presented in Figure 2.2. As a first step, NO_2 slant column densities are defined as the integrated amount of NO_2 along the average photon path from the Sun through the atmosphere back to the sensor [11, 41]. Next, based on the output from the
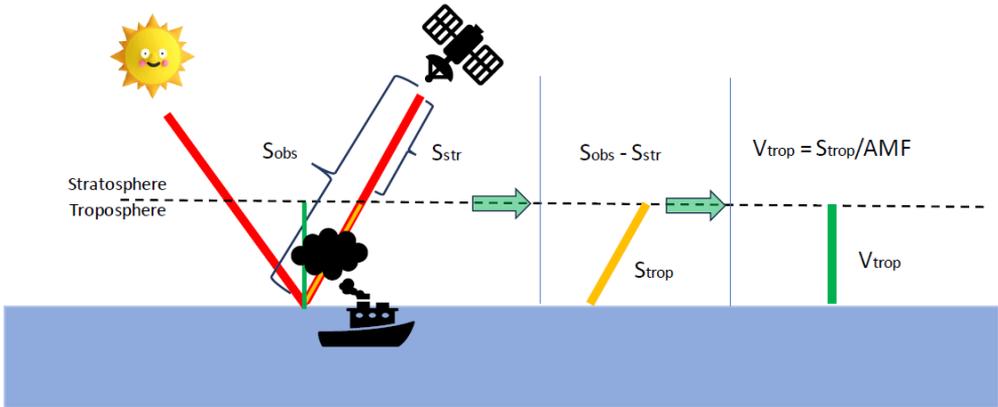
Figure 2.2: Illustration of the retrieval algorithm of NO_2 vertical tropospheric column density. S stands for Slant Column Density, V – Vertical Column Density. Visualization inspired by [105].

data assimilation system, the slant column is split into stratospheric and tropospheric components [29]. As the last step, using a tropospheric air mass factor (AMF), the tropospheric slant columns are transformed into tropospheric vertical column densities. The AMF accounts for the path length that sunlight travels through the atmosphere before reaching the satellite sensor, normalizing it by the amount of sunlight that would reach the surface under direct overhead conditions. Calculation of AMF to a large extent depends on the emission inventories and chemical transport models, which, in turn, rely on information about historical concentrations of emissions, including NO_2 [31]. Starting from Chapter 4, in all chapters of this thesis, we will base our analysis on tropospheric vertical column densities, as this variable ensures the best enhancement of NO_2 plumes. In Chapter 3, however, we study the sensitivity of the TROPOMI data-based detection system with respect to the NO_2 ship plume detection. Therefore, the historical information contained in AMF, and, in the resulting vertical column densities may cause information leakage. To prevent such a situation, the study presented in Chapter 3 will be based on tropospheric slant column density data.

2.1.2   Ship-related data

The second data source used in this study is information on ship positions. To coincide the detected NO_2 plumes with the emitting ships, the information on the positions of the ships at the moment of the satellite overpass is compulsory for this study. The used data on the positions of the ships comes from the Automatic Identification System

(AIS) transponders. Since 2002, all commercial sea-going vessels have the obligation to carry on board an AIS transponder [76] which transmits information about the position, speed, heading (direction), a unique identifier (MMSI), and the type of the ship.

At the moment, there is no open-access AIS with the spatiotemporal coverage and data quality required for this study. The data, however, can be accessed through several commercial providers. For the scope of this study, the AIS data as well as information about the dimensions of ships were provided by ILT, which has access to commercial databases.

In Figure 2.3, we present an example of TROPOMI data with the indicated positions of ships in the area starting from 2 hours before, until the moment of the satellite overpass. We can see that while the beginning of a ship's trajectory often corresponds with the origin of the plume, some significant deviations can be observed for the rest of the trajectory line. This happens because after the plume has been emitted by the ship, it is carried away in the direction of the prevailing winds. Therefore, for the efficient allocation of the plume with the ship emitter, information about the speed and the direction of the wind is required.

2.1.3 Wind data

Throughout the thesis, we use wind data from the European Center for Medium range Weather Forecasts (ECMWF). The wind fields (wind speed and wind direction) are the results of operational model analyses at a spatial resolution of $0.25°^1$, the temporal resolution of 6 hours and altitude of 10 meters. In [41], the wind data at 10 meters altitude was considered sufficient for ship-plume matching. Starting from the TROPOMI product version upgrade from 1.2.2 to 1.3.0 on March 27, 2019, the ECMWF 10-meter wind data for coinciding time is available as a support product in the TROPOMI data file [101].

2.2 Feature engineering

After data integration, the second step of the general workflow is feature engineering. We define feature engineering as follows:

Definition 2.2. Feature engineering is the process of extracting features from raw data.

---

[1]For the analyzed area the spatial resolution of $0.25° \times 0.25°$ translates to $\approx 23.4 \times 27.6$ km^2.
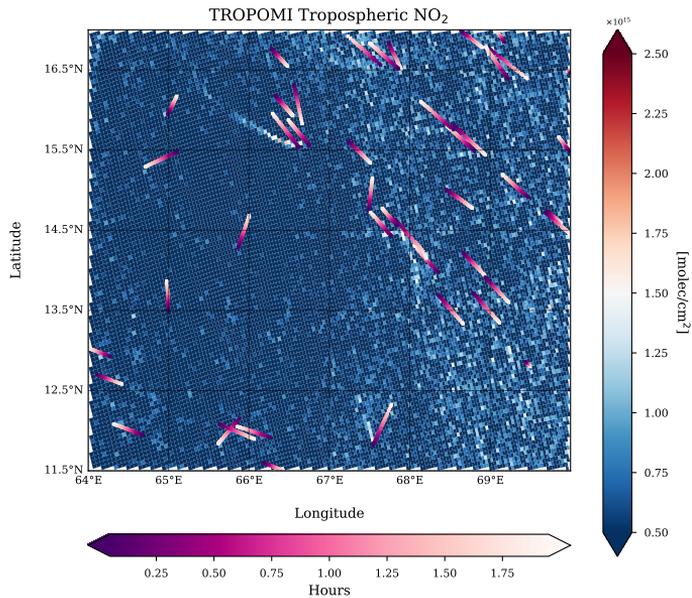
Figure 2.3: The NO$_2$ vertical tropospheric column density. Date: April 11th, 2020. Region: the Arabian Sea. Magenta lines indicate ship tracks based on information from AIS data. On the right-hand side of the map, an outflow effect from the variety of land bases NO$_2$ sources can be noticed.

Definition 2.3. Features are the set of characteristics associated with the data [75]. Other names of features are variables or attributes.

The aim of feature engineering is to transform the integrated dataset to be used by a machine-learning method for a particular task. Since in each chapter, we address different tasks, the applied methods of feature engineering will differ as well. The examples of feature engineering techniques that were used throughout the thesis include an assignation and further geometric transformation of the Region of Interest, data aggregation (through calculation of various statistics), encoding of spatial information, and categorical data encoding. They will be explained in the respective chapters.

2.3 Model selection and optimization

The next step of the workflow is the selection of a machine-learning model suitable for a given problem. This process includes the selection of the best-performing algorithm and the optimization of its hyperparameters. We define hyperparameters as follows:

Definition 2.4. Hyperparameters of a machine-learning algorithm are the parameters that steer the behavior of the learning process. The hyperparameters cannot be learned by the algorithm from its experience E (c.f. Definition 1.1) and need to be set by a researcher [55].

In the next subsections, we explain how we perform model selection and optimization in this thesis.

2.3.1 Machine-learning metrics

When performing model selection, we define beforehand evaluation metrics suited to the problem at hand. Several machine-learning metrics are used in this thesis and are defined below depending on whether we perform a binary classification or a regression task.

Binary classification metrics

In the context of binary classification, each classification result is assigned to one of the four categories:

- True positives (TP): the output data points with a positive class label correctly identified by the classifier.

- True negatives (TN): the output data points with a negative class label correctly identified by the classifier.

- False positives (FP): the output data points incorrectly identified by the classifier as positive.

- False negatives (FN): the output data points incorrectly identified by the classifier as negative.

The assignment to one category depends on the output of the model, that is the computed probability and a probability threshold. For instance, if the model gives a probability of 0.78 for a given data point and the threshold is set to 0.5, the data point will be labeled as positive. If the original label of the data point was positive, it is then correctly classified and is considered as TP. Using these categories, we can define performance metrics for a binary classifier.

First, we introduce a *precision-recall curve* – a graphical evaluation technique depicting precision as a function of recall where:

$$Precision = \frac{TP}{TP + FP} \tag{2.1}$$

$$Recall = \frac{TP}{TP + FN}, \tag{2.2}$$

Such a curve is obtained by using multiple probability thresholds to obtain multiple precision and recall points. When comparing the precision-recall curves of two models, if the precision and recall points of one curve are all above the points of the other, then the corresponding model is considered better than the other.

In the case of intersecting curves, to rank the models, we calculate an area under the precision-recall curve. We call this metric *average precision*. For a classifier that classifies all the data points correctly, the value of the *average precision* will be equal to 1. For a random guess classifier, the value of the *average precision* is equal to the ratio of positive samples in the dataset. Throughout the thesis, the average precision will be used as an evaluation metric for performing model selection and hyperparameter optimization.

The next method that we use in this thesis for the evaluation of the binary classifier is the *Receiver Operating Characteristic (ROC) curve* [35, 36]. The *ROC curve* is a graphical evaluation technique depicting all possible thresholds between the true positive rate (TPR), which is another name for *Recall*, and the false positive rate (FPR) [109], which we define as follows:
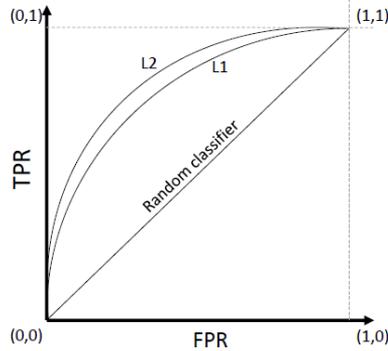
Figure 2.4: ROC curve - schematic example

$$TPR = \frac{TP}{TP + FN}, \tag{2.3}$$

$$FPR = \frac{FP}{FP + TN}, \tag{2.4}$$

An example of the *ROC curve* is presented in Figure 2.4. The classification results are perfect when $TPR = 1$ and $FPR = 0$. The classification results are completely wrong when $TPR = 0$ and $FPR = 1$. A diagonal line from the bottom left to the top right ($TPR = FPR$) corresponds to the results of a random-guess classifier.

Based on the *ROC curve*, we can compute the *Area Under the ROC Curve (ROC-AUC)* metric. The highest achievable score of *ROC-AUC* is equal to 1. In the case of random guessing, the *ROC-AUC* score will be equal to 0.5.

Regression model

For the evaluation of regression model performances, we use two metrics: *Pearson correlation coefficient* and *coefficient of determination* R^2. *Pearson correlation coefficient* ρ is defined as:

$$\rho = \frac{Cov(Y, \hat{Y})}{\sigma(Y)\sigma(\hat{Y})} \tag{2.5}$$

where Cov is the covariance, $\sigma(Y)$, and $\sigma(\hat{Y})$ are standard deviations of real and predicted values of a target variable respectively. The value $\rho = 1$ indicates a perfect linear correlation, value $\rho = -1$ indicates perfect linear anti-correlation and $\rho = 0$ is

the total absence of linear correlation. The second metric, R^2, is defined as:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{2.6}$$

$R^2 \in [0; 1]$, and is a measure of the goodness of fit of a model and is interpreted as the part of the variation of the predicted variable that is explained by the regression model. The $R^2 = 1$ suggests that the predictions obtained with a regression model fit the data perfectly well. We use R^2 as the quality metric for the process of model selection and optimization.

2.3.2   Hyperparameter optimization strategies

There are two strategies of algorithm selection and optimization of its hyperparameters used in this thesis. The first is a selection of the optimal algorithm among the list of pre-selected candidates while performing a randomized search [10] for the hyperparameters optimization. The benefit of this strategy is that we can explore the performance of pre-selected candidates and quantify the gain achieved from the usage of more complex techniques. The second strategy is to directly solve the so-called CASH problem (Combined Algorithm Selection and Hyperparameter optimization [57]) using automated machine learning (AutoML) [49]. AutoML deals with the automation of the application of machine learning to real-world problems [55]. The CASH problem is the task of selecting a suitable machine-learning algorithm (which can be a combination of several algorithms) for the analyzed dataset, together with the proper pre-processing methods and set of hyperparameters of all components involved, without requiring human intervention [55]. The advantage of using this strategy is that such a technique enables an efficient selection of a machine-learning algorithm and feature preprocessor from a more extensive list of candidates within a limited time frame. This is particularly useful when performance benchmarks are unavailable. The disadvantage of such a technique, however, is that the comparison of the performance of several models cannot be done directly (as weaker candidates are discarded during the process of optimization). In this thesis, we address the CASH problem using TPOT (Tree-based Pipeline Optimization Tool) [77] – a Python package for automatic selection of machine learning pipelines based on genetic programming (GP) [58].

In order to combine the process of model performance evaluation with the process of algorithm selection and optimization of its hyperparameters, we apply a *nested cross-validation* scheme [96, 18]. The general setup of *nested cross-validation* is as
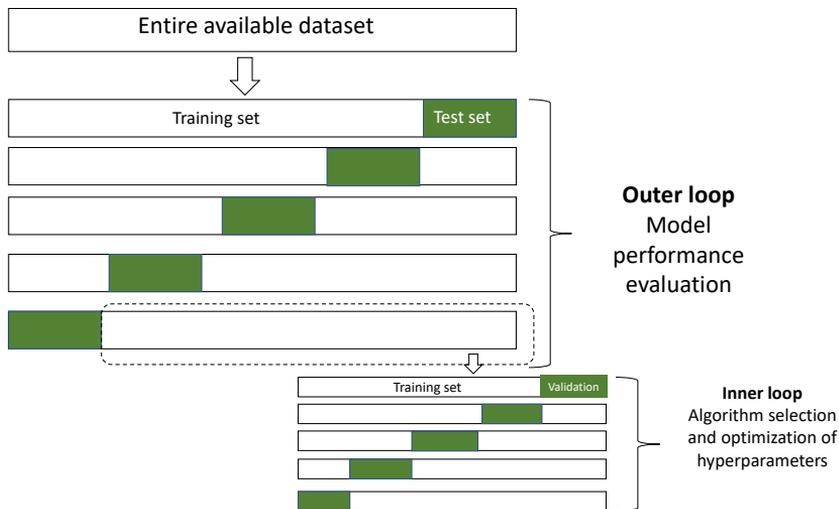
Figure 2.5: Schematic representation of nested cross-validation. In the inner loop, the generated training and validation sets are used to find an optimal set of the hyperparameters of the model. In the outer loop, we generated a series of test sets that are used for the model performance evaluation.

follows: In the outer loop of cross-validation, the entire dataset is split into K subsets (folds). The model is trained on K-1 subsets, while the remaining subset is used for the model evaluation. This procedure is repeated K times. Within each iteration of the outer loop, an inner cross-validation loop is performed. The training data from the outer loop is further split into K-1 subsets for training and one subset for validation. Different model hyperparameters are tested using the training and validation sets in the inner loop. The model with the best performance on the inner loop validation set is selected. The selected model from the inner loop is then evaluated on the test set from the outer loop. For a visual explanation, see Figure 2.5. Note that, in the field of statistics, another naming convention for resulting splits of the data is used (validation is then called test set, and vice-versa). The main advantage of the nested scheme of cross-validation is the prevention of information leakage coming from using the same data for the evaluation of model performance and tuning the hyperparameters, which takes place in case straightforward cross-validation is applied [18].

2.4   Results evaluation

The last step of the general workflow is a comparison of estimated values of NO_2 with some kind of independent measurement. However, as mentioned earlier, for the task of ship emission monitoring, TROPOMI is the only way of measurement above the open sea. The "ground truth" data for this task is not available. To overcome this, we will use a theoretical ship emission proxy E_s as a reference value defined as follows:

$$E_s = L_s^2 \cdot U_s^3, \tag{2.7}$$

where L_s is the length of the ship in meters (m), and U_s is its speed in meters per second (m/s). The details of the derivation of the given measure can be found in [41], where the proxy was introduced. As it is noted in [41], the advantage of E_s in comparison to other ship emission proxies (e.g. [33]) is that it can be calculated based on AIS data only, while other existing emission proxies require ship information that is not in the AIS data and is not available publicly. This, however, will result in some loss of the quality of emission approximation.

To sum up, in this Chapter, we described general process steps that will be used throughout the thesis. We call it the general workflow. The applied methodological details of each of the described steps may differ depending on the task at hand and will be described in each chapter separately.

2.4.   Results evaluation

Chapter 3

Sensitivity analysis for the detection of NO$_2$ plumes from seagoing ships using TROPOMI data

Abstract The marine shipping industry is among the strong emitters of nitrogen oxides (NO_x) – a substance harmful to ecology and human health. Monitoring of emissions from shipping is a significant societal task. Currently, the only technical possibility to observe NO_2 emission from seagoing ships on a global scale is using TROPOMI data. A range of studies reported that NO_2 plumes from some individual ships can be visually distinguished on selected TROPOMI images. However, all these studies applied subjectively established pre-determined thresholds to the minimal speed/length of the ship – variables that to a large extent define the emission potential of a ship. In this Chapter, we investigate the sensitivity limits for ship plume detection as a function of their speed and length using TROPOMI data. For this, we train a classification model to distinguish TROPOMI image patches with a ship, from the image patches, where there are no ships. This way, we exploit ground truth ship location data to potentially exceed human visual distinguishability. To test for regional differences, we study four regions: the Mediterranean Sea, Biscay Bay, Arabian Sea, and Bengal Bay. For the Mediterranean and the Arabian Sea, we estimate the sensitivity limit to lie around a minimum speed of 10 knots and a minimum length of 150 meters. For the Biscay Bay – around 8 knots and 100 meters. We further show that when focusing the analysis on the biggest emitters (junctions of several ships in the area), the detectability can be improved up to above 0.8 ROC-AUC. Finally, we show that increasing the size of the dataset, beyond the dataset used in this study, yields further improvements in the detectability of smaller/slower ships. The rate of improvement in both experiments is dependent on the region studied.

3.1   Introduction

As it was mentioned in the Introduction of this thesis, the TROPOMI/S5P is the first satellite-based instrument that gives the possibility to visually detect NO$_2$ plumes from some individual seagoing ships [41]. This is due to significantly higher than its predecessor spatial resolution of the instrument. Such an improvement in the quality of the remote-sensing-based atmospheric monitoring allows to consider the TROPOMI/S5P instrument as a potential solution for the task of global and continuous monitoring of the emissions produced by seagoing ships [90]. However, in order to fully understand the potential of the TROPOMI for a given task, the first step is to estimate the limitations in terms of the sensitivity of the detection system for NO$_2$ plumes from seagoing ships using TROPOMI data.

To tackle the problem, we prepare image patches – small, regular-sized sections of the TROPOMI measurement (image). We use the created image patches to train a machine-learning classification model. The task of the model is to distinguish image patches with at least one ship from the image patches where there are no ships. The labels of the model were created using AIS ship location data, and, therefore, are independent of the distinctivity of ship plumes by a human. This way, we formulate the research questions of the study as follows:

- **RQ1**: What is the minimum speed and length of a seagoing ship so that the NO$_2$ plume from it can be detected with the detection system using TROPOMI data?

- **RQ2**: To what extent can the detectability of NO$_2$ plumes be improved if only the biggest emitters are taken into account? With the biggest emitters, we mean the biggest ships operating at the highest speeds, or several smaller or slower ships operating in proximity to each other.

- **RQ3**: Is there a potential for improvement of detectability of NO$_2$ plumes from the slow/small ships if more data were used to train the used classification model?

We conduct this study on four regions of interest: Mediterranean Sea, Biscay Bay, Arabian Sea, and Bengal Bay (the coordinate scope see in Table 3.1 and Figure 3.1). The study areas are directed towards the Europe – Middle East – Asia trade route, with selected areas representing low background pollution and common occurrence of clear skies.

The rest of the Chapter is organized as follows: In Section 3.2, we explain how the data was pre-processed in order to obtain datasets used for machine learning

models. In Section 3.3, we introduce the experimental setup for each stage of the study and present the obtained results. We discuss the obtained results in Section 3.4 and conclude in Section 3.5.
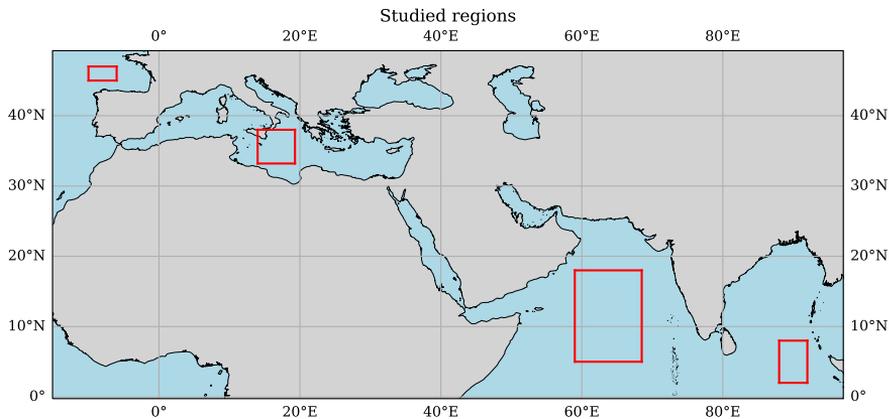
Figure 3.1: Four studied regions (from left to right): Biscay Bay, Mediterranean Sea, Arabian Sea, Bengal Bay.

Region	Longitude [deg]	Latitude [deg]	Studied period
Mediterranean	(14, 19.3)	(33.2, 38)	(31-03-20; 28-02-23)
Biscay Bay	(-10, -6)	(45, 47)	(01-04-20; 28-02-23)
Arabian Sea	(59, 68.5)	(5, 18)	(31-03-20; 30-11-22)
Bengal Bay	(88, 92)	(2, 8)	(03-06-20; 31-12-22)

Table 3.1: Geographical coordinates and analyzed periods defining the study scope for each region.

3.2 Dataset

The supervised learning task that is addressed in this study is to distinguish image patches with a ship plume on them. In this Section, we describe the process of the preparation of the dataset for the given supervised learning task. We first describe the undertaken steps of data pre-processing. We then introduce the features used for the model training and define the target variable.

Region	Ship image	No ship image
Mediterranean	16%	18%
Biscay Bay	48%	52%
Arabian Sea	49%	52%
Bengal Bay	54%	54%

Table 3.2: Percentage of data from the original dataset lost when a *qa value* of .75 is applied for filtering.

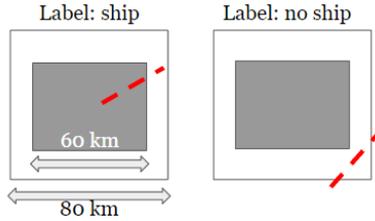Figure 3.2: An illustration of the set-up used for counting the number of ships per image patch. White square – image patch. Grey square – a central part of the image patch. Red dashed lines – an example of ship trajectory starting from 2 hours before until the moment of the satellite overpass. Only ships, whose trajectories cross the central part of the image patch are considered to be present in the area covered by a patch.

3.2.1   Data preprocessing

The first step of data preparation is regridding[1]. This is done so that for each region we have pixels with the same spatial coverage. The regridded pixel size for each region is approximately equal to 4×5 km^2. Following the set-up used in the previous studies [63, 64], for the regridding, we only use pixels with cloud coverage below 0.5, wind speed lower than 10 m/s, and *qa value* above 0.5 [93]. This level of *qa value* filtering was shown to be sufficient for the identification of NO$_2$ plumes from individual ships and is a trade-off between a high standard of data quality, and an attempt to preserve as many data points as possible. In Table 3.2, the reader can find an assessment of the data loss in case *qa value* filtering was set to the level of 0.75 – the level suggested in the TROPOMI manual [31].

As a next step, we split the studied area into non-overlapping patches of equal size 80×80 km^2. The selected size of the image corresponds to a distance that the fastest ships in the dataset will cover in 2 hours. The observation period of 2 hours

---

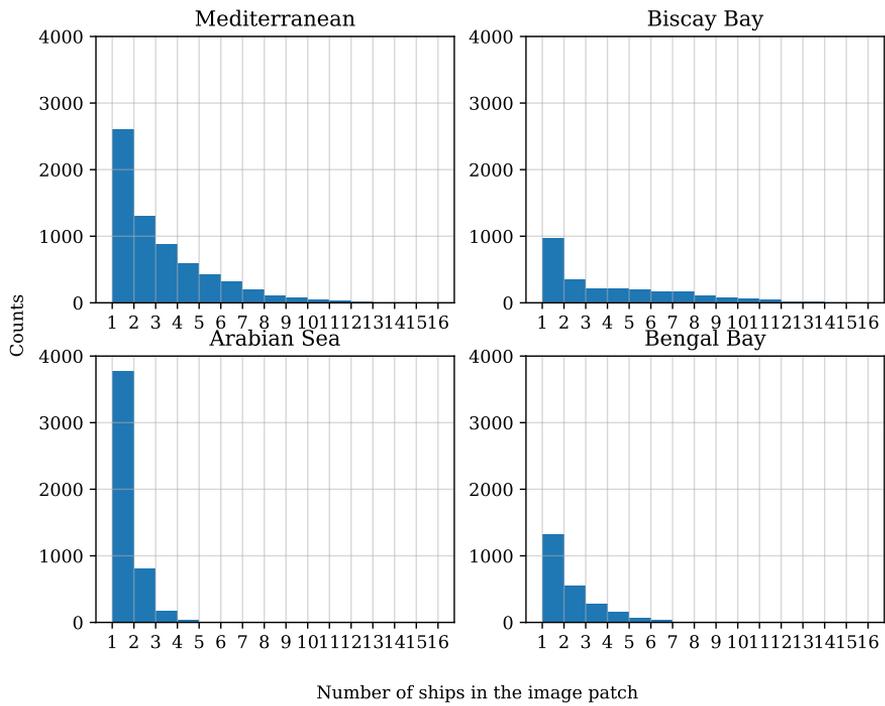[1]The regridding is performed using the Python package HARP v.1.13.

Figure 3.3: Distribution of the number of ships per image patch for the studied regions.

Region	Ship image	No ship image
Mediterranean	6652	9693
Biscay Bay	2641	2812
Arabian Sea	4804	24594
Bengal Bay	2444	6848

Table 3.3: Class-wise distribution of image patches for each studied region. The rate of imbalance depends on the traffic density in the region.

was motivated by the fact that due to the physical dispersion and limited lifetime of NO$_2$ within plumes, the detectability of ship plumes will fall sharply after 2 hours [107]. For each image patch, we calculate how many ships were in the central area of the patch within 2 hours before the overpass of the satellite. The central area of the patch is defined as 60×60 km^2 square. We do not take into account ships that do not pass through the central area of the image patch, as the probability that their plume will be located within the image patch is very low. An example is presented in Figure 3.2. The resulting distribution of the number of ships per image patch for each studied region can be found in Figure 3.3. Please note the regional differences in the distribution of ships among patches. The Arab Sea typically has a high number of patches with a single ship. The Biscay Bay, in comparison to other regions, has the highest number of patches with a high number of ships on it. These patterns illustrate the difference in shipping density among the studied regions.

3.2.2   Feature engineering

To study the sensitivity of the TROPOMI instrument with respect to the detection of NO$_2$ plumes from seagoing ships, we prepare a dataset for supervised machine learning. The NO$_2$ trace gas variable of our interest is *Tropospheric Slant Column Density – SCD trop* [31]. As mentioned in Chapter 2, the SCD variable is suitable for satellite sensitivity study [41] as its derivation is not based on airmass factor – a variable estimated based on, among others, historical NO$_2$ concentration within a certain area.

The objective is to distinguish image patches that cover the area where there are no ships, from image patches covering the area with at least one ship on it. Since this is a binary problem, the value of the output label is 1, if there is at least one ship that is faster than 6 kt, which is approximately 11.1 km/h and longer than 90 m in the area covered by an image patch. The output label is 0, if there is no ship in the area,

or the ship is shorter than 90 m or slower than 6 kt. The values of 90 m and 6 kt are sufficiently low to be well below detectable limits as will also follow from this study. Table 3.3 shows the resulting distribution of classes for studied regions. Examples of image patches without (label 0) and with at least one ship on it (label 1) are presented in Figure 3.4. We can see that not all image patches with a ship actually contain a visually distinguishable plume. This is because the NO_2 plumes produced by some ships are below the sensitivity limit of the TROPOMI instrument, or we are not able to distinguish it visually.

We address the classification problem with a multivariate classifier. Therefore, we represent the TROPOMI image patches in terms of a set of features - a statistical representation of the image patch. More specifically, for the regridded pixels of each image patch, we calculate the following statistics: *min(SCD), mean(SCD), median(SCD), max(SCD), std(SCD)*, where SCD stands for NO_2 slant column density. To give information about the level of plume dispersion, we add wind-related variables *zonal wind velocity (wind zon), meridional wind velocity (wind med)*, which represent the speed of the wind from the west to east and from south to north respectively. Finally, we add features *sensor zenith angle, solar zenith angle* and *solar azimuth angle* to represent the viewing geometry of the satellite. Values for wind information and satellite geometry are the average values of the pixels within the image patch. The resulting feature set is presented in Table 3.4. In Figure 3.5, the reader can find histograms of the dataset features for the studied regions. Clearly, the features related to the properties of ships cannot be included in the feature space, because the presence of a ship has to be established. Moreover, we deliberately do not include any features in the feature set related to the geographic locations of a given patch. This is because shipping lanes may bias the model. The dataset used in this study as well as the code used for generating the presented in this study results are available publicly as a reproducibility capsule [60]. Prior to the application of a machine learning model, all features were standardized using a median-interquartile range scaling[2] – a scaling technique that allows to reduce a negative impact of the outliers in the dataset [32].

3.3   Experiments and results

In this Section, we describe the experiments and show the results obtained. We start with the introduction of the classification model – we present model selection and hyperparameter optimization results. For the selected model, we provide the explain-

---

[2]RobustScaler implemented in scikit-learn v.1.2.2.
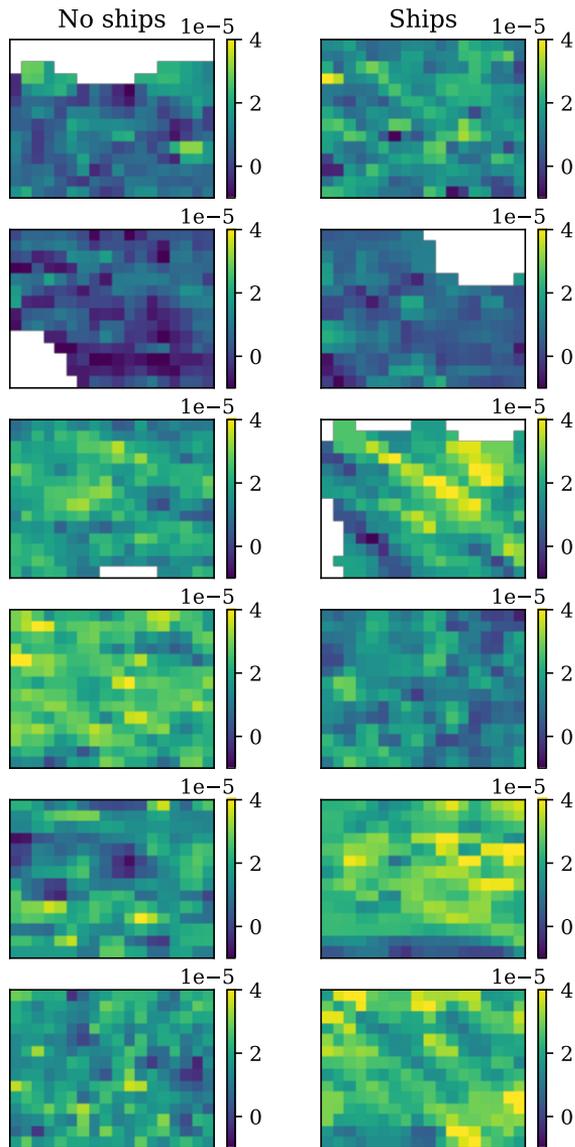
Figure 3.4: Examples of image patches without a ship and with at least one ship on it. The presented image patches were randomly sampled from the dataset of the region Biscay Bay.

Figure 3.5: Histograms of the variables from the dataset.

Feature type	Feature name
NO$_2$ slant column density	min(SCD)
	mean(SCD)
	median(SCD)
	max(SCD)
	std(SCD)
Wind information	zonal wind velocity
	meridional wind velocity
Satellite geometry	sensor zenith angle
	solar zenith angle
	solar azimuth angle

Table 3.4: List of features used for classification model.

ability analysis. Next, in the consecutive subsections, we explain and provide the results of the experiments addressing the three research questions of this study.

3.3.1   Classification model

Experimental setup

As a first step, we compared the performance of several multivariate classifiers and selected the one that is going to be used in the remaining part of the Chapter for the sensitivity analysis. We studied four machine learning classifiers of increasing complexity: Logistic regression, Support Vector Machine (SVM) with the radial basis function (rbf) kernel, Random Forest[3], and Extreme Gradient Boosting[4] (XGBoost) [22]. All selected models are robust to noise and can be efficient even given the relatively small size of datasets. To make sure that we exploit the maximum potential of a given machine learning model, we optimized the hyperparameters of each studied model. The hyperparameters were optimized using a random search[5] technique with the objective metrics - *average precision*. The used search space of the hyperparameters for each of the models studied as well as the results of the hyperparameters optimization can be found in the original paper [61]. To be able to simultaneously perform the hyperparameter optimization and evaluation of the model performance, we use 5-fold nested cross-validation [96, 18] (for the explanation of the concept and visual example see Section 2.3). To maintain the same percentage of samples of a certain label in the training and test set, the cross-validation was based on *stratified K-fold* splits [47, 42].

Results

The classification results are presented in Table 3.5. Comparing the performances between different classifiers, we can see that the XGBoost classifier yielded the best results for most of the regions – we used this classifier for the remaining experiments of this study. Comparing the results between regions, we start with ROC-AUC. The highest achievable score of ROC-AUC is equal to 1. While the ROC-AUC score that will be obtained in case of random guessing is 0.5. The ROC-AUC score is calculated based on the ROC curve. For the XGBoost classifier, it is presented in the right-hand side plot of Figure 3.6. The scores for Biscay Bay and the Mediterranean Sea are higher than for the Arabian Sea and Bengal Bay. One of the reasons for this difference

---

[3]All above-mentioned models are implemented in Python scikit-learn v.1.2.2.

[4]XGBoost v. 1.7.0

[5]Implemented in Python scikit-learn v.1.2.2.

Region	Model	Average Precision	ROC-AUC
Mediterranean	XGBoost	**0.636 ± 0.013**	**0.712 ± 0.011**
	Random Forest	0.629 ± 0.018	0.706 ± 0.016
	SVM (rbf)	0.615 ± 0.015	0.694 ± 0.013
	Logistic	0.448 ± 0.008	0.546 ± 0.009
Biscay Bay	XGBoost	**0.704 ± 0.021**	**0.713 ± 0.015**
	Random Forest	0.620 ± 0.025	0.652 ± 0.022
	SVM (rbf)	0.573 ± 0.020	0.589 ± 0.014
	Logistic	0.523 ± 0.013	0.541 ± 0.018
Arabian Sea	XGBoost	0.226 ± 0.007	0.610 ± 0.008
	Random Forest	**0.229 ± 0.006**	**0.618 ± 0.006**
	SVM (rbf)	0.195 ± 0.004	0.545 ± 0.007
	Logistic	0.169 ± 0.003	0.498 ± 0.008
Bengal Bay	XGBoost	**0.379 ± 0.017**	**0.601 ± 0.01**
	Random Forest	0.364 ± 0.016	0.601 ± 0.010
	SVM (rbf)	0.346 ± 0.006	0.560 ± 0.016
	Logistic	0.289 ± 0.015	0.542 ± 0.016

Table 3.5: Results of the optimization of the classification models' hyperparameter. The reported results were obtained on the hold-out test sets based on nested 5-fold cross-validation [96, 18]. The bold font indicates the performance of the best model for a given region.
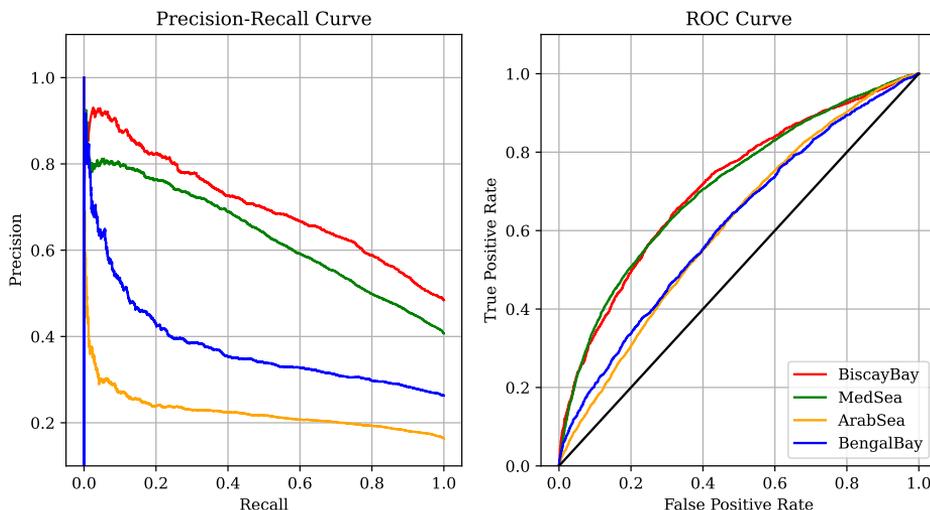

Figure 3.6: Precision-recall and ROC curves for the studied regions. The black line in the right panel – performance of a random guess classifier.

might be that the regions Biscay Bay and the Mediterranean Sea have a higher overall number of ships per image patch (and, therefore, a higher percentage of potentially well-recognizable plumes) than the two remaining regions, c.f. Figure 3.3. Next, we compare the scores of average precision. Also in the case of this metric, a perfect classifier would have a score of 1.0, while a random guess classifier would have an average precision score equal to the ratio of positive samples in the dataset. The average precision score is calculated based on a precision-recall curve, which is presented in Figure 3.6, left-hand-side plot. Due to the different rates of class imbalance of datasets from different regions, the average precision scores from the Table are difficult to compare directly. However, analyzing the precision recall-curves, we can conclude the following: the performance of the classifiers on Biscay Bay and Mediterranean Sea regions are very close to each other and the difference between the obtained average precision scores is mainly caused by a slightly different class imbalance. The lower average-precision scores for the regions Bengal Bay and Arabian Sea are also to a big extent a result of the fact that those datasets contain fewer image patches with a ship than two other regions. However, in the case of Bengal Bay, for the lower rates of recall, we can observe quite high values of precision. This signalizes the fact that there is a set of images that the model can quite confidently correctly recognize. This is not the case for the Arabian Sea, which implies better performance of the classification model on the Bengal Bay region in comparison to the Arabian Sea. For all regions, it is important to underline that the reported performances of the models were negatively affected by the presence of ships whose size and speed are known to be too small or slow to be detected by the TROPOMI instrument, which is a cause of the topic of this research, that is the study of the detection limits.

Explainability analysis

As a next step, we would like to understand which of the used features are the strongest indicators of the presence of a ship in the area for the XGBoost model. For this, we perform the explainability analysis using the SHapley Additive exPlanations (SHAP) [70] summary plots (see Figure 3.7). The plots indicate the strength of the impact of a value of a certain model feature on the model outcome (positive or negative) for individual samples from the test set. The red and blue colors show the effects of a certain feature's high and low values respectively.

We can see that for the Mediterranean Sea, and Biscay Bay, the feature having the strongest impact on the decision of the model the most is *scd std*, representing the standard deviation of stratospheric column density within the image patch. In the
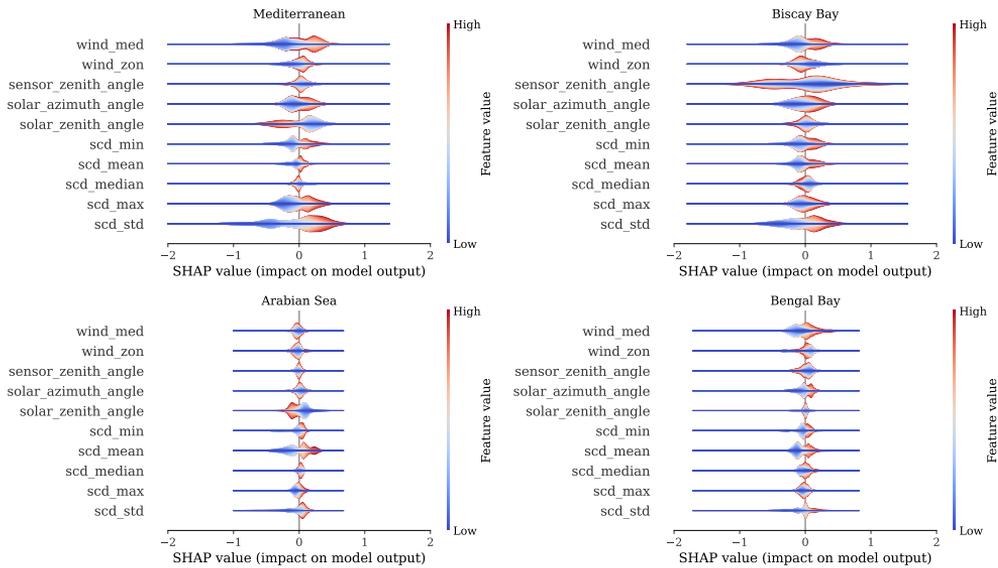
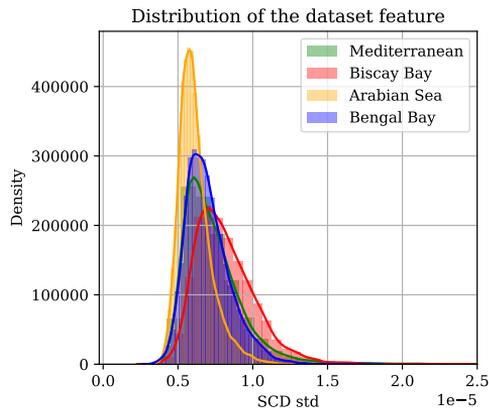Figure 3.7: SHAP violin plots on concatenated test sets for each studied region.

Figure 3.8: Distribution of the variable *scd std* for four studied regions. For the Arabian Sea, the distribution is noticeably more narrow than for other regions.

case of the Mediterranean Sea, *scd max* and *solar zenith angle* also play significant roles. Interestingly, the direction of the meridional wind also has a strong influence on the model's decision in the Mediterranean Sea. From the plot, we see that the negative meridional wind corresponds to strong negative model responses, potentially due to land outflow from Europe affecting ship plume visibility. In the Arabian Sea and Bengal Bay regions, the strongest impact on the model response is attributed to the values of the feature *scd mean*. Notably, for the Arabian Sea, high values of *scd std* do not necessarily indicate the presence of a plume, possibly because as we can see from Figure 3.8, standard deviations of NO_2 concentrations in this region are typically lower compared to others. Low values of *scd std*, however, are used by the model as a strong suggestion of the absence of a plume in the image patch. Finally, one can notice that for Biscay Bay, the feature *sensor zenith angle* is of great importance. However, since we do not see a clear split into high/low values for positive/negative model outcomes, the influence of the feature on the model response will depend on the values of other features [40, 47]. From this experiment, we can conclude that the same machine learning models applied to different studied regions not only yield different quality of results but are also driven by different sets of features.

3.3.2   Sensitivity limits estimation

In this Subsection, we address the first research question: What is the minimum speed and length of a seagoing ship so that the NO_2 plume from it can be detected with the detection system based on TROPOMI data? With the detection system we mean a sequence of steps needed to automatically detect an NO_2 plume from a ship on a TROPOMI image patch. The first step of this sequence is a measurement performed by the TROPOMI sensor. The last step is the application of a trained machine-learning model on the set of unseen image patches with the aim of distinguishing patches covering the area with a ship. In [41], it was shown that the length and the speed of the ship are the main factors determining the emission potential of the ship. Following the considerations presented in [41], in order to decrease the level of problem complexity, we represent the length/speed of the studied ship in terms of one variable – the ship emission proxy E_s [41], as defined in Section 2.4. For the purpose of this study, we define the sensitivity limit of the detection system for NO_2 plumes from seagoing ships using TROPOMI data for a given region as the level of ship emission proxy E_s, starting from which the classification model can distinguish image patches without a ship from image patches with a ship.

Region	Average Precision	ROC-AUC
Mediterranean	0.538 ± 0.036	0.518 ± 0.038
Biscay Bay	0.539 ± 0.053	0.513 ± 0.067
Arabian Sea	0.563 ± 0.035	0.560 ± 0.031
Bengal Bay	0.564 ± 0.054	0.540 ± 0.060

Table 3.6: Model performance when only considering the one-ship patches with the emission proxy below 10% quantile.

Given the provided definition of the sensitivity limit, our initial investigation evaluates the classification model's performance using image patches with the lowest total emission proxy. For this, we first exclusively chose patches covering a single ship. Then, from the selected subset, we further narrowed our selection to those patches with an emission proxy falling below the 10% quantile of all one-ship patches. To ensure comparability of performance metrics between areas and samples with different ship proxy values, we took a sample with an equal number of patches with and without a ship covered by the patch. To make sure that all image patches with and without ships that satisfy the above-provided criteria are used for the model training and evaluation, we repeated the sampling procedure 5 times. Subsequently, we conducted a 5-fold cross-validation for each set of sampled data points. The averaged results over the five folds are presented in Table 3.6. The outcomes indicate that none of the regions allowed for distinguishing patches with a ship, as the ROC-AUC/Average precision values obtained were not significantly higher than 0.5. Consequently, we infer that the ships with the lowest emission proxies in our dataset fall below the sensitivity limit of the detection system for NO_2 plumes from seagoing ships using TROPOMI data.

In the next experiment, we checked what the emission proxy threshold for the ship plumes detectability is. Here, we again considered only image patches with one ship on it. We then gradually removed ships with the lowest emission proxy from the dataset, analyzing the changes in the model performance. The applied emission proxy thresholds were determined as a range of quantiles starting from 10% and gradually increasing by 10%, until it reaches 90%. If after reaching a certain level of threshold, the number of patches with a ship (label 1) went below 300, the experiment was terminated and the next thresholding levels were not tested[6]. The criterion of 300 patches was established based on the number of patches with a ship left after a 90%

---

[6]This way, the highest applied threshold for Biscay Bay was 70% and for Bengal Bay 80% quantile.
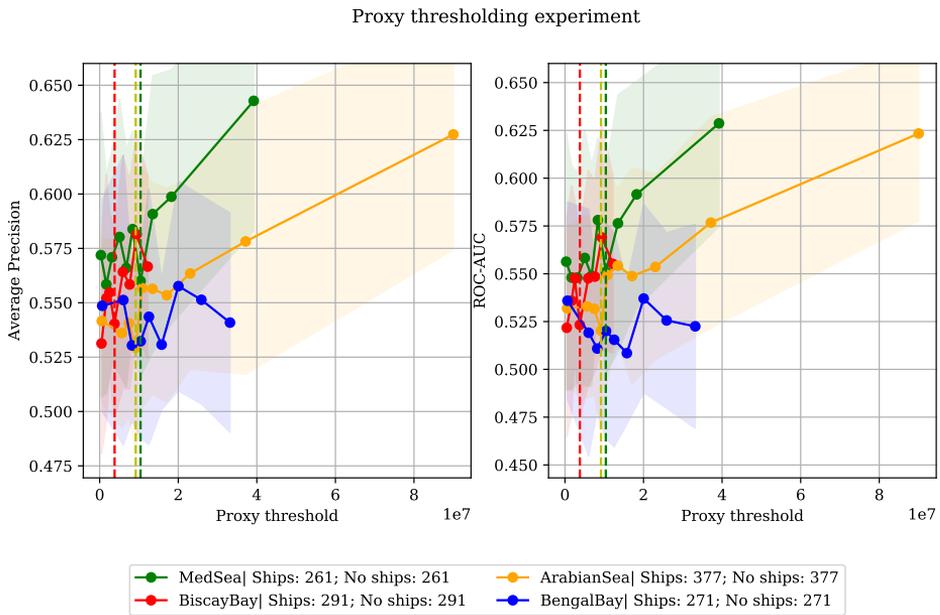
Proxy thresholding experiment

Figure 3.9: Step-wise removal of the patches (containing one ship) with the lowest emission proxy. Dashed lines indicate estimated levels of sensitivity limits for the Biscay Bay, Mediterranean, and Arabian Seas. To assure the comparability of the results, a similar size of training/test datasets was used at each threshold level.
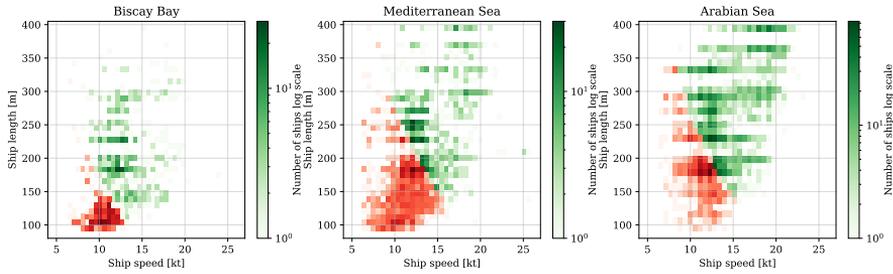
Figure 3.10: 2D histograms of speed and lengths for ships that are above (green) and below (red) the estimated sensitivity limits for the Biscay Bay, Mediterranean, and Arabian Seas.

threshold applied for the region with the highest number of one-ship patches available (Arabian Sea). Clearly, by removing the image patches with the proxy values below a certain threshold, we decreased the size of the dataset. To eliminate the potential effect of the dataset size on the model performance, throughout the experiment, we kept the dataset size constant. To achieve this, for each applied threshold, we sampled the number of data points equal to the number of data points available for the highest applied threshold. As in the previous experiment, we repeated the sampling procedure 5 times. For each set of sampled data points, we performed a 5-fold cross-validation.

The results of the experiment are presented in Figure 3.9. We can see that for the lowest thresholds, for all four regions, the average performance quality did not change. This means that the removed ships were still below the sensitivity level of the detection system for NO_2 plumes from seagoing ships using TROPOMI data. From a certain threshold (indicated with dashed lines on the plot), however, the model performance started to increase. The level of the ship emission proxy threshold starting from which we observe the improvement of the performance of the model is the sensitivity limit of the detection system for NO_2 plumes from seagoing ships using TROPOMI data for a given region. For the Mediterranean and the Arabian Sea, the sensitivity limit in terms of ship emission proxy was established to be around $1 \times 10^7 m^5/s^3$. For the Biscay Bay, the sensitivity limit is lower and is around $3.8 \times 10^6 m^5/s^3$. To get the intuition around these numbers, we return to the values of speed and length of the ship. To achieve this, for the regions of the Biscay Bay, Arabian, and Mediterranean Seas, in Figure 3.10, we present 2D histograms of the speed and length of ships that are above (green color) and below (red color) the estimated sensitivity limits. From the histograms, we conclude that to distinguish NO_2 plumes, the minimum speed of the ship for the Arabian and Mediterranean Seas should range between 10 and 15

kt depending on the length of the ship. Ships that are slower than 10 kt or shorter
than 150 m are below the sensitivity limit. For Biscay Bay, the limit lies around 8
kt and 100 m. For Bengal Bay, the sensitivity limit cannot be determined since the
available amount of data did not allow us to raise the proxy threshold high enough to
see the increase in the performance of the model. However, when comparing the curve
dynamics of the Bengal Bay with other regions, the obtained pattern suggests that
the sensitivity limit for this region is higher than for the Arabian and Mediterranean
Seas.

3.3.3   On detection of biggest emitters

Our second research question is how the detectability of NO$_2$ plumes can be improved
if only the biggest emitters are taken into account. Our aim here is to understand the
potential of the detectability of NO$_2$ plumes when the total emission proxy is very high.
The high emission proxy can result from a big ship operating at a high speed, or smaller
or slower ships operating in proximity to each other. Therefore, in this experiment,
we considered all image patches (without, with one, or with more than one ship on it).
This way, in some of the image patches, there will be more than one ship with a high
emission proxy present. As in the previous experiment, we gradually removed from
the dataset the image patches with the lowest total emission proxy. Once again we
studied how the removal of the low emitters affects the quality of classification. The
thresholds used for the proxy filtering were determined as quantiles of the proxy values
of the dataset of a given region. For the Mediterranean and Arabian Sea, the applied
quantiles ranged from 0 to 90%. For the Biscay and Bengal Bay, due to the smaller
sizes of the datasets, the applied quantiles ranged from 0 to 80%. In Figure 3.11, we
present the results of the experiment. For each of the studied regions, we can observe
an increase in the model performances. We can see that for the Mediterranean Sea,
for the patches with the highest total emission proxy, the ROC-AUC score can exceed
0.8. For the regions Arabian Sea and Bengal Bay, the level of the results is noticeably
lower. This pattern in the results is similar to what we observed in Subsection 3.3.1.

   As a next step, we checked if the dependency between the applied proxy threshold
and classification performance is impacted by a certain hyperparameter configuration
of the XGBoost model. We would like to know to which extent we can improve the
quality of classification for the image patches with the highest total emission proxy.
For this, we studied two configurations of the dataset. In the first case, we applied the
highest proxy threshold for the given region (the last data point from the corresponding

Proxy thresholding experiment
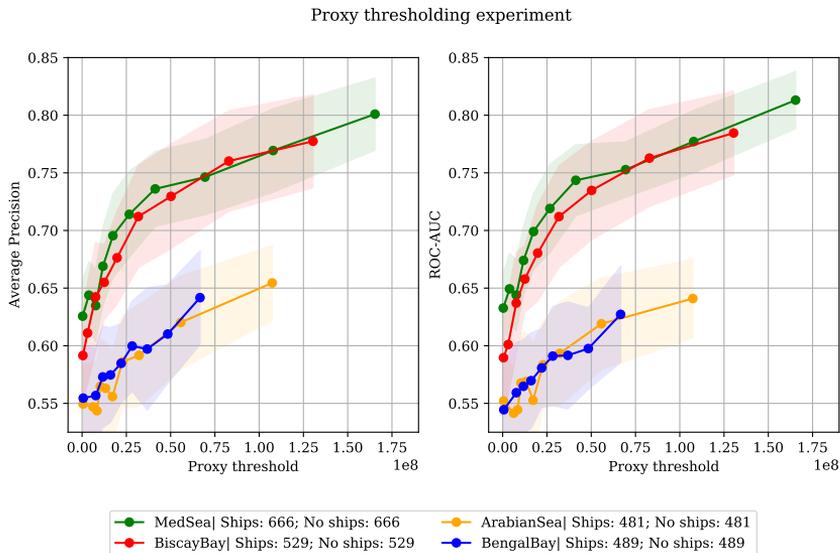
Figure 3.11: Illustration on how the step-wise removal of the image patches with the lowest total emission proxy from the dataset affects the performance of the classification model.

plots of Figure 3.11). In the second case, we did not apply any proxy threshold but kept the dataset size equal to the case when the proxy threshold was applied (the scenario corresponds to the first data point of the corresponding plots of Figure 3.11). For each of the datasets, we performed optimization of the hyperparameters of the classification model, in the same way as it is explained in 3.3.1. We then compared the performance of the models for both scenarios. The results are presented in Figure 3.12. For all studied regions, we can see that the quality of detecting NO_2 plumes from ships can be improved if only the image patches with the highest total emission proxy are considered. Based on this, we conclude that the dependencies shown in Figure 3.11 are not the results of a particular model configuration, but rather a property of data. However, we can see that the optimization of the hyperparameters of the model did not result in the improvement of the model performance.

3.3.4 Potential improvements in small ship detectability

In this Subsection, we address the third research question of the study. Namely, we investigate whether there is a potential for improvement of detectability of NO_2 plumes from the slow/small ships if more data would be used for the training of the

(a)

(b)
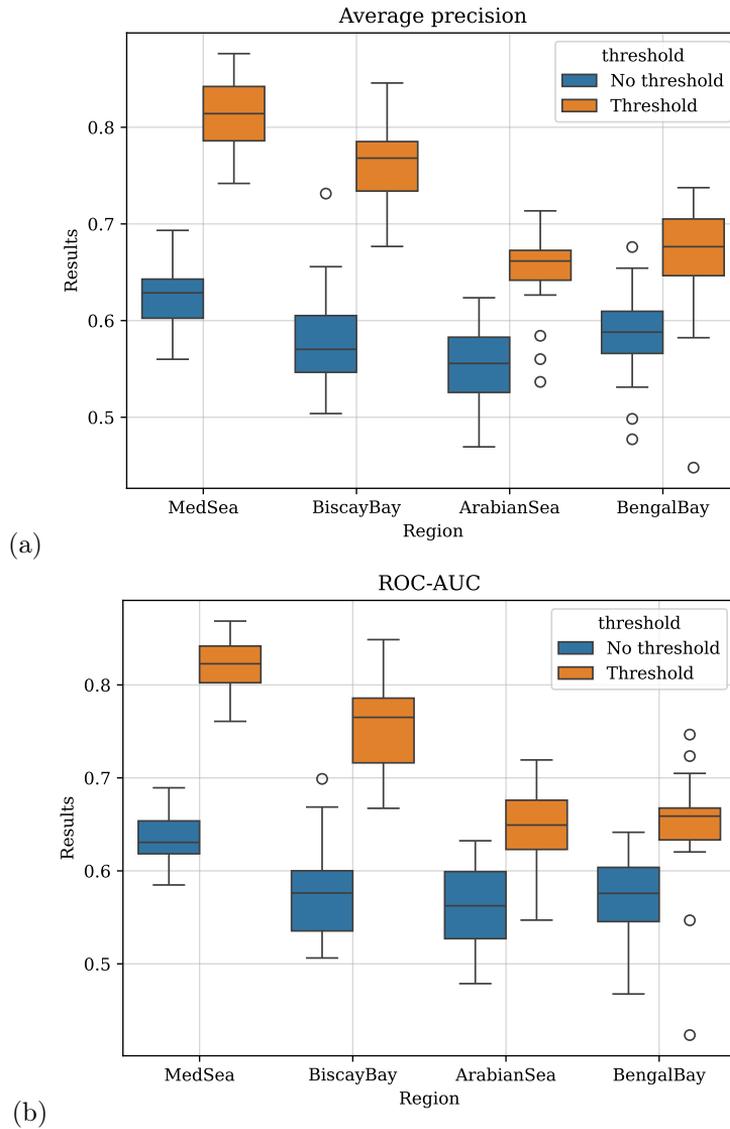
Figure 3.12: Comparison of the performance of the model when all ship images are in the dataset and when only images with the proxy above the predetermined proxy threshold are used.
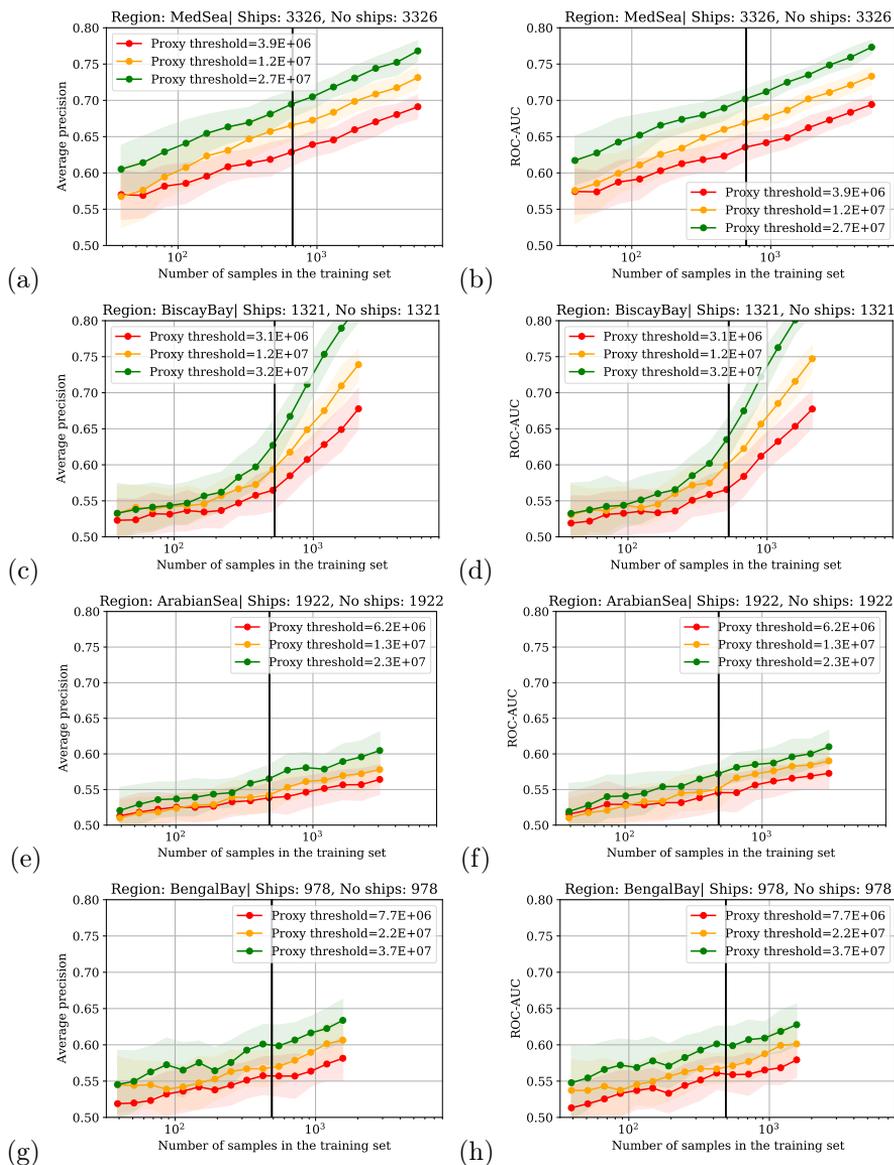
Figure 3.13: Learning curves for different levels of the applied thresholds. The black line indicates the dataset size that was used for the experiments reported in Figures 3.11, 3.12.
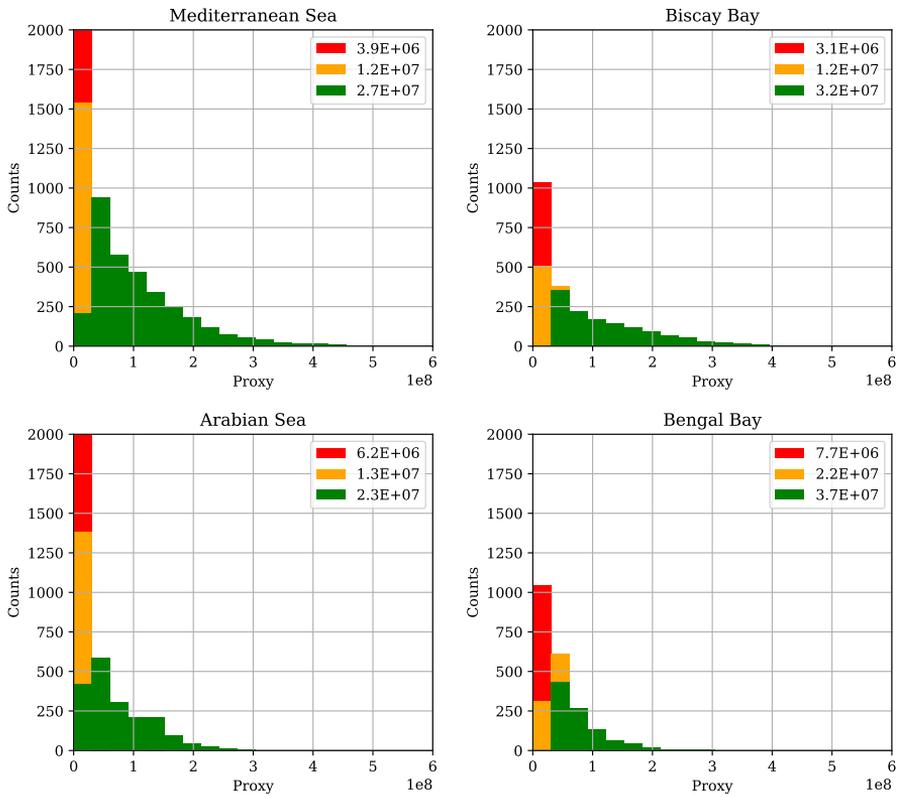
Figure 3.14: Change of the ship proxy distribution after applying thresholds as in Figure 3.13.

classification model. For each region, we selected three proxy thresholding levels and studied the change in the model performance with the growth of the size of the dataset used for the model training. We focus here on the low thresholds. The used thresholds were set as 10%, 30%, and 50% quantiles of the proxy value for the Mediterranean Sea and Biscay Bay. For the Arab Sea and Bengal Bay, the applied thresholds were 10%, 40%, and 60% due to the fact that the model performances on the lowest quantiles were indistinguishable. Similarly to the previous experiment, the maximum size of the dataset was defined by the number of data points in the dataset with the proxy value higher than the highest among the three applied thresholds.

The resulting learning curves for each of the studied regions are presented in Figure 3.13. We can see that for all studied regions, the results shown in Figure 3.11 can be improved by using more data for model training. We also observe that for the regions Biscay Bay and Mediterranean Sea, more data results in a more significant increase in performance, than for the Arabian Sea and Bengal Bay. To explain this, in Figure 3.14, we present the distribution of the variable ship emission *Proxy* for each consecutive threshold applied. The histograms show that for the Biscay Bay and the Mediterranean Sea, there are many more image patches with high values of total emission proxy than for the Arabian Sea and Bengal Bay. As a result, even after removing from the dataset the image patches with the lowest total emission proxy, for such regions as the Arabian and Bengal Bay, the models are still trained on significantly lower total emission proxies than the models for the Biscay Bay and the Mediterranean Sea.

3.4 Discussion

The main objective of this study was to investigate the sensitivity limits of a detection system for NO_2 plumes from seagoing ships using TROPOMI data, considering the speed and length of the ships that we expressed through the means of ship emission proxy. By the detection system, we mean a sequence of steps starting from the signal measurement by the sensor, followed by data retrieval, and finally the application of the developed methodology of automated detection of ship plumes. Each of these steps influences the numbers obtained in this study.

To be able to address the problem of sensitivity estimation, we build a methodology based on machine-learning classification models. This approach allowed us to effectively exploit the TROPOMI signal and contextual information while automatically separating the image patches into those, where the NO_2 plumes can and cannot

be detected. The choice of a multivariate model enabled us to take into account features important for satellite sensitivity, such as wind and satellite/solar viewing angles. Studying several machine learning classifiers of increasing complexity, we found that the XGBoost model yielded the best performance across most regions. This shows the importance of the application of complex machine-learning models for the effective identification of TROPOMI image patches with NO$_2$ plumes from ships with a relatively low number of features.

With the first research question (**RQ1**), we attempted to determine the minimum speed and length of seagoing ships for which the TROPOMI data-based detection system can detect NO$_2$ plumes. We first showed that while the smallest ships considered in our dataset are below the detection limit of the system, once reaching a certain level of ship speed/size, the signal becomes detectable. Second, for the Mediterranean Sea and the Arabian Sea, we estimated sensitivity limits of approximately $1 \times 10^7 m^5/s^3$. For Biscay Bay, the obtained limit lies around $3.8 \times 10^6 m^5/s^3$. Comparing the obtained numbers with the ship emission estimation provided in [41], we can see that our detection system allows us to correctly recognize some plumes with concentrations close to the background concentrations estimated for the Mediterranean Sea. The obtained values of emission proxy translate to the minimum detectable speed of 10 kt and minimum detectable length of 150 m for the Mediterranean and Arabian Seas and 8 kt and 100 m for Biscay Bay. Unfortunately, due to the insufficient amount of data, the sensitivity limits for the Bengal Bay region could not be determined.

With the second research question (**RQ2**), we examined the potential improvement in NO$_2$ plume detectability when considering only the biggest emitters. With our results, we numerically confirmed that restricting the analysis to faster/larger ships leads to enhanced detectability of NO$_2$ plumes. For the Mediterranean Sea region, the performance of the classification model can exceed 0.8 ROC-AUC and average precision scores. This finding suggests concentrating the focus on the larger emitters, could potentially increase the efficiency of the application and accuracy of ship emission monitoring using the TROPOMI instrument. Our analysis also revealed distinct differences in model performance quality between regions. Notably, the Mediterranean Sea and Biscay Bay consistently show better performance compared to the Arabian Sea and Bengal Bay. We can see that these variations could be attributed to variations in ship traffic density between the regions. Additional factors that potentially can influence the performances of the models are measurement conditions (e.g., number of cloudy days), differences in data quality between regions (c.f. Table 3.2), and different scales of temperature fluctuations or concentration of ozone in the background. The

last two factors affect the lifetime of NO_2. However, an in-depth understanding of this problem requires a separate study and we leave it as future work.

Our investigation into the third research question (**RQ3**), regarding the potential for improving NO_2 plume detectability from slow or small ships by utilizing more training data, again showed the variability of the results across the regions. For the Mediterranean Sea and Biscay Bay regions, an increase in data volume led to a notable enhancement in model performance. While, for the Arabian Sea and Bengal Bay, the impact of increased data, even though present, was less pronounced. One of the established reasons was the fact that for European regions we had a higher ratio of data points with a high value of emission proxy in the dataset than for the Bengal Bay and Arabian Sea. Nevertheless, the obtained results indicate that the accuracy of currently determined detection limits is perhaps constrained not by the methodology or the sensor, but by data availability.

Implications and future work

The insights gained from this study have important implications for satellite-based ship emission monitoring. By identifying sensitivity limits and optimal ship characteristics for detectability, our findings guide the scope of future studies on ship's NO_2 estimation using TROPOMI data and give an overview of the potential application of the TROPOMI instrument for ship emission monitoring. Moreover, the obtained results can be used as a benchmark sensitivity level for future satellite missions, such as, for instance, TANGO [67].

In future research, it would be valuable to explore factors beyond ship speed and length that influence detectability, such as temperature regimes, clouds, background ozone concentrations, effect of the sunglint or satellite viewing angle. Moreover, it would be valuable to perform an in-depth study explaining the observed multi-regional differences in ship plume detectability. Finally, studying different types of machine-learning architectures or including more data features in the used datasets can provide additional insights into understanding if the ship plume detectability limits can be lowered further by means of potential improvement information extraction from image patches. A possible candidate is Convolutional Neural Networks (CNN), as it was done in [38] for the detection of visually distinguishable ship NO_2 plumes. However, [63, 64] provide indications that CNN architecture might not be a suitable option for the detection of plumes that are poorly distinguishable on the TROPOMI data.

3.5   Conclusions

In this study, we investigated the sensitivity limits of the TROPOMI data-based detection system with respect to the detection of NO_2 plumes from individual seagoing ships. To the best of our knowledge, no previous research has examined this aspect, making our findings novel and significant in understanding the capabilities of the TROPOMI instrument. Our results are obtained through the analysis of four regions of interest (the Mediterranean Sea, Biscay Bay, Arabian Sea, and Bengal Bay) and can be summarized as follows:

1. We quantified the sensitivity limits of a detection system for NO_2 plumes from seagoing ships using TROPOMI data in terms of the length and speed of a ship beyond which the NO_2 plumes from individual ships cannot be distinguished anymore.

2. We also numerically showed that, as expected, the ships with higher emissions (through either greater length or speed) are more easily detected. We demonstrated such an effect by analyzing model performances with the removal from the dataset ships with the lowest emission proxy. This is agnostic to the model or studied region.

3. Then, we demonstrated that the detection of the NO_2 plumes from the ships with lower emission proxy can be improved, once more training data are added.

4. Finally, we obtained different levels of results between the studied regions. We showed that for different regions a machine learning model not only yields different levels of results but also uses different features as main indicators of the presence of a plume in an image patch. A discrepancy is noticeable when comparing the Arabian Sea and Bengal Bay to the Mediterranean Sea and Biscay Bay.

To sum up, our findings suggest that, while efficient monitoring of seagoing ships from the TROPOMI satellite is possible, the quality of ship plume detectability depends on many factors. We believe that our results provide guidelines for establishing the research scope for future studies on NO_2 ship plume detection as well as contribute to the successful application of satellite-based instruments for the monitoring of NO_2 emission from seagoing ships.

3.5. Conclusions

Chapter 4

Automated assignation of a ship Region of Interest for estimation of NO$_2$ emission from individual ships using satellite data

Abstract As of 2021, more demanding NO_x emission requirements entered into force for newly built ships operating on the North and Baltic Sea. Even though various methods are used to assess the emission from ships in ports and off the coastal areas, monitoring over the open sea has been infeasible until now. In this Chapter, we present an automated method for the evaluation of NO_2 emissions produced by individual seagoing ships. We use the spatial association statistic local Moran's I in order to improve the distinguishability between the plume and the background. Using the Automatic Identification Signal (AIS) data of ship locations as well as incorporated uncertainties in wind speed and wind direction, we present a method for automatic association of the detected plumes with individual ships. We evaluate the quality of ship-plume matching by calculating the Pearson correlation coefficient between the values of a model-based emission proxy and the estimated NO_2 concentrations. For five of the six analyzed areas, our method yields improved results against the baseline approach used in a previous study.

4.1 Introduction

Once the sensitivity limits of the detection system using TROPOMI data are estimated, we focus our attention on using the TROPOMI data for the quantification of NO$_2$ emission from individual ships. In [41], the authors introduced the first attempt to quantify isolated ship plumes that can be identified by visual inspection of daily data. However, the NO$_2$ traces from the majority of the ships in the area are not sufficiently stronger than the NO$_2$ background concentration. As a result, only plumes of larger ships were assessed in that study. In addition, the authors acknowledged that their approach requires multiple manual steps. In order to be able to apply the TROPOMI instrument data for the global and continuous monitoring of the NO$_2$ emissions from seagoing ships, an automated method for ships' NO$_2$ estimation is needed.

In this Chapter, we present a heuristic for automated evaluation of NO$_2$ concentrations resulting from NO$_x$ emissions produced by individual seagoing ships. We start with the enhancement of the contrast between NO$_2$ plumes and the background. We then introduce a method for automated assignation of the Region of Interest (RoI) to a studied ship. Finally, we apply a thresholding method for the separation of ship plumes from the background and estimation of ships' NO$_2$. The presented approach allows the quantification of the ships' emission, even if the produced plume cannot be distinguished visually so that the performance of more and smaller ships can be assessed in a single satellite overpass. The obtained results are benchmarked against the method proposed in [41].

With this study, we address the following research questions:

- **RQ4**: How to assign a TROPOMI signal associated with a certain plume to a potential emitting ship?

- **RQ5**: To what extent can the NO$_2$ plumes be segmented in the TROPOMI data using a simple thresholding method?

The rest of the Chapter is organized as follows: In Section 4.2, we present our methodology. We start with the introduction of the applied image enhancement method in Section 4.2.1. We then present a developed approach to ship-plume assignment in Section 4.2.2. We further explain how we estimate and evaluate the quality of estimation of ship's NO$_2$ in Section 4.2.3. The results of the study are presented in Section 4.3, and conclusions can be found in Section 4.4.

4.2 Method

In this Section, we describe all steps of the proposed approach for automated evaluation of NO_2 emission from individual ships. We start with the introduction of the technique used for the enhancement of the TROPOMI data. We then present a proposed approach for the delineation of the region of interest of the studied ship. Finally, we explain how we evaluate the obtained results of ship NO_2 estimation.

4.2.1 Image enhancement
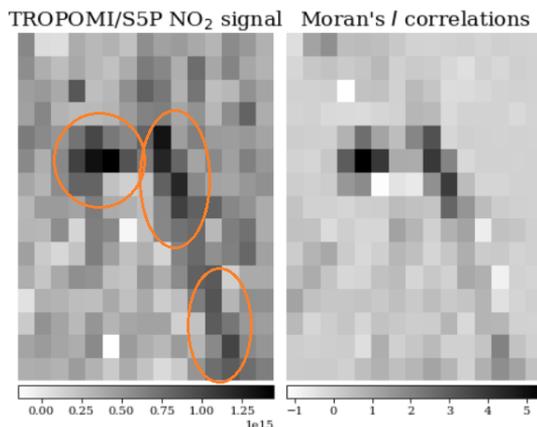
Figure 4.1: Reduction of the background noise as a result of application of the local Moran's I. Orange circles indicate ship plumes that can be distinguished on the plot. NO_2 is in [molec/cm^2].

The first step is an enhancement of the TROPOMI data. To increase the contrast between the ships' plumes and the background, we used spatial association statistic local Moran's I [5] — one of the most often used methods for hot spot detection [78]. The local Moran's I spatial auto-correlation statistic allows the enhancement of the intensity of high-value pixels located in a cluster while suppressing isolated concentration peaks randomly occurring in the background. We characterize a ship plume as a cluster of pixels adjacent to each other with a concentration higher than the background average. This way, calculating the spatial auto-correlation on TROPOMI image, we combine image denoising with the enhancement of the relevant part of the image. Figure 4.1 illustrates an example where the applied image enhancement procedure notably reduced the background noise and increased the contrast between the plume and the background, improving the detectability of the ships' traces.

Formally, the local Moran's I spatial auto-correlation statistic is defined as follows:

$$I_i = \frac{(x_i - \mu)}{\sigma^2} \sum_{j=1, j \neq i}^{N} w_{ij}(x_j - \mu), \tag{4.1}$$

where i is the pixel of an image, x_i is the value of the respective pixel, N is the number of analyzed pixels of a *ship plume image* (in our case 18×18), μ is a mean value of all N pixels, σ^2 their variance, and w_{ij} is the value of an element in a binary spatial contiguity weight matrix W at location j with regards to the analyzed pixel i. The value of an element of the binary spatial contiguity matrix w_{ij} is 1 for pixels that are considered to be the neighbors of the analyzed pixel i, and 0 otherwise. For the study, the queen spatial contiguity [43], which is the 3×3 8-connected neighborhood of the analyzed (central) pixel was applied. The value I_i becomes the value of the corresponding pixel of the resulting enhanced image.

4.2.2 Ship-plume assignment

Determining the spatial correspondence between the TROPOMI signal and the location of the ships is a challenging task. The emitted plume is displaced by the prevailing winds so that observed NO$_2$ concentrations no longer coincide with the tracks of the ships. At the same time, linear transformation of the ship trajectory that is solely based on available wind data (e.g. [41]) might be inaccurate due to the wind-related uncertainties (see Figure 4.2b). Such a method is further used for the benchmarking of our approach. We refer to it as *ship track shift*. To overcome the problem of inaccurate ship-plume matching, we propose to assign to each ship a Region of Interest (RoI), the so-called *ship sector* (Figure 4.2d). The RoI of a ship defines the region within which the produced NO$_2$ is expected based on ship speed, wind speed, direction, and uncertainties.

To define a *ship sector*, we start with estimating the trajectory of the ship – a *ship track* – using AIS ship data, from some time before, until the moment of the satellite overpass (c.f. Figure 4.2a). For each of the analyzed regions, we experimentally determined the optimal time prior to the satellite overpass during which the AIS data was used for plume trace localizing. For the Mediterranean Sea area, this time was equal to 1 hour, and for the Arabian Sea to 40 minutes. The difference in the considered time can be explained by the fact that the time needed for $NO_x \rightarrow NO_2$ transformation is mostly determined by the atmospheric conditions [94], which in the area of the Arabian and Mediterranean Sea are different.

We now move to the introduction of the second step of *ship sector* definition. Following [41], we assume that the plume emitted by a ship has moved in accordance with wind direction by a distance $d = v \times |\Delta t|$, where v is the local wind speed for a coinciding time, and $|\Delta t|$ is a time difference between the time of the satellite overpass and the time of a given AIS ship position. In this way, we obtain a trajectory that we call a *wind-shifted ship track*. An illustration of a *wind-shifted ship track* is depicted in Figure 4.2b. Both wind speed and wind direction are assumed to be constant for the whole time during which we study the plume.

The assumption of constant speed and direction of the wind may create uncertainties in the expected position of the plume of the ship. Therefore, in the third step, we calculate the *extreme wind-shifted tracks*, by adding the margin of wind-related uncertainty to each side of the *wind-shifted ship track* – c.f. Figure 4.2c. The applied values of wind speed and wind direction uncertainties are sub-optimal and are equal to 2 m/s and 20° respectively. The *extreme wind-shifted tracks* define the borders of the RoI of the analyzed ship that we refer to as a *ship sector*.

The radius of the *ship sector* is determined as a maximum distance from the position of the ship at the moment of the satellite overpass to the position of the ship at the earliest moment of the observation (1 hour for Mediterranean Sea, 40 minutes for Arabian Sea) in accordance to *ship track*, *wind-shifted ship track*, or *extreme wind-shifted tracks* (the furthest point is taken into consideration). The *ship sector* delineates the area within which we study the plume produced by the analyzed ship. In Figure 4.2d an example of a resulting RoI that we call a *ship sector* is presented.

4.2.3 NO$_2$ estimation and model performance evaluation

Within each *ship sector*, we separated pixels that were determined as a plume by thresholding the enhanced image. The applied threshold is equal to the 25th percentile of the values of the pixels lying inside the sector. The average value of NO$_2$ pixels determined as a plume held as the evaluation value of NO$_2$ for the corresponding ship and was used for the comparison with the emission proxy defined in Section 2.4.

The quality of ship-plume assignment was reported in terms of the Pearson correlation between the assigned value of NO$_2$ and the emission proxy E_s (see Chapter 2) for all ships in an analyzed area. Note that the used simple model of emission proxy does not reflect all factors that are needed for a precise estimation of ship emission potential; as a result, it is not expected to lead to a perfect correlation with experimentally determined values of NO$_2$.
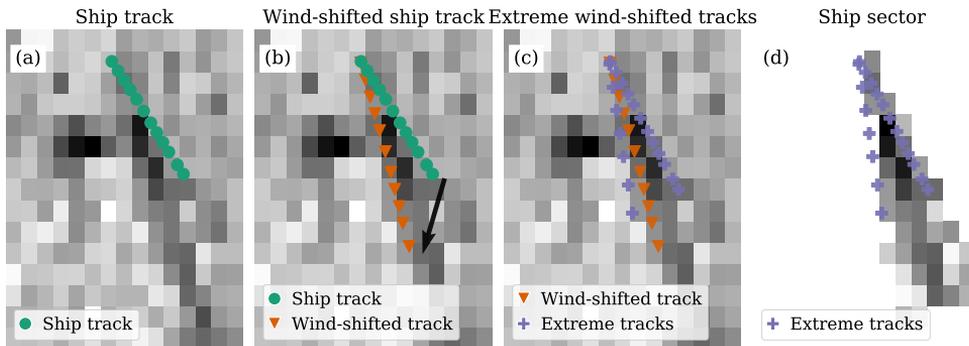
Figure 4.2: Ship sector definition pipeline. Background – the TROPOMI NO$_2$ signal around the analyzed ship. Two ship plumes can be distinguished at the central part of the image, and one, less intense – in the bottom right of the image. Only one is of interest here. (a) *Ship track* – estimated, based on AIS data records. (b) *Wind-shifted ship track* – a ship track shifted in accordance with the speed and direction of the wind. It indicates the expected position of the ship plume. A black arrow indicates the wind direction. (c) *Extreme wind-shifted ship tracks* – calculated, based on wind information with assumed uncertainties; define the borders of the *ship sector*. (d) A resulting *ship sector* – an ROI of an analyzed ship.
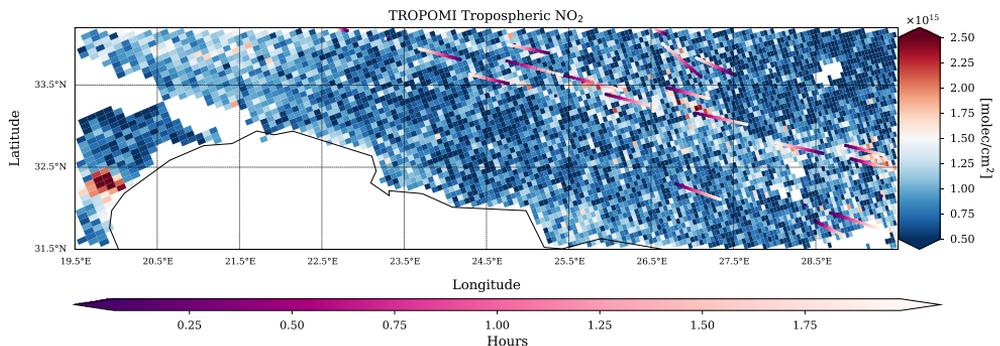
Figure 4.3: The NO$_2$ tropospheric vertical column density. Region: Mediterranean Sea, restricted by the Northern coasts of Libya and Egypt from the south and the South coast of Crete from the north. Date: April 2nd, 2019. Magenta lines indicate ships' tracks based on information from AIS data. The native local size of the TROPOMI pixels is presented in the Figure.

Date	Region	Long min	Long max	Lat min	Lat max
02-04-19	Mediterranean	21.5	29.5	32	34.5
07-08-19	Mediterranean	19.5	27	33	35
27-09-19	Mediterranean	20	24.74	32.5	35
11-04-20	Arabian	64	69	14	18
01-05-20	Arabian	64	71	14	18
03-05-20	Arabian	64	70	12	18

Table 4.1: The boundaries and dates of observation of six analyzed scenes in the Mediterranean and the Arabian Sea.

4.2.4 Data selection

For the experiments, we chose six days with suitable weather conditions (wind < 6 m/s, low level of background pollution) in the Arabian and Mediterranean Sea (see Table 4.2). An example of an analyzed area can be found in Figure 4.3. The boundaries of the analyzed areas are provided in Table 4.1. The differences in the coordinates of the boundary boxes can be explained by shifts in the TROPOMI orbit coverages. To ensure the high quality of the satellite signal, the following filtering criteria were applied to TROPOMI data: $qa\_value > 0.5$, $cloud\ fraction < 0.5$. For a detailed description of the variables see [93]. Some pre-processing steps were applied to the AIS ship data. First, only ships with an overall length exceeding 150 m and speed above 12 kt[1] were taken into account. As was shown in the previous Chapter, the emission levels of smaller or slower ships are expected to be below the detection level of the satellite. Second, ships of which the location was known for less than 25 minutes prior to the TROPOMI overpass were excluded from the study. Finally, if the plume of a ship undergoes intersection with any other plume of sufficient concentration, the ship was excluded from the study.

4.3 Results

The results of the experiments are summarized in Table 4.2. We compared the results obtained with our approach with the baseline method of ship track shift [41]. For five out of the six analyzed regions, the method proposed in this study led to an improvement in the quality of the linear correlation between the estimated NO_2 values and the emission proxy of the corresponding ship. The weighted mean value of the

---

[1]kt - knot, a unit of speed equal to a nautical mile per hour. 12 kt $\approx$ 6.2 m/s.
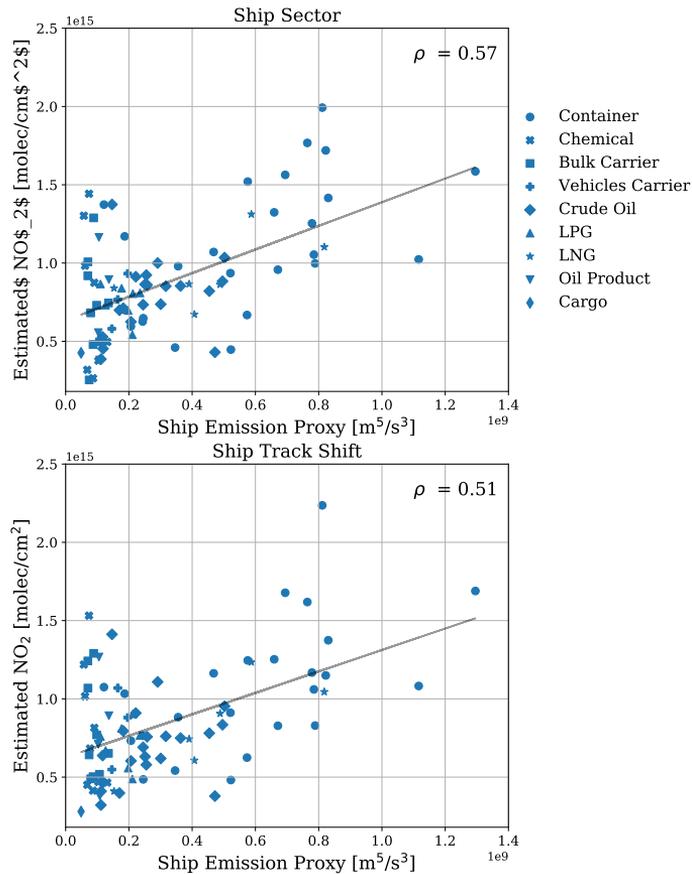
Figure 4.4: Aggregated correlation between the assigned NO$_2$ values and the corresponding values of the emission proxy for the method proposed in this study (top) and the baseline approach (bottom). Markers indicate the types of ships.

Date	Region	Ship sector	Ship track shift	Detected	Undetected
02-04-19	Med.Sea	**0.81**	0.24	8	1
07-08-19	Med.Sea	**0.61**	0.60	17	1
27-09-19	Med.Sea	**0.78**	0.77	9	1
11-04-20	Arabian	**0.83**	0.67	15	1
01-05-20	Arabian	0.65	**0.71**	20	0
03-05-20	Arabian	**0.80**	0.63	16	2
	AVG/SUM:	**0.73**	0.63	85	6
	STD:	**0.09**	0.13		

Table 4.2: The Pearson correlation coefficients of estimated NO_2 with the corresponding value of the ship's emission proxy for each of the analyzed scenes. The **ship sector** column presents the results achieved with the method introduced in this study. The **ship track shift** column shows the results obtained with the baseline method [41].

Pearson coefficient (the weights are the number of ships analyzed within each area) increased from 0.63 to 0.73, whereas the weighted standard deviation decreased from 0.13 to 0.09. Finally, we calculated an aggregated correlation of NO_2 estimations for all six days analyzed with the associated emission proxies. The results are presented in Figure 4.4. This experiment allowed us to assess the generalization properties of the proposed approach. We can see that the level of achieved correlation decreased for both methods. This was expected due to the presence of smaller ships, whose detectability is highly dependent on the instrument's sensitivity at a given moment of time. Nevertheless, the approach proposed in our study still assures higher quality of the results.

4.4 Conclusions

In this Chapter, we introduced the first approach for the automated evaluation of NO_x emissions of individual seagoing ships with TROPOMI data. We applied a spatial association statistic, local Moran's I, to enhance the separability between the ships' plumes and the background of comparable concentration. We then proposed our method for automated assignation of the RoI to the analyzed ship. We call the proposed RoI a *ship sector*. It is defined based on the uncertainties in the speed and direction of the wind data. We assume that the emission plume produced by a studied ship will be located within the assigned *ship sector*. With this, we addressed the **RQ4** of the thesis: How to assign a TROPOMI signal associated with a certain plume to a

potential emitting ship? We then performed a separation of plume-related pixels from the background and estimated the amount of NO$_2$ associated with a given ship. Since this is the first attempt of NO$_2$ estimation from individual ships using satellite data, the plume of the ship within the *ship sector* was segmented using the local threshold (**RQ5**). To validate the proposed approach, we used a theoretically derived ship emission proxy. We compared the results obtained using our method with the approach proposed in a previous study. The comparison showed that our method leads to the increment of the linear correlation between estimated values of NO$_2$ and model-based emission proxy for five of the six analyzed areas, as well as for an aggregated experiment, where the Pearson correlation was calculated for all six analyzed areas at once. We, however, see that the achieved levels of Pearson correlations cannot be characterized as very high. Partially, this can be explained by the fact that the emission proxy used for the validation of the approach does not account for all factors influencing the emission level of the ship. Nevertheless, this also suggests the need for a more complex method of ship plume segmentation.

To conclude, the proposed approach assures more precise quantification of local NO$_2$ concentrations caused by NO$_x$ emissions of individual ships than the previous (and the only existing study). Moreover, the method introduced in this study does not require any manual steps which is a significant improvement over the current state-of-the-art. As a future research direction, we would like to propose the application of more complex techniques (e.g. machine learning) for the task of ship plume segmentation within the presented here ship's RoI – *ship sector*. Such a methodology will be discussed in the next chapter.

4.4. Conclusions

Chapter 5

Ship plume segmentation with supervised machine learning

Abstract To deploy a remote sensing-based global emission monitoring system, an automated procedure for the estimation of NO_2 emissions from individual ships is needed. The extremely low signal-to-noise ratio of the available data as well as the absence of ground truth makes the task very challenging. Here, we present a methodology for the automated segmentation of NO_2 plumes produced by seagoing ships using supervised machine learning on TROPOMI/S5P data. We show that the proposed approach leads to a more than a 20% increase in the average precision score in comparison to the methods used in previous studies and results in a high correlation of 0.834 with the theoretically derived ship emission proxy. This work is a crucial step toward the development of an automated procedure for global ship emission monitoring using remote sensing data.

5.1   Introduction

In the previous Chapter, we have introduced the method for efficient assignation of a RoI to a studied ship. However, the segmentation of the ship plume within the assigned RoI was performed on the basis of a local threshold. This simple method provides the first baseline for the task but has a list of disadvantages. Namely, the threshold was established on the basis of the only variable (NO_2 concentration). It also assumes the linear separability between the signal coming from the plume and the background. All this results in insufficient flexibility of the method and consequent low quality of ship-plume segmentation.

In this Chapter, focus our attention on the development of a method for efficient segmentation of ship plumes. Among the main challenges of this task are low temporal sample rate and spatial resolution resulting in an extremely low signal-to-noise ratio. In addition, there is a high risk of interference of the ship plume with other NO_x sources and a high frequency of occurrence of plume-like objects that cannot be associated with any ship. Finally, the ground truth for this task is not available. To increase the number of potentially distinguishable plumes, we enhance the contrast between the ship plumes and the background. In order to overcome the above-mentioned challenges, we present a methodology that allows addressing the problem of automated ship plume segmentation with supervised machine learning. The developed method of feature engineering allows for the application of multivariate machine learning models. This, in turn, allows us to account for multiple factors that help differentiate a plume produced by a ship of interest from all the other plumes in the ship's neighborhood, circumventing the listed limitations.

With the aim to increase the number of potentially distinguishable plumes, we enhance the contrast between the ship plumes and the background. The used enhancement technique allows for a differentiation between the ship plumes and random co-occurring concentration peaks in the ships' neighborhood. The application of the image enhancement technique also allows for an improvement of the low signal-to-noise ratio. Then, to focus the area of analysis on the region where the ship plume is expected to be located, we use the presented in Chapter 4 concept of ship's RoI – the *ship sector*. Subsequently, we normalize the *ship sector* and divide it into sub-regions. This way, we distinguish the plume of interest from all the other NO_2 plumes or land-origin outflows that potentially might be located within the *ship sector*. Based on the *ship sector* division, we create a set of spatial features that characterize the location of the NO_2 plume within the *ship sector*. Due to the absence of other sources

of ground truth, each pixel of the *ship sectors* we manually label as a "plume" or "not a plume". Trained on the manually labeled data, a machine learning model will enable us to automatically segment plumes in unseen images. We study five robust machine learning models of increasing complexity and compare their performance with the threshold-based methods used in previous studies. To validate the developed pipeline, we compare the estimated based on the result of segmentation amount of NO_2 to the theoretically derived ship emission proxy [41].

In this Chapter, we address the following research questions:

- **RQ6**: Can we improve the segmentation quality of NO_2 plumes from individual ships using supervised machine learning?

- **RQ7**: Does the machine learning-based segmentation allow for the detection of NO_2 plumes that cannot be recognized visually?

The rest of this Chapter is organized as follows: In Section 5.2.1, we start with an explanation of data selection and data preparation steps. We then provide a description of the developed methodology in Section 5.2.2. In Section 5.3, the reader can find the results of the study, which are followed by the conclusions in Section 5.4 and discussion in Section 5.5.

5.2 Materials and methods

5.2.1 Data preparation

In this Section, we explain the steps of data selection and preparation that were performed in the process of the preparation of the dataset used in this study. First, to generate images of regular size, we regridded[1] the original TROPOMI data into a regular-size grid of size of $0.045° \times 0.045°$, which for the pixel in the middle of the analyzed area translates to approximately 4.2×5 km^2. To assure the good quality of the used TROPOMI measurements, we applied the following filtering criteria to TROPOMI data: $qa\_value > 0.5$, $cloud\ fraction < 0.5$. In Chapter 3, we showed that such filtering criteria assure a good trade-off between data quality and data availability.

In this study, we analyzed 68 days of TROPOMI measurements from the period between 1 April 2019 and 31 December 2019. The analyzed days were mostly sunny – the distribution of the variable *cloud fraction* for the scope of this study is provided in

---

[1]For the data regridding HARP v.1.13 Python package was used.

Distribution of the variable from the dataset
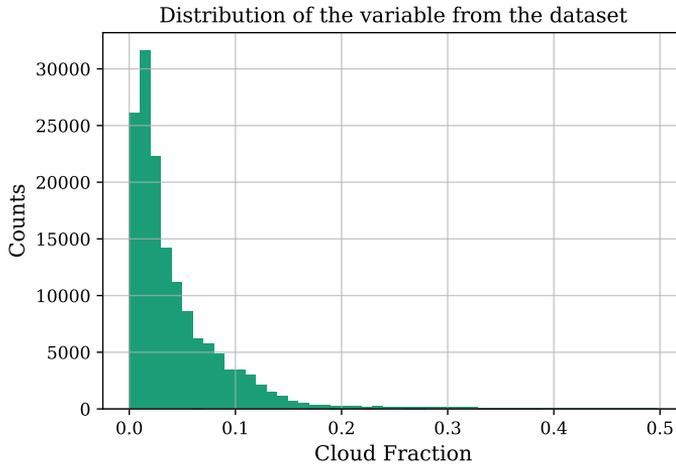
Figure 5.1: Distribution of the variable *cloud fraction* for the dataset used in this study.

Figure 5.1. The studied data product: tropospheric vertical column of nitrogen dioxide [31]. Data version: 1.3.0. For the analysis, we chose an area in the Mediterranean Sea, similar to the one studied in Chapter 4 (for an area outline c.f Figure 4.3). The area is restricted by the Northern coasts of Libya and Egypt from the south and South coast of Crete from the north[2]. Apart from the fact that it was already studied in the previous studies, this region was selected because of the presence of a busy shipping lane connecting Europe and Asia, the high frequency of occurrence of sunny days, and relatively low levels of NO_2 background concentrations, which are favorable conditions for the analysis.

With the aim of reducing the number of images where the ship plume cannot be visually detected, in our study, we only focus on ships with a speed that exceeds 14 kt. If two ships move in immediate proximity to each other, only the ship with the highest speed was taken into consideration. From the analysis were also excluded ships that are not involved in global trade, such as Yachts, Leisure Vessels, or Research Vessels. In Figure 5.2, the information about the dates used for this study as well as the number of ships per day studied is depicted. The differences in the number of ships per studied day can be caused by bad weather conditions on the measurement day.

---

[2]lon: [19.5°; 29.5°], lat: [31.5°; 34.2°].
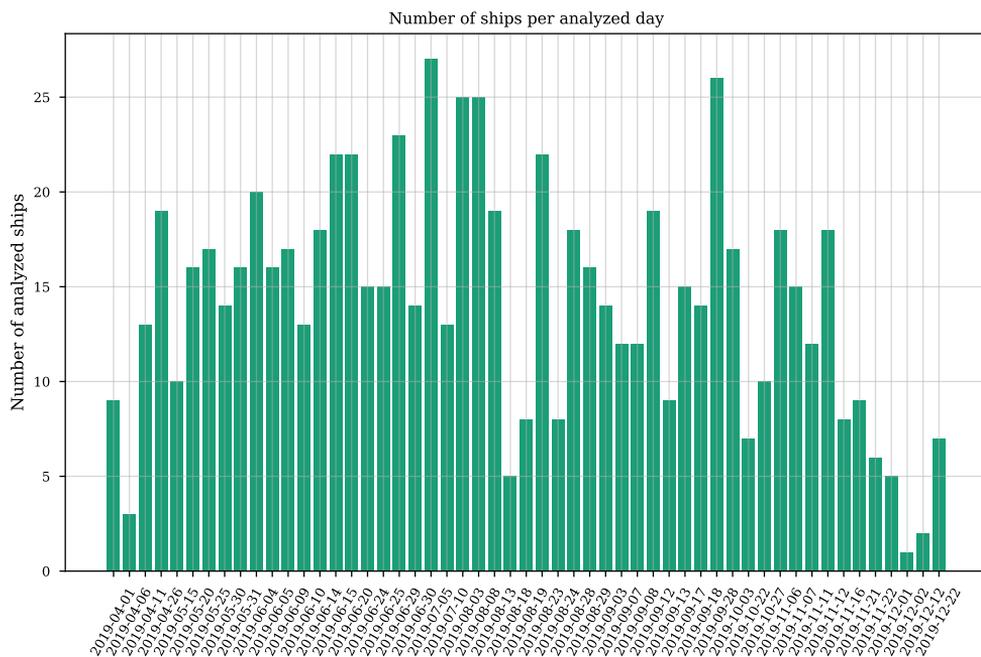
Figure 5.2: A list of days used for the dataset creation and the number of ships per day studied.
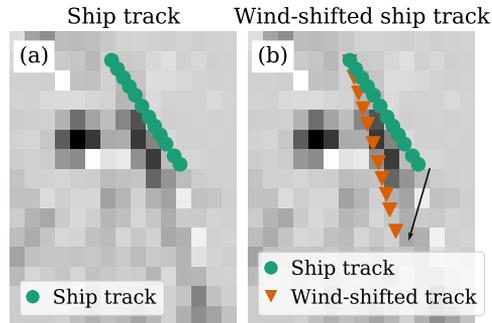
Figure 5.3: Both panels: *ship plume image* with indicated ship tracks. Panel a): *Ship track* – estimated, based on AIS data records. The ship track is shown for the time period starting from 2 hours before the satellite overpass until the moment of the satellite overpass. Panel b): *Wind-shifted ship track* – a ship track shifted in accordance with the speed and direction of the wind. The *wind-shifted ship track* indicates the expected position of the ship plume. A black arrow indicates the wind direction. For both presented images, the size of the pixel is equal to 4.2×5 km^2

5.2.2 Method

In this Subsection, we present the developed methodology. Taking advantage of the characteristics of the analyzed ship as well as wind conditions in the studied region, our approach allows the segmentation of NO$_2$ plume produced by the particular ship of interest distinguishing it from all the other concentration peaks in the surrounding area. The results produced by the proposed approach are easily interpretable and thus can be used as a reliable source of information by ship inspectors.

The method is built upon the concepts introduced in Chapter 4. Therefore, with the aim of not repeating ourselves, for some definitions, the reader will be referred to the above-mentioned chapter. The method presented in this Chapter consists of the following steps: definition and enhancement of a ship plume image, definition of a ship sector that allows the further restriction of the analyzed area, normalization of the defined ship sector, and split of the normalized sector into sub-regions that, finally, give the possibility to retrieve the set of necessary features. These steps are described below.

Ship plume image definition and enhancement

As a first step of our method, we define an area within the immediate proximity of an analyzed ship. We call it a *ship plume image*. For this, we utilize the knowledge of a ship's position summarized in its *ship track* and *wind-shifted ship track*, as defined in Chapter 4, Section 4.2.2. In Figure 5.3, an illustration of *ship plume image* with indicated *ship track* and *wind-shifted ship track* is presented. Based on *wind-shifted ship track*, the area of the *ship plume image* is determined as follows: the average coordinate of the studied *wind-shifted ship track* defines the center $(longitude_{centr}, latitude_{centr})$ of the *ship plume image*, the borders of the image are defined as $longitude_{centr}, latitude_{centr} \pm 0.4°$[3]. This particular size of a *ship plume image* was determined in order to allow for optimal plume coverage for the most typical range of ship speeds (14kt - 20 kt)[4]. Given the size of the pixel grid, such an offset results in an image of a maximum dimension of 18×18 pixels.

To improve the quality of the TROPOMI data, in the data pre-processing step, on each of the analyzed *ship plume images* we apply spatial auto-correlation statistic local Moran's I [5]. The formal introduction of the method the reader can find in Chapter 4. There we showed that the application of this technique substantially improves separability between the ship plume and the background.

Ship sector

Parameter	Value
Trace track duration	2 hours
Wind speed uncertainty	5 m/s
Wind direction uncertainty	40°

Table 5.1: Parameters applied for ship sector definition.

A plume produced by a ship at a given moment will be displaced, over time, in the direction of the wind in the analyzed area. Having the wind information available, we restrict the analysis to the part of the *ship plume image*, where the probability of finding the plume of the ship is the highest. We perform the area restriction by defining the RoI of an analyzed ship – a *ship sector*, defined in accordance with the description provided in Section 4.2.2, Chapter 4. By defining a *ship sector*, we assume

---

[3]For the area in Mediterranean Sea, in horizontal direction 0.4° ≈ 37.4 km, in vertical direction 0.4° ≈ 44.2 km.

[4]kt - knot, a unit of speed equal to a nautical mile per hour. 14 kt ≈ 26 km/h. 20 kt ≈ 37 km/h.

that the plume produced by a studied ship will lie within the *ship sector* boundaries. Only pixels lying within the *ship sector* are taken into consideration in further analysis. Parameters related to the *ship sector* definition can be found in Table 5.1.

Feature engineering

In order to obtain a multivariate description of the *ship sector* pixels, we encode the spatial information into a set of generic features. First, we perform a *ship sector* normalization to make spatial information in the sector comparable between the different sectors. We define a *normalized sector* by standardization of the orientation and the scale of the original *ship sector*. In this way, the position of the plume within the *ship sector* becomes invariant to the heading (direction) and speed of the ship, as well as to the direction and speed of the wind.

We standardize the orientation of a *ship sector* by rotating to 320° (This particular value of sector rotation angle was chosen for the convenience of visualization and has no influence on further modeling) so that the angle of the polar coordinate of the corresponding *wind-shifted ship track* is the same for all ships (see Figure 5.4). Assuming S is a set of ship sectors in the dataset, formally, the rotation coordinates of a *ship sector* are defined in the following way:

$$\forall s \in S, \quad \forall i \in s: \quad lon\_rot_{s,i} = r_{s,i} \cdot cos(\alpha_{s,i} + \Theta_s), \quad lat\_rot_{s,i} = r_{s,i} \cdot sin(\alpha_{s,i} + \Theta_s), \tag{5.1}$$

where lon_rot_i and lat_rot_i are the polar coordinates of the pixel i within the rotated *ship sector*, $r_{s,i}$ is the radial distance of the pixel i from the origin of the *ship sector* s (in our case, sector origin corresponds to the position of the ship at the moment of satellite overpass), $\alpha_{s,i}$ is a counterclockwise rotation angle of the pixel i from the axis x (*longitude*) of the *ship sector* s, $\Theta_s = \beta - \alpha_s$ is a counterclockwise rotation angle that will be applied for the orientation change of each pixel i of the *ship sector* s, α_s is a rotation angle of a *ship sector* s that corresponds to the counterclockwise rotation angle of the pixel $i_{s,max}$ with the radial distance from the origin $r_{s,max} = max(r_s)$, $\beta = 320°$ is a new rotation angle of each *ship sector* s after the rotation.

We standardize the *ship sector's* scale so that the horizontal and vertical coordinates of the rotated *ship sector* are rescaled into the range [0, 1] by applying a min-max
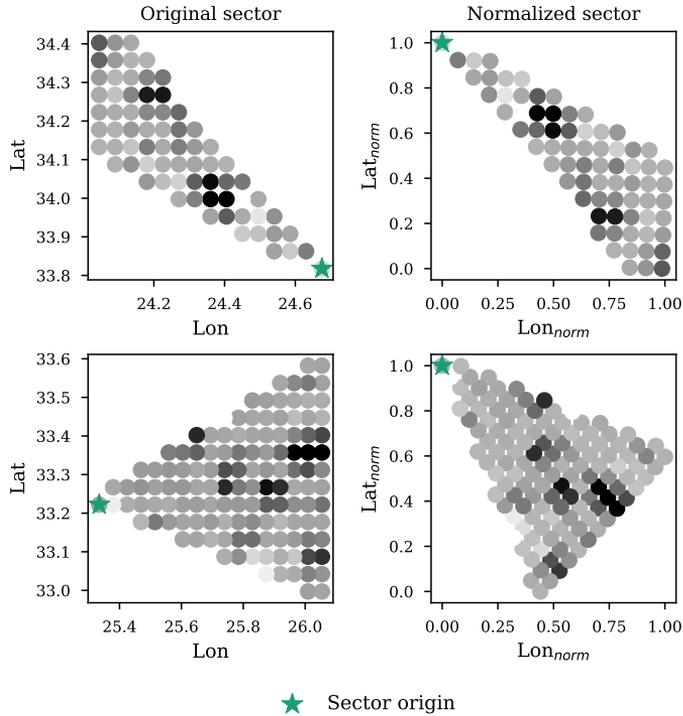
Figure 5.4: Sector normalization. We rotate the *ship sectors* so that all resulting sectors have the same orientation equal to 320° independently of the original direction of the ship's heading. We then rescale the image so that the range of both coordinates is between 0 and 1. The gray area in each figure indicates a *ship sector*. The *ship sector* origin indicator shows the position of the ship at the moment of the satellite overpass. Two examples of original and rotated sectors are shown: one in the top row, and one in the bottom row.

scaler on the horizontal and vertical coordinates of the pixel:

$$lon\_norm = \frac{lon\_rot - min(lon\_rot)}{max(lon\_rot) - min(lon\_rot)},$$

$$lat\_norm = \frac{lat\_rot - min(lat\_rot)}{max(lat\_rot) - min(lat\_rot)}$$

(5.2)

The second step of the feature construction procedure is the division of the *normalized sector* into a set of sub-regions that enable encoding spatial information of the pixels within the *normalized sector*. First, we define *levels* of the *normalized sector* by splitting it into six sub-regions on the basis of the radial distance of the pixel from
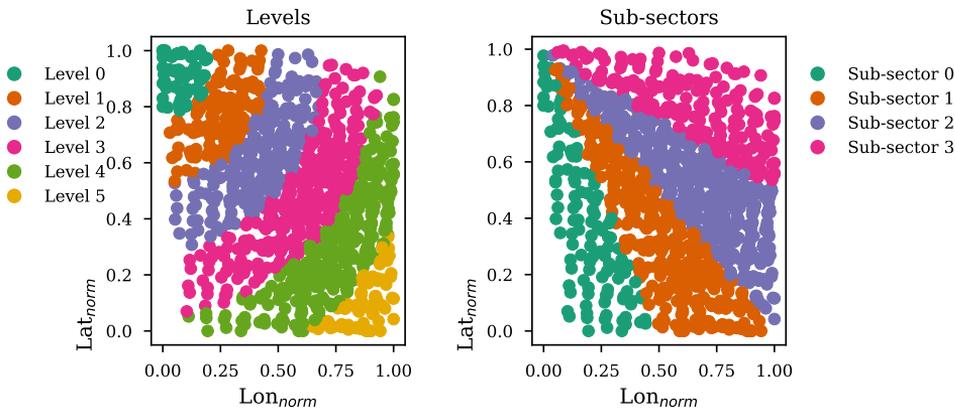
Figure 5.5: Levels and sub-sectors. We perform a feature construction by dividing the *normalized sector* into sub-regions: *levels* and *sub-sectors*. For the convenience of visualization, data points from one day of analysis were used for the preparation of the figure.

the origin of the sector. Then, we define *sub-sectors* by splitting the *normalized sector* into four sub-regions on the basis of the pixel's rotation angle. As a result, the position of each pixel within the *normalized sector* image can be characterized in terms of two values: a *level* and a *sub-sector*. An illustration of the *normalized sector* divided into a set of *levels* and *sub-sectors* is presented in Figure 5.5.

5.2.3   Experiment design

Here, we describe the experimental setup used in this study: first, we describe the dataset used for the training of the multivariate models. Then we explain the models used for the benchmarking, provide a list of used multivariate classifiers, and describe the methods used for hyperparameters optimization.

Dataset composition

Following the steps provided in the previous subsections, we created 754 images and cropped them to an area of the *ship sector*. The *ship sector* images were enhanced by Moran's I operator and manually labeled so that they can be used for training machine-learning models. Not all *ship sector* images contained a visually identifiable NO_2 plume. Moreover, due to the dispersion and chemical transformation of a ship plume, some parts of the plume will always be under the detection limit of

Class-wise data distribution
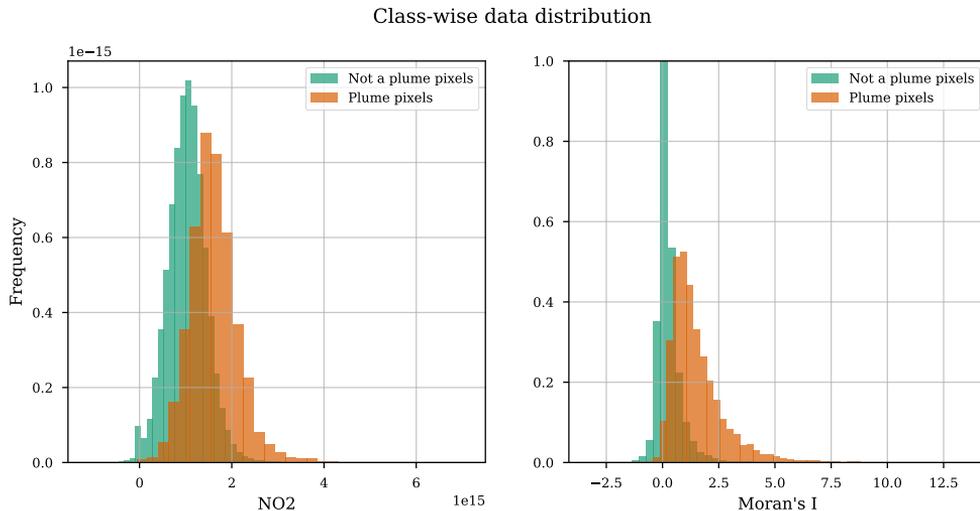
Figure 5.6: Class-wise distribution of the two main features of the dataset: NO_2 and Moran's I.

	No plume	Plume
Number of pixels	68646	6980
Number of images	208	535

Table 5.2: A number of measurement points per class in the dataset.

the TROPOMI instrument and therefore, indistinguishable. Thus, labeling errors are possible. To minimize the chance of mistakes the labeler was supported with several representations of the area of interest: the original not enhanced NO_2 tropospheric vertical columns for the area of a *ship plume image*, the enhanced with the Moran's I area of a *ship plume image*, and NO_2 tropospheric vertical columns for the full studied area in Mediterranean Sea with the positions of the neighboring ships. The descriptive statistics of the resulting dataset are provided in Figure 5.6. In Table 5.2, the information on the data distribution within the two classes of the dataset is shown. All mentioned numbers correspond to the full dataset before the training/test set division.

Multivariate models

To exploit the potential of multivariate modeling, we used several classifiers of increasing complexity: Logistic Regression, Support Vector Machines with linear kernel

[34], Support Vector Machines with radial basis kernel [21], Random Forest[5][14], and Extreme Gradient Boosting (XGBoost)[6][22]. The above-mentioned models are multivariate and thus are able to benefit from the set of prepared features. Namely, the set of spatial features developed with the method is described in Subsection 5.2.2, along with ship and wind-related features. All models selected for the experiment are highly robust. Therefore, the potential mistakes in human labeling, if present in reasonable amounts, should still allow for models' proper training.

The first feature of the model is enhanced by *Moran's I* values of the pixels that were translated into a one-dimensional feature vector. As can be inferred from the definition of Moran's I statistic (see Equation 4.1), the application of Moran's I may result in the creation of additional high-value pixels resulting from the enhancement of clusters of low-value pixels. To mitigate the negative impact of this side effect, apart from the Moran's I, the feature set was composed of the corresponding value of *NO_2*. This way, a supervised learning model will be able to differentiate between high and low-value enhanced NO_2 clusters. Other features used by the model are *Wind Speed*, *Wind Direction*[7], *Ship Speed*, and *Ship Length*. Finally, the position of an analyzed pixel within the *normalized sector* in terms of *levels* and *sub-sectors* was translated into the feature vectors using one-hot encoding. The resulting feature set was composed of 17 features in total. For the full feature list, see Figure 5.10. The used binary label indicates whether the given pixel is a part of the ship plume or not.

For the model fine-tuning and model performance evaluation, a 5-fold nested cross-validation [96, 18] with randomized search [10] was used. The *average precision* score was used as a target function for optimization.

Benchmarks

To quantify the performance improvement gained by the usage of multivariate supervised models, we performed ship plume segmentation by applying a thresholding method on a single selected feature. First, we applied a thresholding method on the tropospheric vertical column of NO_2 TROPOMI product regridded in accordance to the description in Section 5.2.1. No image enhancement technique was applied. This simplest way of plume-background separation was used, among the others, in [85] for the quantification of NO_2 emission from the international shipping sector. In [41], the separation of pixels related to NO_2 plumes from individual ships was also performed

---

[5]All above-mentioned models were implemented in Scikit-learn v. 0.24.2 package [80].

[6]Implemented in xgboost Python package v. 1.3.3.

[7]Wind Direction feature vector was encoded into its sine and cosine components, in order to enable a continuous feature space for various wind directions.
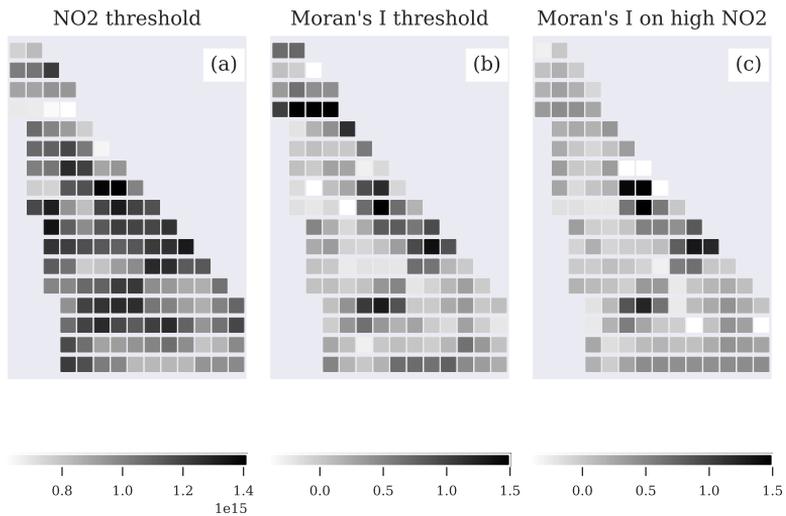
Figure 5.7: Input data example for univariate threshold-based benchmarks. a) Input data for a benchmark method *NO$_2$ threshold*. b) Input data for a benchmark method *Moran's I threshold*. At the top of the *ship sector* the reader can find an example when a cluster of low value NO$_2$ was mistakenly enhanced by Moran's *I*. c) Input data for a benchmark method *Moran's I on high NO$_2$*. For all presented images, the size of the pixel is equal to 4.2×5 km2.

based on solely TROPOMI NO_2 data. In this Chapter, we will refer to this bench-marking method *NO_2 threshold*. Visualization of the input data for this thresholding technique can be found in Figure 5.7(a).

As a second benchmarking method, following the suggestion made in Chapter 4, we performed a ship plume segmentation based on images enhanced with Moran's *I* statistic. The TROPOMI image enhancement allows effective separation of a greater amount of NO_2 plumes. However, as it can be inferred from the definition of Moran's *I* statistic (c.f. 4.1), the application of Moran's *I* statistic may result in the enhancement of low-value clusters that are not part of a plume. Visualization of the input data for this benchmarking technique is presented in Figure 5.7(b). In the rest of the article, we call this method *Moran's I threshold*.

To overcome the problem of enhancement of low-value clusters by Moran's *I*, we propose to assign the value 0 to all pixels of the image with intensity lower than the median of the given *ship sector* picture, and afterward apply the Moran's *I* enhance-ment. This is the third benchmarking method used in this study. We call it *Moran's I on high NO_2*. Visualization of the results of the application of Moran's *I* only on high NO_2 values can be found in Figure 5.7(c). As presented in Figure 5.7, for all three benchmarking methods only pixels that lie within the *ship sector* area were taken into account for segmentation.

NO_2 validation metrics

So far, we have been measuring models' performance based on manually created labels. To evaluate the uncertainty hidden in human labeling, a reference value is required. Due to the fact that there are no on-site emission measurements available at the scale of this analysis, it is therefore necessary to use a ship emission proxy to represent the reference value. Similarly, as in previous chapters, we use a theoretically derived NO_x emission proxy E_s as defined in 2.4.

The ship emission proxy is calculated for each ship of the test sets. We compare the obtained values of emission proxy with the estimated on the basis of segmentation results amount of produced NO_2. We estimate the amount of produced NO_2 by summing up NO_2 concentration within the pixels classified as a "plume" by each of the studied models. For the comparison between the emission proxy and the estimated amount of NO_2, Pearson linear correlation was used.

Model	AP	ROC-AUC
Linear SVM	0.609±0.063	0.935±0.009
Logistic	0.610±0.064	0.936±0.010
RBF SVM	0.742±0.031	0.951±0.008
Random Forest	0.743±0.030	0.952±0.008
XGBoost	**0.745±0.030**	**0.953±0.007**
NO_2 threshold	0.375±0.062	0.823±0.017
Moran's I threshold	0.493±0.063	0.912±0.011
Moran's I on high NO_2	0.607±0.056	0.922±0.010

Table 5.3: Results on the test set with 5-fold cross-validation. Bold font indicates the best-obtained result. Under the dashed line: results obtained from univariate threshold-based methods that, in this study, we considered as benchmarks.

5.3 Results

In this Section, we present the results of our study. We begin with the presentation of the results of the plume segmentation model in Subsection 5.3.1. Appropriate segmentation quality is necessary for a correct estimation of NO_2 produced by ships. In Subsection 5.3.2, we validate the concept presented in this Chapter. In the Subsection, we compare the obtained on the basis of segmentation model results of ship NO_2 estimation with the theoretical ship emission proxy.

5.3.1 Plume segmentation

In Table 5.3, we report the results of the pixel classification based on a 5-fold cross-validation for all models and benchmarks studied. Figure 5.8 provides the corresponding precision-recall curves, obtained by averaging the scores over all cross-validation test sets. In Figure 5.10, we visualize the model coefficients for the linear models studied, as well as the impurity-based feature importance coefficients for the tree-based models (Random Forest and XGBoost). The obtained results can be summarized as follows:

(i) From Table 5.3, Figure 5.8, as well as Figure 5.9 we can conclude that nonlinear classifiers clearly outperform both linear classifiers and threshold-based univariate benchmarks. Both used measures: AP score and ROC-AUC resulted in a similar rank of the studied classifiers. With XGBoost, Random Forest, or RBF SVM models, a very high level of precision can be achieved. For the task of ship plume segmentation, our biggest interest lies in the correct segmentation of the most representative pixels of the

Figure 5.8: Precision-recall curve based on 5-fold cross-validation. Dashed lines indicate the results obtained from univariate threshold-based methods that, in this study, we considered as benchmarks.

Receiver Operating Characteristic (ROC) Curve

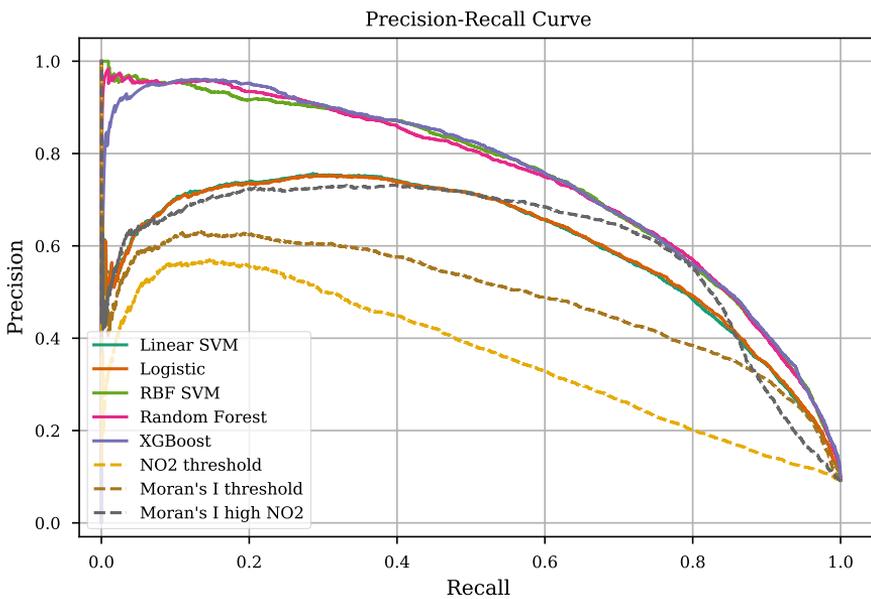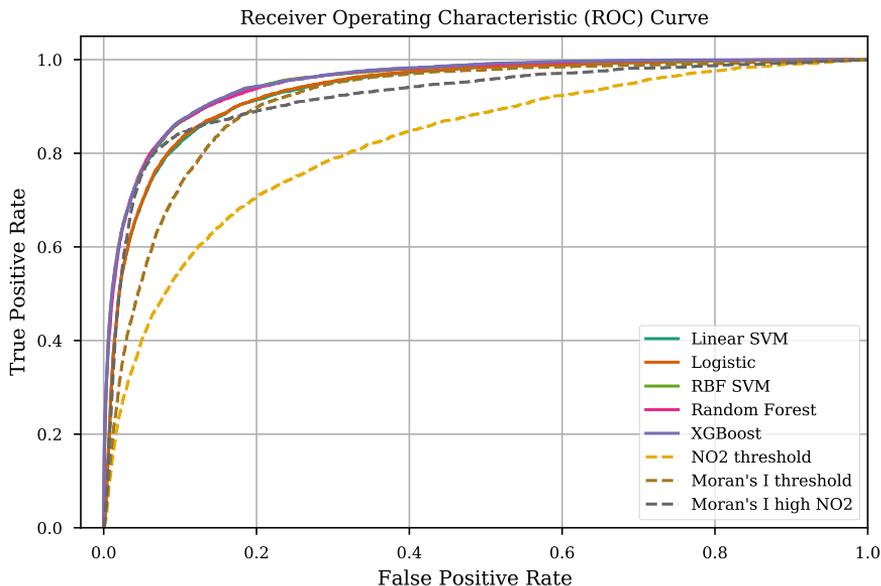Figure 5.9: Receiver Operating Characteristics (ROC) curve based on 5-fold cross-validation. Dashed lines indicate the results obtained from univariate threshold-based methods that, in this study, we considered as benchmarks.
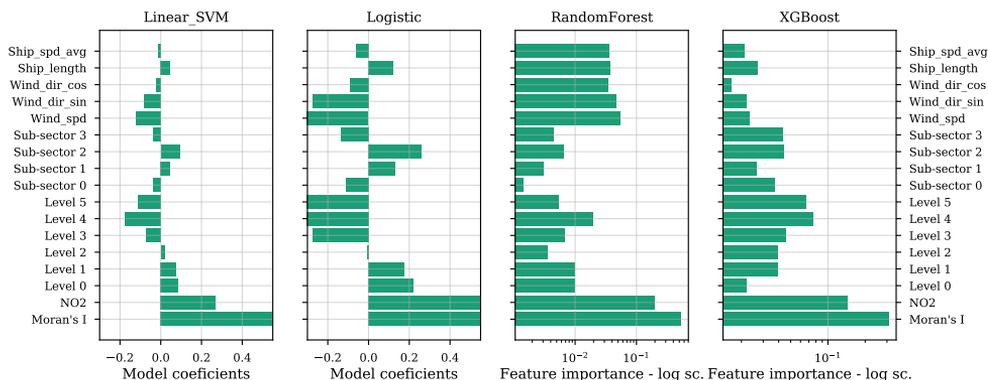
Figure 5.10: Coefficients of the features in the decision function of the linear models and the impurity-based feature importance values for tree-based models.

Segmentation Method	Pearson Correlation	Number of detected plumes
XGBoost	0.834	371
Manual Labeling	*0.781*	*334*
Random Forest	0.775	436
NO$_2$	0.774	334
Logistic	0.766	452
Linear SVM	0.765	452
RBF SVM	0.757	447
Moran's I on high NO$_2$	0.733	422
Moran's I	0.681	448

Table 5.4: Results on the comparison between the estimated amount of NO$_2$ and theoretically derived NO$_x$ ship emission proxy. Sorted in accordance with the achieved level of Pearson correlation. Italic font indicates baseline results.

ship plume. Thus, the obtained level of recall we consider as reasonably satisfactory. From Table 5.3, we can also see that the level of the standard deviation of AP scores for multivariate non-linear models is significantly lower than for linear or univariate models. This suggests that the results obtained with the nonlinear classifiers are more robust.

(ii) From Figure 5.10, we can see that Linear SVM, Logistic Regression, Random Forest, and XGBoost multivariate models utilize the spatial information provided by sub-sectors and levels. The complexity of the RBF SVM model does not allow the direct calculation of the importance of the utilized features. Even though due to the different nature of the models, the coefficients' values depicted in Figure 5.10 cannot be compared directly, the relative differences between the models' features go along with our intuition on where the plume produced by an analyzed ship should be located within a *normalized sector*. For instance, high negative coefficients for the linear models that correspond to the features *Level 4* and *Level 5* suggest that even if a high-value pixel does occur in those regions of the *normalized sector*, it was most probably produced by a source other than the analyzed ship. On the other hand, the high positive coefficients corresponding to a feature *Sub-sector 2*, tell us that if a high-value pixel occurs in the middle of the sector, it is most probably a part of the plume produced by the studied ship.
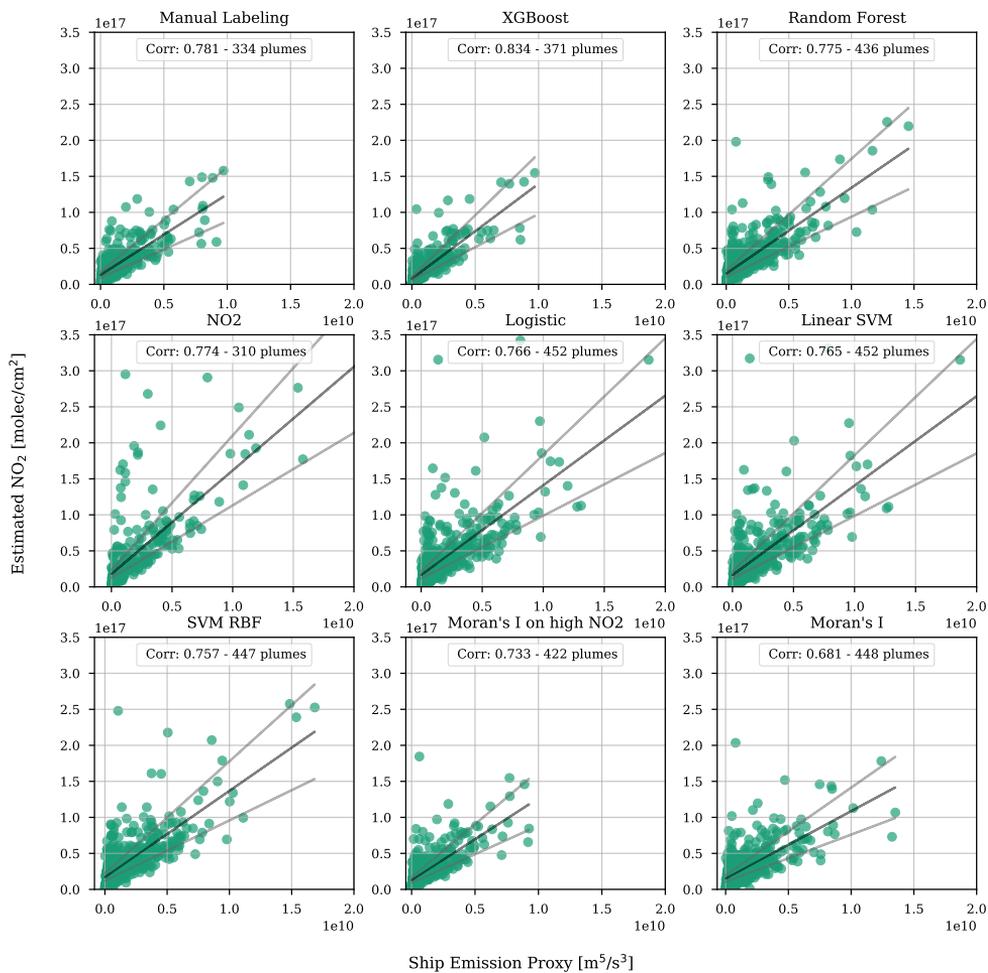
Figure 5.11: Pearson correlations between estimated (based on classification results) values of NO$_2$ emitted by each ship on a given day and a theoretical ship emission proxy. Black lines indicate a fitted linear trend. Grey lines show 30% deviations from the fitted linear trend.
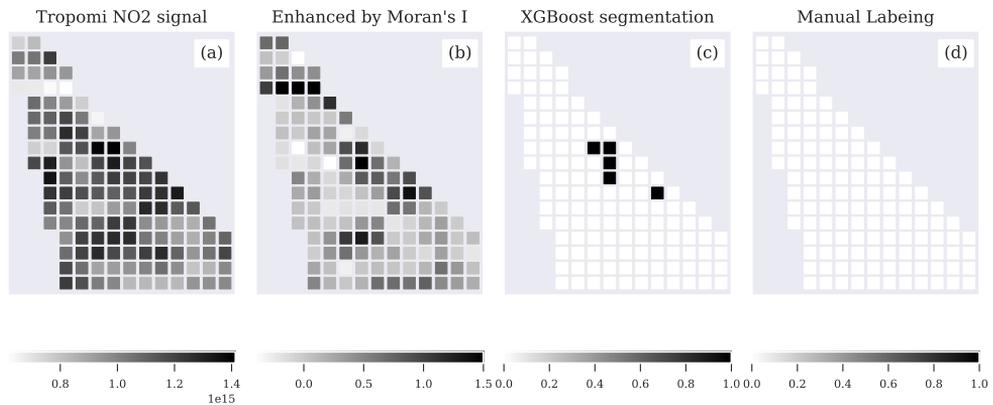
Figure 5.12: XGBoost classifier allows for the segmentation of plumes that were not recognized by the labeler. (a) TROPOMI NO$_2$ tropospheric vertical column density. Units: mol/m^2. The variable was a part of the input to machine-learning models. Ship plume is difficult to distinguish by the human eye. (b) TROPOMI NO$_2$ image enhanced by Moran's I. The variable was a part of the input to machine-learning models. After enhancement, the ship plume can be recognized better. At the top of the *ship sector* can be found an example when a cluster of low value NO$_2$ was enhanced incorrectly. (c) Results of segmentation of XGBoost model. Black pixels indicate pixels classified by the model as a "plume". (d) Human labels. The absence of black pixels means that there were no pixels within the area labeled as a plume. For all presented images, the size of the pixel is equal to 4.2 × 5 km2. Measurement date: June 24th, 2019. Ship type: Tanker. Ship length: 230 m. Average speed within the studied time scope: 14.27 kt.
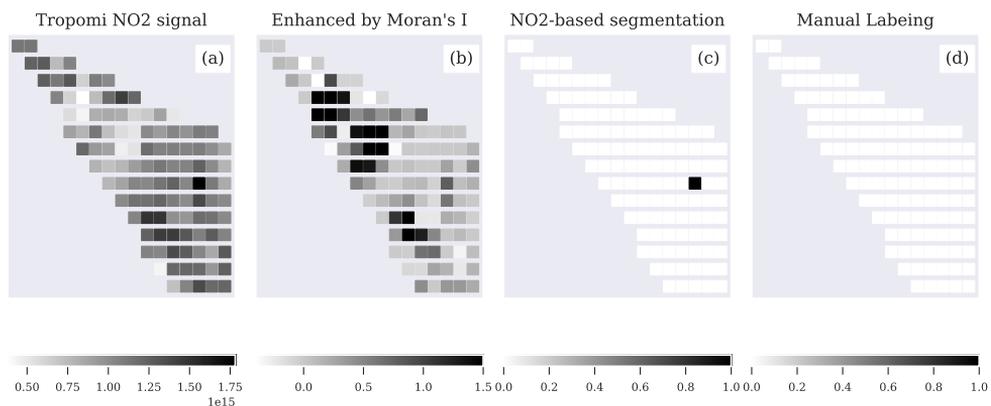
Figure 5.13: NO$_2$-based thresholding allows for distinguishing plumes cumulated within one pixel of the TROPOMI image. (a) TROPOMI NO$_2$ tropospheric vertical column density. Units: mol/m^2. (b) TROPOMI NO$_2$ image enhanced by Moran's *I*. At the top left of the *ship sector* can be found an example when a cluster of low value NO$_2$ was enhanced incorrectly. (c) Results of segmentation of *NO$_2$ threshold* method. A black pixel is a pixel that was identified by a model as a plume. (d) Human labels. The absence of black pixels means that there were no pixels within the area labeled as a plume. For all presented images, the size of the pixel is equal to 4.2 $\times$ 5 km2. Measurement date: June 9th, 2019. Ship type: Tanker. Ship length: 285 m. Average speed within the studied time scope: 15.4 kt.

5.3.2   Validation with emission proxy

Figure 6.6 provides the correlation plots of NO_2 values estimated for a given ship on a given day based on the segmentation results of a given model and the theoretically derived NO_x ship emission proxy E_s. Table 5.4 gives information on the achieved level of Pearson correlation and the number of plumes that were segmented by a certain model. Here, our baseline result is the level of Pearson correlation and the number of plumes that were identified by Manual Labeling. We can see that the majority of the models detected more plumes than the labeler. However, in all cases apart from XGBoost, the higher number of segmented plumes caused the decrement in the correlation score. The XGBoost model, on the other hand, was able to detect more plumes than the manual labeler, while achieving the highest correlation score. Such a result allows us to form a hypothesis that the developed machine-learning-based methodology is able to segment plumes better than a human labeler. An example of a case where the XGBoost classifier identifies a plume better than the human labeler can be found in Figure 5.12. More experiments are, however, required in order to make final conclusions.

The highest contrast between the scores of the performance metrics and the correlation with the emission proxy can be noted for the NO_2 threshold benchmark model. This is due to the fact that the ship plumes composed out of one pixel in our dataset were not labeled as plumes. The substantially high correlation with the emission proxy suggests that the single-pixel plumes were, nevertheless, identified by the method correctly. An illustration of such an example is provided in Figure 5.13.

5.4   Conclusions

In this Chapter, we presented a new supervised-learning-based method for the automatic evaluation of emission plumes produced by individual ships using TROPOMI data. The experiments were performed using NO_2 measurements from the TROPOMI instrument. We started with the enhancement of the TROPOMI data in order to increase the contrast between the ship plume and the background. The applied image pre-processing technique enhances the intensity of high-value pixels located in a cluster (plume) and suppresses random concentration peaks in the background. We then automatically assigned a *ship sector* to each analyzed ship, which excludes from the analysis parts of the image where the plume of the studied ship cannot be located based on wind conditions and speed/direction of the ship. As a next step, we pre-

sented a feature engineering method consisting of the normalization of the *ship sector* and its division into smaller sub-regions. Each sub-region has a different probability of containing a plume produced by the ship of interest. This way, we differentiate the plume produced by the ship of interest from all the other plumes potentially located within the *ship sector*. The set of newly created spatial *ship sector*-based features allows us to perform ship plume segmentation using multivariate machine-learning models. The application of the multivariate models gives the possibility to support the ship plume segmentation process with a set of additional one-dimensional features such as ship characteristics and speed.

We integrated several data sources into a multivariate dataset. We manually labeled the data, so that the problem of individual ship-plume segmentation can be addressed with supervised learning. We trained a set of robust linear and nonlinear multivariate classifiers and compared their performance with the segmentation results of thresholding-based univariate benchmarks. All studied non-linear classifiers showed superior results in comparison to both linear models and univariate benchmarks. With the XGBoost model, we were able to achieve more than a 20% increase in the segmentation average precision in comparison to the best benchmark univariate model. This allows us to answer positively the **RQ6** of this thesis.

We validated the proposed methodology using an independent measure, i.e. a theoretically derived NO_x ship emission proxy that we use as a reference value. For the comparison, we estimated the amount of NO_2 produced by each of the analyzed ships and calculated the Pearson correlation of the obtained results with the ship emission proxy. We compared the obtained correlations and the number of plumes segmented by each of the studied models with the results obtained from manual segmentation. We showed that with the XGBoost model, we are able to segment more plumes while achieving a 6.8% higher correlation with the emission proxy than when the plumes were segmented manually. That might suggest that the proposed method is able to find plumes that are hardly or not detectable by the human eye (**RQ7**).

5.5 Discussion

The presented approach opens new perspectives for the application of remote sensing in the domain of ship emission monitoring. However, there are several points on the generalization of results, the methodology, and the TROPOMI detection limit we would like to address here.

Firstly, we would like to discuss the possibility of the application of the proposed

Distribution of the variables from the dataset

Figure 5.14: Distribution of the dataset features for the images, where there were no visible ship plumes distinguished, and for the images, where there was a visually distinguishable ship plume.

Variable Name	No plume image	Image with a plume
Wind speed [m/s]	5.47± 2.31	5.27±2.00
Ship speed [kt]	16.83± 2.01	17.41 ± 2.04
Ship length [m]	279.92±86.64	303.99 ± 82.79

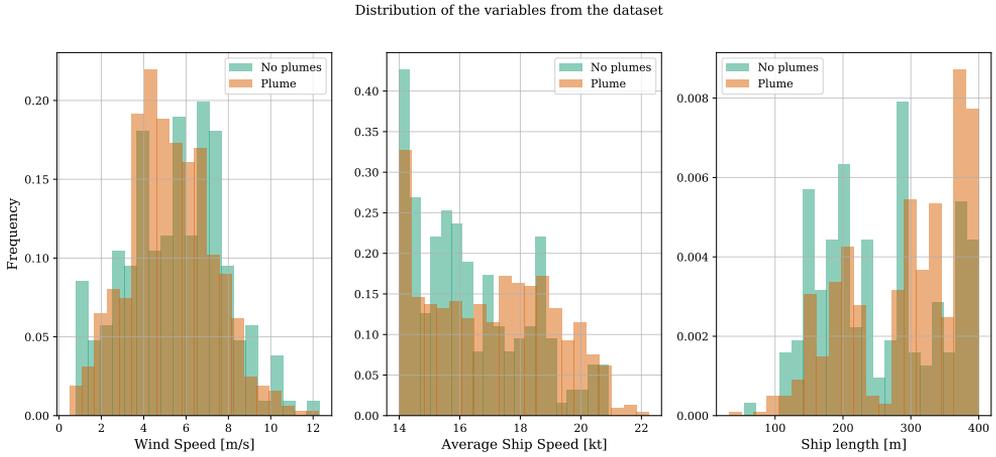Table 5.5: Average and standard deviation for the dataset features for the images, where there were no visible ship plumes distinguished, and for the images, where there was a visually distinguishable ship plume.

methodology to other regions. In this study, we presented a general approach that allows for the application of machine-learning models for more efficient, automated segmentation of plumes from individual ships using TROPOMI data. All steps of feature preparation can be performed on the data from any region of the globe. Nevertheless, the machine-learning models will have to be retrained on the region-specific datasets.

Secondly, not all regions will be equally suitable for the performance of ship emission monitoring with remote sensing. In particular, at the moment there is no scientific evidence that under the thick layer of land-based emission outflow, it will still be possible to differentiate plumes produced by ships. Therefore, areas that lie in close proximity to big cities, ports, or industrial objects are currently challenging to analyze.

The next point is the validation approaches used in this study. For the training of the machine-learning model, we used human labels. Human labeling is the basis of all machine-learning methods and this study pioneers ship plume segmentation with more efficient supervised learning based on human labeling. However, the dispersion and chemical transformation of a ship plume, as well as its non-rigid structure mean that there are always some parts of this plume that are at or beyond the visible detection limit of the combination of the TROPOMI instrument and the retrieval algorithm. This can cause errors in labeling as is demonstrated in Figure 5.12. Such mistakes if present in reasonable amounts should not affect the performance of the model, but, if the number of labeling errors is too high, the machine-learning model will not be able to learn properly, and thus, the resulting performance will be very poor. The fact that non-linear models were able to easily outperform thresholding-based benchmarks suggests that the models were able to use the provided labels for training and thus, the labeling error rate is low. Nevertheless, an independent measure of the method evaluation is needed. Since the interest of our study centers around seagoing ships, the in-situ measurements cannot be considered as a potential way of method validation. The option of on-board measurement of fuel samples, cannot be performed at the scale of the study. Therefore, a theoretical measure of ship emission potential which is ship emission proxy turns out to be the only available option of a reference value for the results of this study.

The usage of the ship emission proxy, however, has its limitations. Namely, the used ship emission proxy does not take into account many factors that influence the expected level of emission for a given ship. Nonetheless, the used proxy allows us to rank the emission potential of the analyzed ships properly.

As a following, we would like to discuss the fact that even though only fast ships

were taken into consideration in this study, the number of ships for which the plume was possible to distinguish is higher than the number of ships for which the plume was invisible for the labeler. This study focuses on observing emission sources at the edge of the detection limits of the TROPOMI instrument. It is, therefore, likely that under certain circumstances ship plumes remain undetected. We can only in part explain under what circumstances plumes are not visible. With the data presented in Figure 5.14 and Table 5.5, we show that, as expected, smaller and slower ships are more often not detected. Similarly, for high wind speeds – the detection is more challenging due to the high dilution of the ships' emissions and therefore low concentrations (the evidence can also be found in Figure 5.14 and Table 5.5). Regarding the lower detectability at lower wind speeds that can also be observed in Figure 5.14, we find some accordance with the findings from [86], where it is described how the wind speed impacts the reflectivity of the sea surface due to the shape of the waves, which in turn influences the sensors' sensitivity. However, this topic needs further study in the satellite retrieval community.

To sum up, the method presented in this study is a big step towards automated and global ship emission monitoring with remote sensing and should not be devalued by the above-mentioned limitations. Firstly, one can train a machine-learning model per region as commonly done in remote sensing. In addition, the region can serve as a feature of the model itself to make it invariant to geographic locations. Moreover, adding such variables as month, solar radiation, or temperature will make the model invariant to the seasonal changes that might be more severe at northern latitudes. Secondly, main ship routes go through both more and less suitable regions for the satellite observations. Thus, a selection of the more convenient regions will still allow us to use our approach for efficient monitoring of the emission levels produced by ships that follow those routes. Moreover, the obtained good results both in terms of segmentation quality and comparison with the emission proxy suggest that labeling has been of substantial quality. The proposed methodology also opens new research directions. For instance, human labeling can be replaced with chemical plume dispersion models, which will further improve the labeling quality and make the proposed methodology even more effective. Finally, the problem of visibility of ship plumes that have been unrevealed with the presented study, once solved, will give us a great overview of the capabilities of TROPOMI sensors.

Chapter 6

Automatic detection of anomalously emitting ships

Abstract

In the previous chapter, we introduced the first method for a large-scale ship NO_2 estimation – a supervised machine learning-based segmentation of ship plumes on TROPOMI image patches. However, both challenging data annotation and insufficiently complex ship emission proxy used for the validation limit the applicability of the model for ship compliance monitoring. In this Chapter, we present a methodology for the automated and scalable selection of potentially non-compliant ships using a combination of machine learning models on TROPOMI data. It is based on a proposed regression model predicting the amount of NO_2 that is expected to be produced by a ship with certain properties operating in the given atmospheric conditions. The model does not require manual labeling and is validated with TROPOMI data directly. The differences between the predicted and actual amount of produced NO_2 are integrated over observations of the ship in time and are used as a measure of the inspection worthiness of a ship. To add further evidence, we compare the obtained results with the results of the previously developed segmentation-based method. Ships that are also highly deviating in accordance with the segmentation method require further attention. If no other explanations can be found by checking the TROPOMI data, the respective ships are advised to be the candidates for inspection.

6.1 Introduction

The current state-of-the-art of large-scale methods for NO_2 ship plume modeling use thresholding or supervised machine-learning-based segmentation of TROPOMI image patches to attribute the measured NO_2 to individual ships [62, 63]. We presented those methods in Chapter 4 and Chapter 5 correspondingly. The latter methodology is an automated procedure improving significantly upon previously used manual methods. However, due to the low signal-to-noise ratio of TROPOMI measurements, ship plumes are often hard to delineate, which makes the process of manual data annotation time-consuming and potentially erroneous. The absence of ground truth for a given task requires an alternative measure of validation. One possibility is the usage of theoretical models for ship emission approximation – ship emission proxy [33, 41]. An example of such a proxy is the one that was utilized by us in previous chapters (explained in Section 2.4) in [41]. However, the proxies (and this one in particular) do not cover the full list of factors that can potentially influence the levels of ship emissions (e.g. amount of cargo on board, local meteorological conditions), which does not allow a proper quantification of the effects of the errors coming from manual labeling. Consequently, the possibilities of the application of this approach to the task of monitoring NO_2 emissions from individual ships are limited.

In this Chapter, we propose a robust method for automated selection of anomalously NO_2 emitting seagoing ships, addressing the last research question of the thesis:

- **RQ8**: How to identify ships that are potential anomalous emitters using TROPOMI data?

The presented approach does not require data labeling and is validated using TROPOMI data directly. Moreover, our method is based on the integration of multiple observations, which gives a more complete perspective on ship performance. This is achieved by training a specifically designed regression model, which predicts the amount of NO_2 that is expected to be observed by the TROPOMI instrument for a given ship operating in certain atmospheric conditions. The difference between the predicted and actual amount of observed NO_2 is integrated over the available number of ship observations. The integrated difference we consider a measure of inspection worthiness of the ship.

For the training of the regression model, we use the concept of ship Region of Interest (RoI) defined in Chapter 4. We apply Automated Machine Learning (AutoML) (for the explanation of the concept, c.f. Section 2.3) to optimize the machine-learning-based regression pipeline for the NO_2 prediction. To assure the robustness of the

proposed method, we compare the results obtained with the regression model with
the method for ship plume segmentation [63] introduced in Chapter 5. Ships that
are also ranked as highly deviating in accordance with the ship plume segmentation
model are nominated as anomalous emitters and require further attention. We visually
check the TROPOMI data for objective explanations of anomalous results. If no other
explanations are found, the ships are advised to be candidates for further inspection.

The rest of this Chapter is organized as follows: In Section 6.2, we describe the data
sources used in this study. In Section 6.3, we introduce the developed methodology,
which is followed by the results presented in Section 6.4. In Sections 6.5 and 6.6, the
reader can find the discussion and final conclusions respectively.

6.2   Data

In this study, the variable of interest is NO_2 tropospheric vertical column density
– VCD_{trop} [31]. As described in Chapter 2, the VCD_{trop} column is the result of
a transformation of SCD (slant column density) using the air mass factors (AMS)
calculated, among the others, on the basis of historical emission inventories [31]. This
results in the fact that the plumes located in the regions of historical shipping lanes will
be enhanced by the retrieval algorithm [30]. To minimize the impact of the potential
bias, such variables as background NO_2 SCD, AMF, surface albedo, and sun/satellite
geometry will be used as model features for ship NO_2 estimation.

In this Chapter, we analyze the same region[1] in the eastern Mediterranean Sea
as in Chapter 5. The study period is 20 months, starting from 1 April 2019 until 31
December 2020. To obtain the image patches of regular size, we perform regridding[2]
of the original TROPOMI data into a grid of regular size $0.045° \times 0.045°$, which for
the studied area translates to approximately 4.2×5 km^2 [63]. The following quality
filers were applied on the TROPOMI data: only pixels flagged with $qa\_value > 0.5$
[93] are taken into consideration. In addition, since the TROPOMI observations of
scenes covered with clouds should not be considered valid, we filtered out from the
data pixels with a cloud fraction higher than 0.05. With this level of cloud filtering,
we lost approximately 35% of ship observations.

In order to prevent the occurrence in our dataset of ships below the detection
limit, we focus our analysis on the seagoing ships that are longer than 150 meters

---

[1]The studied region is restricted by the following coordinates: long: [19.5°; 29.5°], lat: [31.5°;
34.2°].

[2]The regridding is performed using the Python package HARP v.1.13.
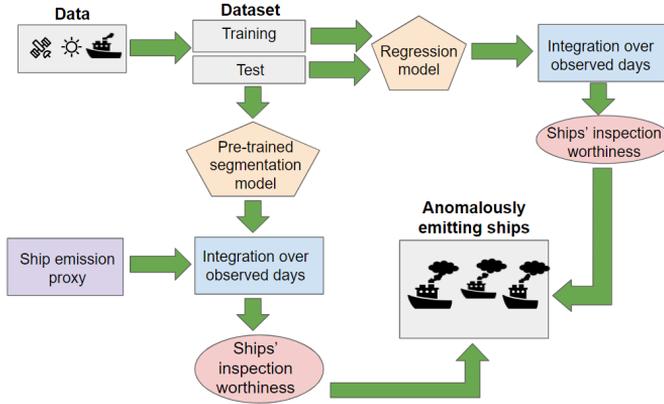
Figure 6.1: High-level diagram of the proposed methodology.

and faster than 12 kt, which is slightly faster than the TROPOMI detection limit established in Chapter 3. Another situation we want to prevent is when too many ships contribute to the creation of the detected NO_2 plume, as in this case, quantification of individual contributions is extremely challenging. Thus, we remove the ships, whose trajectories within 2 hours before the satellite overpass, intersect with more than 3 other neighboring ships. This is a trade-off between a sufficient size of the dataset and the complexity of the problem of the quantification of individual contributions. Among all ship types present in the dataset, for the detection of anomalously emitting ships, we focus our attention on two ship types: containers and tankers. Other ship types have not been represented in the dataset in a sufficient amount to obtain statistically significant results.

6.3   Method

In this Section, we present the method for automated detection of ships that produce anomalously high amounts of NO_2. The method is composed of the following steps: we train a regression model for the prediction of the amount of NO_2 within the RoI of the analyzed ship. We calculate the difference between the observed and predicted amount of NO_2 and integrate this value over all observations of the same ship within the studied period. The integrated difference between the real and predicted value of NO_2 we consider as a measure of the inspection worthiness of the ship. We rank the studied ships accordingly. To assure the robustness of the results, we apply the ship

plume segmentation model [63] to the same dataset. We compare the results obtained using the segmentation model with the value of the theoretical ship emission proxy. We consider the results of the comparison to be a measure of the inspection worthiness according to the segmentation model. The ships that are high on the inspection worthiness list of both independently trained and validated machine-learning models are considered to be potentially anomalously emitting. We evaluate the obtained results by visual inspection of the corresponding TROPOMI observations. Figure 6.1 provides a high-level explanation of the proposed method for the detection of anomalously emitting ships. Below, each step of the methodology is described in detail.

6.3.1 Regression model

Here, we describe our proposed regression model as part of a method for the detection of anomalously emitting ships. Firstly, we provide a formal definition of the proposed way for ship NO_2 estimation with the regression model. Then, we introduce the details of training and optimization of the machine-learning methodology proposed in this study.

Formalization of the problem

For a given ship $s \in S$ on a given day $d \in D$, the real amount of NO_2 observed by TROPOMI is calculated as:

$$NO_{2;d,s} = \sum_{i \in RoI_{d,s}} VCD_{NO_2;i} \qquad (6.1)$$

where VCD_{NO_2} is the value of the retrieved TROPOMI pixel within the RoI of the analyzed ship, where the RoI of the ship is a *ship sector* defined in accordance with the description proposed in Chapter 4 (c.f. Section 4.2.2). We then use a machine-learning model f that based on values of features $X \in \mathbb{R}$ predicts the expected amount of NO_2: $\hat{NO}_{2;d,s} \in \mathbb{R}$.

$$\hat{NO}_{2;d,s} = f(X_{d,s}) \qquad (6.2)$$

The list of features X can be found in Table 6.1. In Figure 6.3, we provide histograms of the features. As a next step, we calculate $diff_{d,s}[\%]$ – a percentage difference between the predicted and observed amount of NO_2. Finally, assuming
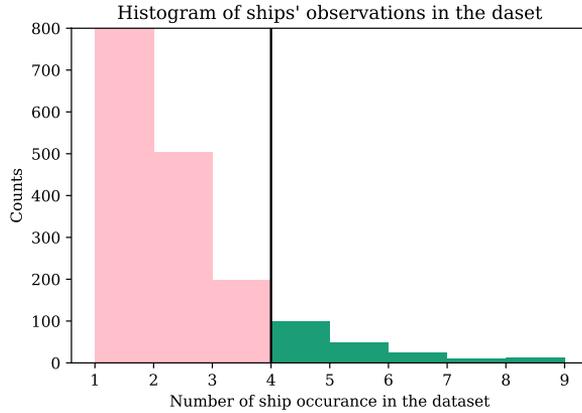
Figure 6.2: Histogram of occurrences of the same ship in the created dataset. The black line indicates the set level of *min_obs_nb*. Only ships that have been observed more than *min_obs_nb* = 4 days are taken into account for the detection of anomalously emitting ships.
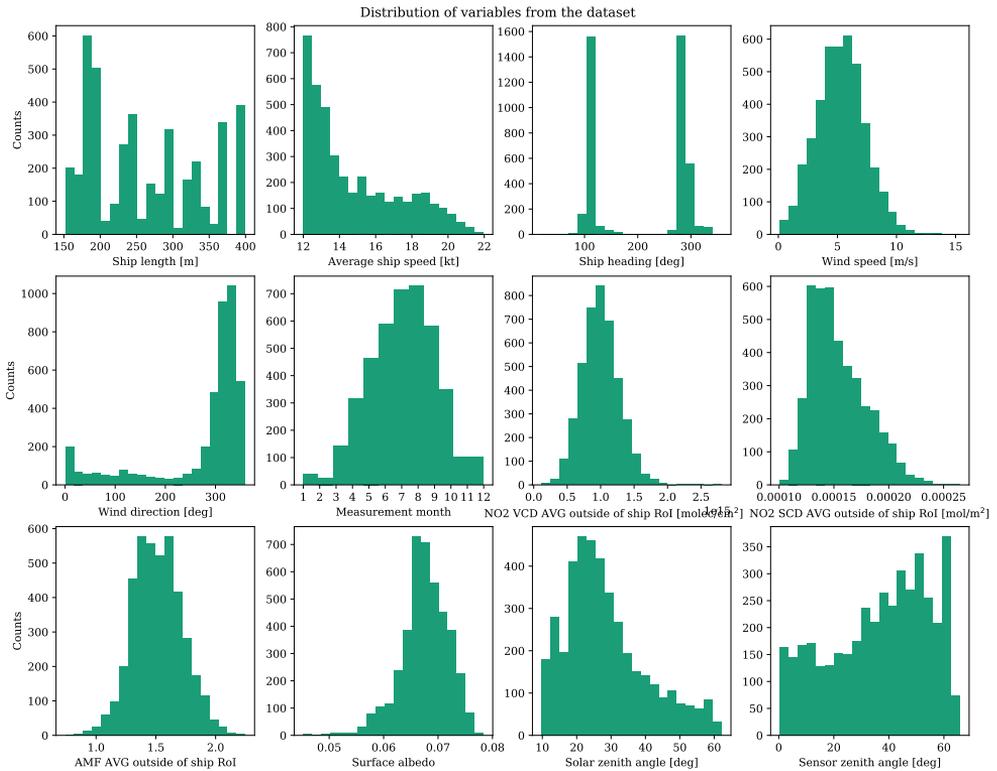
Figure 6.3: Distribution of variables of the dataset.

$|D_s|$ is the number of days when the ship s was observed, min_obs_nb is the minimum number of days we require the ship to be present in the dataset, for each ship $s \in S : |D_s| \geq min\_obs\_nb$, we integrate the obtained differences over the observed number of days calculating arithmetic mean $\mu(diff_{d,s})$ and standard deviation $\sigma(diff_{d,s})$. To ensure that our ship profile is representative to make the decision about being anomalously emitting and taking into consideration data availability (see Figure 6.2), we set the threshold as $min\_obs\_nb = 4$.

A high value of $\mu(diff_{d,s})$ represents a situation when the observed value of NO_2 was repeatedly underestimated by the model. This means that the amount of NO_2 observed was consistently higher than can be expected given the ship's characteristics and operational atmospheric conditions. In other words, $\mu(diff_{d,s})$ is a measure of the inspection worthiness of the ship in accordance with the regression model IW_s^{regr}. The value $\sigma(diff_{d,s})$ is a measure of the consistency of the obtained results. Since the TROPOMI measurement results have a lower limit and do not have an upper limit, a very high $\sigma(diff_{d,s})$ can only occur from the fact that very high values of NO_2 were assigned to a ship that on a regular basis does not produce that much – only high NO_2 outliers can cause a high standard deviation. Such a situation is not of our interest. Therefore, ships with outlying values of $\sigma(diff_{d,s})$ will be removed from the analysis. The value of $\sigma(diff_{d,s})$ is considered to be outlying if $\sigma(diff_{d,s}) > \mu(\sigma(diff_{d,s})) + 2\sigma(\sigma(diff_{d,s}))$, which corresponds to 5% of the highest observations of $\sigma(diff_{d,s})$.

Model optimization

Similarly to the previous chapters, for the selection of the regression and optimization of its hyperparameters, we use a 5-fold nested scheme of cross-validation. Within the outer loop of cross-validation we create 5 "hold out" non-overlapping test sets and 5 training sets. Given the considered application, the test sets are used for:

1. Performance evaluation of the regression model.

2. Detection of anomalously emitting ships.

Within the inner loop of cross-validation, we split the training set into training and validation, which are used for the optimization of the regression model performance. The task of model optimization is tackled here with automated machine learning (AutoML) [49] by solving a so-called CASH problem (for an explanation of the concept and benefits coming from its application see Section 2.3). Given the absence of available

Feature type	Feature name
Ship related	Ship length
	Ship speed
	Ship heading
	Gross tonnage
	Ship type
State of the atmosphere	Wind speed
	Wind direction
	Surface albedo
	Solar zenith angle
	Measurement month
Priors for background	Average NO_2 VCD_{trop} outside *ship sector*
	Average NO_2 SCD outside *ship sector*
	AMF outside *ship sector*
	Sensor zenith angle

Table 6.1: List of features used for the regression model. The area outside the *ship sector* is restricted to the ship neighborhood defined as the *ship plume image* in accordance with Chapter 4.

benchmarks for our original dataset, such a technique allows for an efficient selection of a regression model and feature preprocessor from among a wide variety of machine-learning models and feature transformation techniques. As mentioned in Section 2.3, we address the CASH problem using TPOT (Tree-based Pipeline Optimization Tool) [77] – a Python package for automatic selection of machine-learning pipelines based on genetic programming (GP) [58].

The results obtained using the TPOT AutoML library are benchmarked towards the results obtained using the eXtreme Gradient Boosting (XGBoost) [22] regression model with the default hyperparameters settings. The XGBoost model is considered to be a good choice when it comes to tabular data [45], as well as showed the best performance on the same type of data in Chapter 5.

6.3.2   Detection of anomalously emitting ships

In order to ensure the robustness of the proposed method for detecting anomalously emitting ships, we compare the results obtained with the regression model with another, independently trained and validated machine-learning model applied to the same dataset. We intersect the results obtained with both considered models in order to obtain a list of potentially anomalously emitting ships. Hereafter, we explain how

the model introduced in Chapter 5 is added to the presented regression model as a decision support tool, and explain how the results of both models are used to make a decision regarding the candidate selection of anomalously emitting ships.

Segmentation Model

As a support tool for the presented regression model, we use the ship plume segmentation model prepared in accordance with the methodology introduced in Chapter 5. Below, we provide a formal explanation of how we propose to use this method for the detection of potentially anomalously emitting ships.

For a given ship $s \in S$ on a given day $d \in D$, the estimated with the segmentation model amount of NO_2 can be expressed as:

$$\hat{NO}_{2;d,s} = \sum_{i \in RoI_{d,s}} \hat{y}_i \cdot NO_{2,i}, \tag{6.3}$$

where $\hat{y}_i \in \{0, 1\}$ and $NO_{2,i}$ are the output of the segmentation model for the pixel i and the value of the pixel i of the ship s on day d.

To detect potential anomalous emitters, for each ship observation, we calculate the value of the ship emission proxy $E_{d,s}$ (for definition c.f. Section 2.4). For each ship $s \in S : |D_s| \geq min\_obs\_nb$, we aggregate the $\hat{NO}_{2;d,s}$ and $E_{d,s}$ over the days of observation by calculating their arithmetic mean μ. We assume that $\mu(\hat{NO}_{2;d,s})$ is linearly proportional to $\mu(E_{d,s})$. Therefore, we can express it as:

$$\mu(\hat{NO}_{2;d,s}) = \alpha \cdot \mu(E_{d,s}) + \beta + \epsilon_s, \tag{6.4}$$

where α and β are the parameters of the fitted linear equation. We consider ϵ_s the measure of the inspection worthiness of the ship in accordance with the segmentation model IW_s^{segm}. The measure of the consistency of the results is defined as the standard deviation of the estimated values of NO_2, $\sigma(\hat{NO}_{2;d,s})$. The ships for which $\sigma(\hat{NO}_{2;d,s}) > \mu(\sigma(\hat{NO}_{2;d,s})) + 2\sigma(\sigma(\hat{NO}_{2;d,s}))$ are considered to be outlying and will not be taken into consideration.

6.3.3 Merge of two models to identify anomalous ships

In order to identify anomalously emitting ships, we intersect the results obtained with the two independently trained/validated machine-learning models: a newly developed regression model for the prediction of ship's NO_2 within the assigned *ship sector*, and

the ship plume segmentation model presented in Chapter 4. To assure the comparability of the results, we perform a normalization of the inspection worthiness measures obtained from both used methods, defining $norm\_IW_s^{regr}, norm\_IW_s^{segm} \in [0,1]$. The normalization is performed using min-max scaling applied on IW_{regr_s} and IW_{segm_s} such that:

$$norm\_IW_s^{regr} = \frac{IW_s^{regr} - min(IW_s^{regr})}{max(IW_s^{regr}) - min(IW_s^{regr})} \qquad (6.5)$$

$$norm\_IW_s^{segm} = \frac{IW_s^{segm} - min(IW_s^{segm})}{max(IW_s^{segm}) - min(IW_s^{segm})} \qquad (6.6)$$

Providing a decision threshold t, the ship is assigned to the list of anomalously emitting ships in accordance with the following rule:

$$norm\_IW_s^{regr} > t \wedge norm\_IW_s^{segm} > t \iff s \in Anomalous\_emitters, \qquad (6.7)$$

such that:

$$Anomalous\_emitters = \{s_1, ..., s_n\} :$$
$$norm\_IW_{s_i}^{regr} \cdot norm\_IW_{s_i}^{segm} < norm\_IW_{s_{i+1}}^{regr} \cdot norm\_IW_{s_{i+1}}^{segm} \qquad (6.8)$$

The decision about the selection of the used threshold level t is left to the user. In this study, the threshold was manually selected as $t = 0.55$.

6.4   Results

In this Section, we present the obtained results. We first present the results of the regression model optimization. We then show the aggregated results of the application of the regression and segmentation models and perform the selection of potentially anomalously emitting ships. Finally, using a one-way ANOVA analysis of group differences, we inspect the obtained results for the presence of a decision bias resulting from the merge of regression and segmentation models.

6.4.1   Regression model optimization

In Table 6.2, we present the results of the regression model optimization. The application of the TPOT pipeline optimization algorithm allowed us to improve the results

Method	Pearson	R^2
TPOT	0.740 ± 0.058	0.538 ± 0.08
Default XGBoost	0.715 ± 0.057	0.497 ± 0.098

Table 6.2: Regression model results. Hyperparameters applied for AutoML optimization: Maximum evaluation time: 10 min; Population size: 50; Number of generations: 50; Early stopping criteria: 10.

Feature processor	Model
MaxAbs Scaler	Gradient Boosting [39]
MaxAbs Scaler	Gradient Boosting
Polynomial Features (2_{nd} deg.)	XGBoost [22]
Standard Scaler	Gradient Boosting
Standard Scaler	XGBoost

Table 6.3: A model and a feature pre-processor selected by TPOT as optimal at a given iteration of cross-validation.

of both used quality metrics over our benchmark – default XGBoost. In Table 6.3, we provide models and feature pre-processing methods selected as optimal (best performance on validation set) at each cross-validation iteration. The XGBoost model was still one of the most often selected optimal models. The advantage of the AutoML application, in this case, was gained by the possibility of hyperparameters optimization and selection of feature pre-processing method. Another well-performing model was the related Gradient Boosting algorithm.

6.4.2 Detection of anomalously emitting ships

Here, we analyze the results of the application of the regression and plume segmentation model with the aim of detecting anomalously emitting ships. First, for each model, we calculated the measures of the consistency of the results, i.e. $\sigma(diff_{d,s})$ and $\sigma(\hat{NO}_{2;d,s})$, while removing the resulting outlying values from the analysis. Figure 6.4 presents the consistency measures for regression and segmentation models along with the applied cut-off thresholds.

In Figure 6.5, we depict the integrated results of the regression model for each studied ship ($\mu(diff_s)$, $\sigma(diff_s)$) and rank them in ascending order of inspection worthiness, $IW_s^{regr} = \mu(diff_s)$. Ships for which the observed level of NO_2 is substantially higher than the predicted level are the most interesting for us. Figure 6.6 presents the
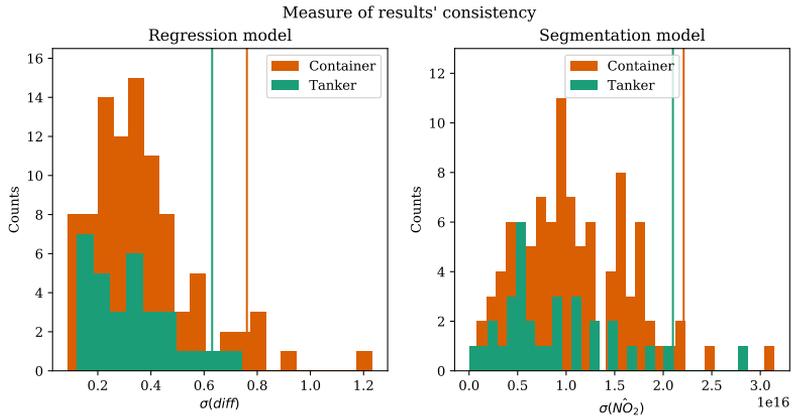
Figure 6.4: A measure of the consistency of results of the regression model $\sigma(diff_s)$ and segmentation model $\sigma(\hat{NO}_{2;d,s})$. The threshold $\mu(\hat{NO}_{2;d,s}) + 2\sigma(\hat{NO}_{2;d,s})$ is indicated with vertical lines.
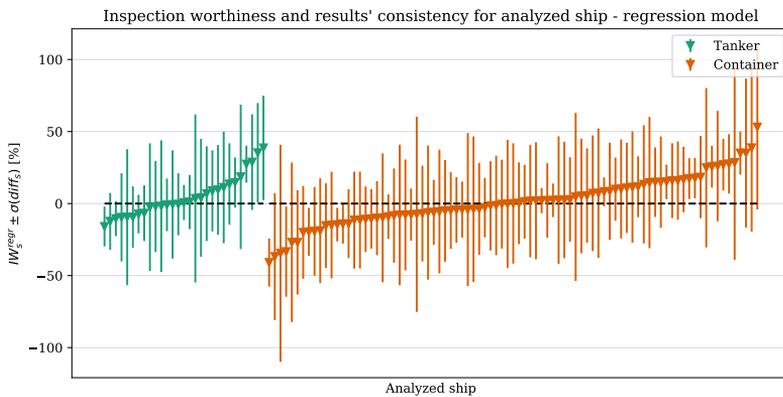
Figure 6.5: The triangle-shaped markers indicate the measure of ship inspection worthiness in accordance with the regression model IW_s^{regr}. The vertical lines indicate $\sigma(diff_s)$ - the measure of the consistency of results for a given ship.
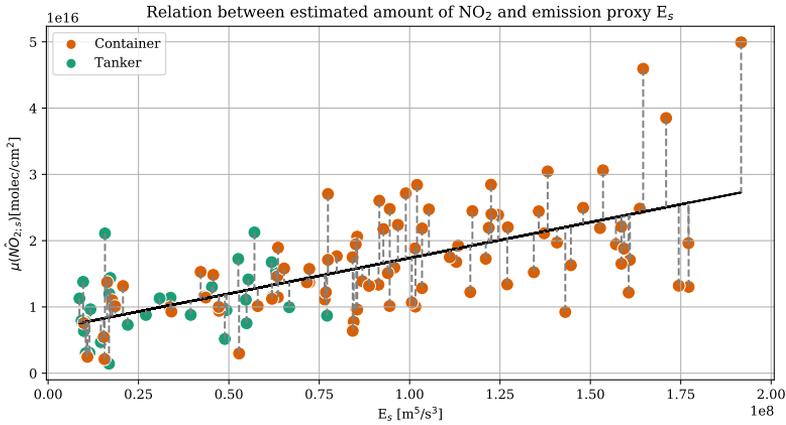
Figure 6.6: Relation between the estimated amount of NO_2 using the segmentation model and ship emission proxy with a fitted linear trend. Gray dashed lines indicate the measure of the ship inspection worthiness IW_s^{segm} according to the plume segmentation model.

resulting relationship between the averaged amounts of $\mu(\hat{NO}_{2;s})$ for each ship and averaged ship emission proxy $\mu(E_s)$. The black line indicates the fitted linear trend. The gray dashed lines indicate the ship inspection worthiness IW_s^{segm}. The ships for which the IW_s^{segm} is the highest are of our main interest.

Next, we combine the errors obtained from the regression and the ship plume segmentation models. Figure 6.7 shows the combined inspection worthiness for the two studied ship types. Black scatter plot markers indicate the analyzed ships. The size of the markers is scaled in accordance with the average value of the ship's emission proxy. Ships located in the green zone of the plots, we consider as weak emitters, because both of the models overestimate the actual level of NO_2. Two yellow zones indicate ships for which one of the models overestimates the actual level of NO_2, while the other model underestimates it. This can be due to the low resistance of the particular machine-learning model to certain types of difficult modeling conditions, or systematic errors. To name a few, the combination with land-based NO_2 sources, a plume accumulated within one TROPOMI pixel, certain atmospheric conditions, etc. Finally, the red zone of a plot indicates ships that are most inspection-worthy according to both models. We call those ships potentially anomalously emitting since throughout twenty months of analysis they were producing more than is expected based on their characteristics and operational atmospheric conditions. Clearly, to make final conclusions, the detected ships should be studied closer.
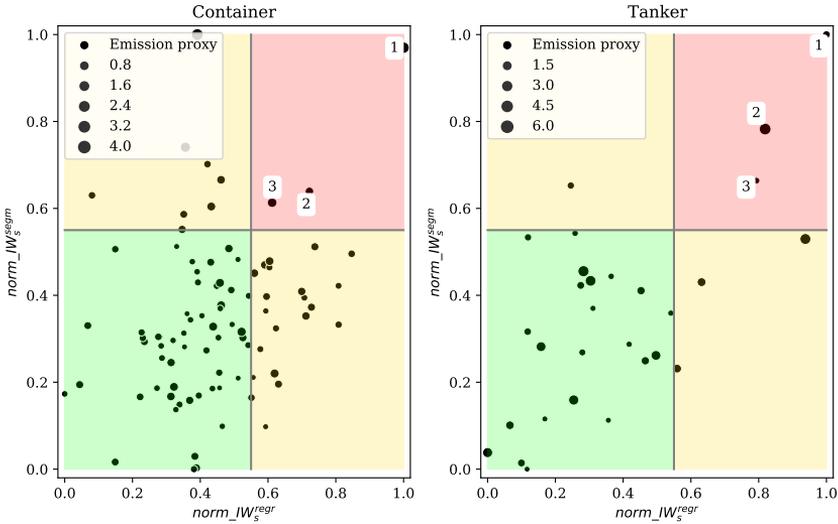
Figure 6.7: Combination of results of segmentation and regression models. Values of the inspection worthiness obtained from each model were normalized using a min-max scaler.

6.4.3 Visual verification of potential anomalous emitters

In order to make final conclusions regarding the ships that were identified by the proposed method as anomalously emitting, as a next step, we visually analyzed the TROPOMI observations related to those ships. Figure 6.8a – c and Figure 6.9a – c provide the TROPOMI image patches for the red-zone containers and tankers respectively. On the image patches from the corresponding dates of TROPOMI observations, we indicate the trajectory of the ship of interest, the other ships in the image patch, and the pixels that were classified as a part of the plume of the ship by the segmentation model.

First, we can see that for each ship, there are image patches where the segmented plume was in fact produced by another ship. This underlines the earlier mentioned constraint that intersecting ship plumes pose a challenge for this type of analysis. Nonetheless, each container ship selected as a potential anomalous emitter has at least two measurement days where there are no other candidates for producing the observed/segmented NO_2 plume. Comparing the values of results consistency (see Table 6.4) for ships selected as anomalous emitters with the data distribution for the
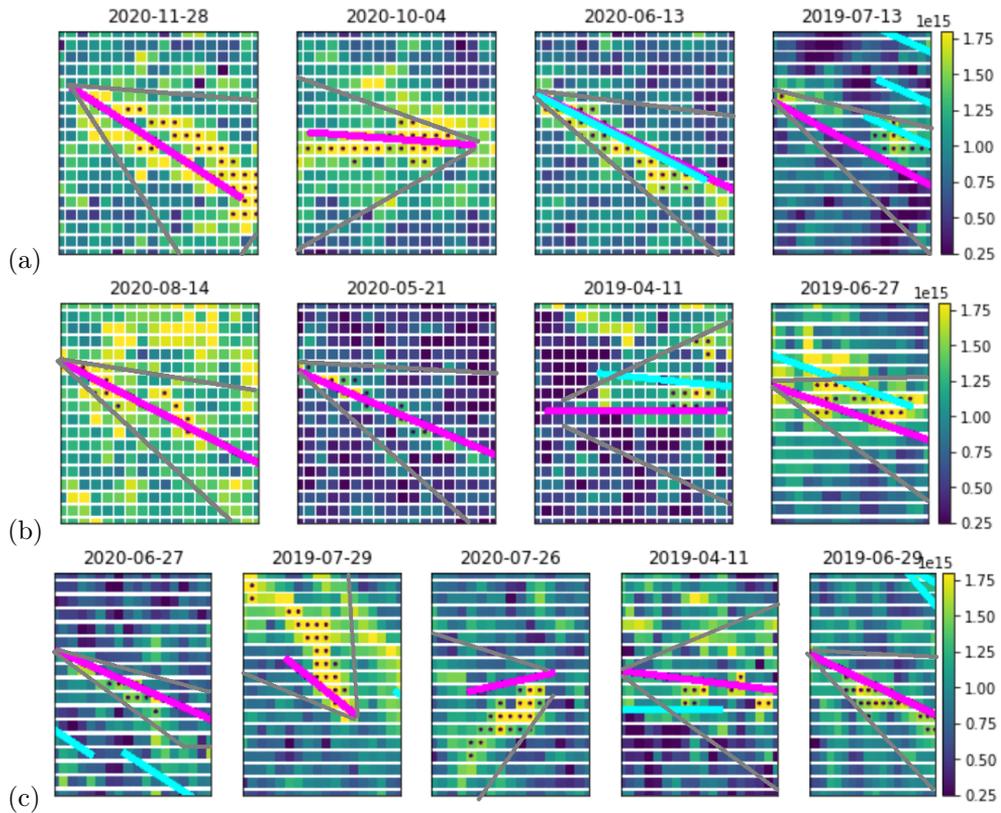
Figure 6.8: Ship type: Container. Lines represent shifted ship tracks. Magenta line – ship of interest. Cyan line – other ships in the area. Grey lines – borders of the *ship sector*. Dots indicate pixels classified by the segmentation model as a plume. a) Outlying ship 1. Ship length: 398 m. Average ship speed: 19.6 kt. Year of built: 2008. b) Outlying ship 2. Ship length: 363 m. Average ship speed: 17.5 kt. Year of built: 2011. c) Outlying ship 3. Ship length: 397 m. Average ship speed: 18.4 kt. Year of built: 2006.
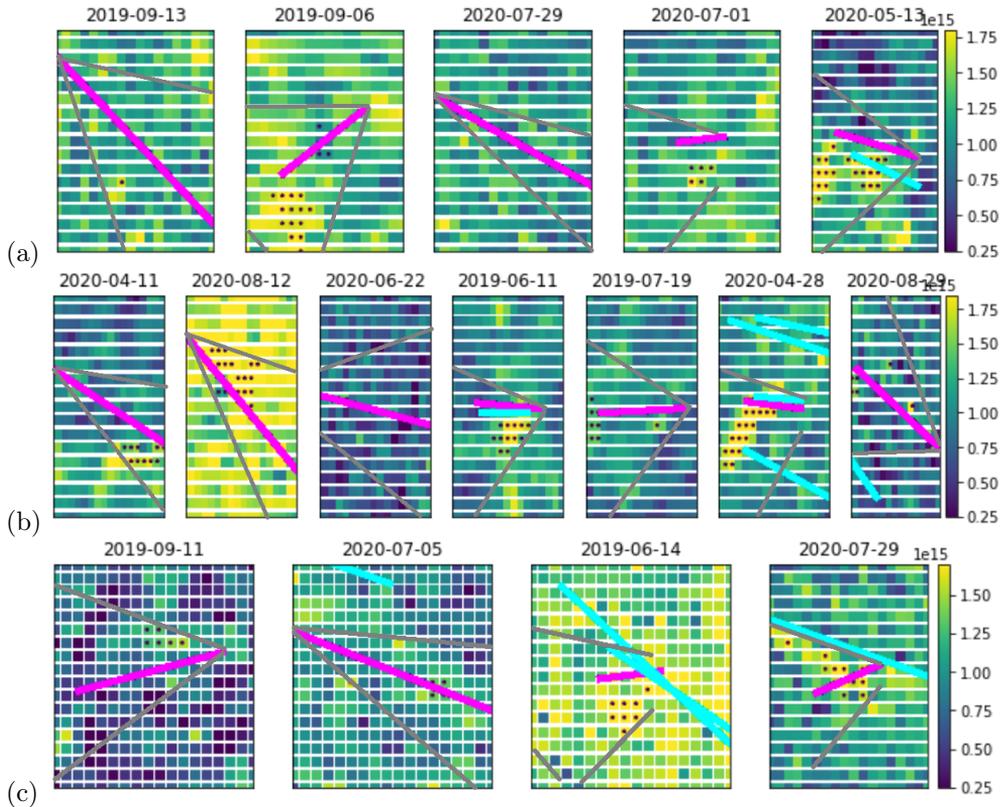
Figure 6.9: Ship type: Tanker. Lines represent shifted ship tracks. Magenta line – ship of interest. Cyan line – other ships in the area. Grey lines – borders of the *ship sector*. Dots indicate pixels classified by the segmentation model as a plume. a) Outlying ship 1. Ship length: 180 m. Average ship speed: 15.3 kt. Year of built: 2016. b) Outlying ship 2. Ship length: 315 m. Average ship speed: 16.1 kt. Year of built: 2008. c) Outlying ship 3. Ship length: 179.5 m. Average ship speed: 13 kt. Year of built: 2017.

Ship type	Ship Id	$\sigma(diff)$	$\sigma(\hat{N}O_2)$
Container	1	0.57	$1.5 \cdot 10^{16}$
	2	0.17	$0.99 \cdot 10^{16}$
	3	0.22	$1.5 \cdot 10^{16}$
Tanker	1	0.36	$2.03 \cdot 10^{16}$
	2	0.33	$1.4 \cdot 10^{16}$
	3	0.12	$0.65 \cdot 10^{16}$

Table 6.4: Measures of results consistency of regression ($\sigma(diff)$) and segmentation ($\sigma(\hat{N}O_2)$) models, for ships identified as anomalous emitter. Ship Ids are in accordance with the numbers assigned in Figure 6.7 for containers and tankers respectively.

whole set of studied ships (Figure 6.4), we can see that values of interest are located in the middle of the data distribution. Therefore, we do not have reasons to remove any of the selected ships from the list of anomalous emitters.

In the case of tankers, the situation is different. For a potential anomalous emitter with Id 1 (c.f. Figure 6.9a), we can see that for two (2019-09-13, 2020-07-29) out of five measurement days, the segmentation model did not segment any plumes. In addition, for one measurement day (2020-05-13), the segmented plume was at least partially produced by another ship. Finally, the obtained $\sigma(\hat{N}O_2)$ is very high and close to the applied cut-off threshold. Therefore, we conclude that the given ship should be removed from the list of potential anomalous emitters.

For the tanker with Id 2, both $\sigma(\hat{N}O_2)$ and $\sigma(diff)$ are within the distributions. However, from Figure 6.9b, we can see that at least two times (2019-06-11, 2020-04-28) the segmented plumes were produced by more than one ship. In three other cases (2020-04-11, 2019-07-19, 2020-08-29), the segmented pieces of plumes partially or fully belong to other emitters. For the measurement day of 2020-06-22, the model did not segment any plume. The one remaining measurement from the profile of a given ship does not justify the addition of that ship to the list of anomalous emitters.

Finally, for the tanker with Id 3, there is one measurement day (2020-07-29) when the segmented plume was at least partially produced by another ship. The rest of the image patches, nevertheless, show visually distinguishable NO_2 plumes that can be attributed to the ship of our interest. Consequently, we do not have reasons to remove a given ship from the list of potential anomalous emitters.

Ship type	Variable	Strong emitters	Weak Emitters
Tanker	Year of built	2013 ± 5	2009 ± 4
	Ship length [m]	224 ± 78	253 ± 66
	Ship speed [kt]	14.8 ± 1.5	14.8 ± 1.6
	Wind speed [m/s]	4.9 ± 0.4	5.0 ± 0.7
	Average IoU	0.07 ± 0.1	0.05 ± 0.06
Container	Year of built	2008 ± 2	2012 ± 5
	Ship length [m]	386 ± 20	340 ± 70
	Ship speed [kt]	18.5 ± 1.	17.1 ± 1.7
	Wind speed [m/s]	4.8 ± 0.5	5.1 ± 0.8
	Average IoU	0.07 ± 0.02	0.04 ± 0.04

Table 6.5: Statistical summary for important factors that influence levels of produced NO_2 for ships that by both models were identified as strong and weak emitters. IoU stands for Intersection over Union.

Ship type	Variable	F statistic	p-value
Tanker	Year of built	2.3	0.13
	Ship length	0.48	0.49
	Ship speed	0.004	0.95
	Wind Speed	0.12	0.72
	Average IoU	0.4	0.53
Container	Year of built	1.7	0.19
	Ship length	0.24	0.27
	Ship speed	1.95	0.16
	Wind Speed	0.53	0.47
	Average IoU	1.32	0.25

Table 6.6: One way ANOVA for the significance of the statistical difference between samples of ships identified as strong and weak emitters. IoU stands for Intersection over Union.

6.4.4 Decision bias

To select the anomalously emitting ships, we combined the results of two independently trained models: a regression model for ship NO_2 estimation and a model of ship plume segmentation. Taking this into account, as a final step of the analysis, we would like to know if such a model fusion did not create any decision bias that would pre-determine the attribution of a certain ship to a class of strong or weak emitters. For this, we decided to study five variables that are interesting from the point of view of result interpretability. Three of the selected variables (ship length, ship speed, and wind speed) were features of both regression and segmentation models. Another two variables (Year of built – stands for the ship built year, and Average IoU – stands for an average score of Intersection over Union of the *ship sector* of the analyzed with the *ship sectors* of other ships[3]) were not a part of any model[4] but can have a potential influence on the attribution of a ship to a class of weak or strong emitters.

To check the potential presence of decision bias, for each studied ship type, we compared the averages of the above-mentioned features (see Table 6.5) and performed a univariate one-way ANOVA test (Table 6.6), analyzing the statistical significance of the differences between the values of the variables from two groups of ships – strong or weak emitters. From the obtained results, we conclude that none of the analyzed variables had a statistically significant influence on attributing a certain ship to a class of strong or weak emitters. This implies the absence of decision bias related to these variables.

6.5 Discussion

In this Chapter, we presented a method for detecting anomalously NO_2 emitting ships by applying a combination of machine-learning-based methods on TROPOMI instrument data (**RQ8**). The provided methodology is an important step toward the automation of the procedures for the selection of ships that should undergo inspection. The application of satellite data for such a task is a substantial advancement, as satellite-based measurements are the only available tools that can access ship emissions in the open sea.

Another advantage of satellite-based observations in contrast to all the other meth-

---

[3]Given two areas of interest, IoU is computed as the surface of their overlap divided by the surface of their joint area.

[4]The variables were tested in the preliminary phase of our regression model experiments but were removed due to the negative impact on model performance.

ods currently used for ship emission monitoring is that satellite observations enable us to observe the emissions over time regularly and remotely. The presented approach exploits this property of satellite-based observations by making multi-day profiles of ship observations. Such an approach allows us to make conclusions based on aggregated statistics of several ship observations rather than based on a single observation only. The disadvantage of such a statistics-based approach is that only systematic high emitters can be captured.

In order to be able to use the proposed approach on a day-to-day basis some technological advancements are needed. First of all, as we can see from Figures 6.8 and 6.9, the correct and complete segmentation of ship plumes remains a challenging task. Additionally, it is challenging to attribute the detected plume to a certain ship. Both challenges will become more feasible when satellite-based observations with an even higher spatial resolution (for instance, TANGO instrument [67]) become available. Moreover, it is still difficult to fully eliminate signal interference. This is mainly due to the high irregularities of both atmospheric chemistry processes and ship trajectories. Also, the problem will become less significant once the higher-resolution data are available.

Another possible improvement is to account for the dynamics of the atmospheric processes within the methodology. The dynamics of the atmospheric processes affects how fast and how much NO_2 will be created out of emitted NO_x. In this study, we implicitly addressed the atmospheric chemistry processes by using features such as the month the observation took place (seasonability) and solar angle. Explicit modeling such as through the introduction of ozone concentration or air temperature features may provide additional insights.

Finally, at the moment, we do not have access to the ground truth data that would allow us to validate the proposed selection of potentially anomalously emitting ships. As we mentioned at the beginning of this Section, the TROPOMI observations are currently the most complete available source of information regarding emissions of ships in the open sea. Once the proposed approach is implemented into a production environment, the feedback received from inspectors can be used for validation and for further optimization of the method.

6.6   Conclusions

In this Chapter, we applied a combination of machine-learning-based methods on TROPOMI instrument data and presented an approach for automatic identification

of potentially anomalously NO_2 emitting ships. Our approach allows the automatic processing of a huge amount of satellite remote sensing data in order to select for the inspection ships that consistently emit more than can be inferred based on their properties and sailing conditions. With the proposed methodology, the selected cases for inspection are based on multi-day observations of ship emissions. With this, we harvest the main advantage of satellite-based observations over the existing approaches for ship compliance monitoring, with which the decisions have to be made on the basis of a single observation only. The proposed methodology provides a potential path toward the development of a scalable recommendation system for ship inspectors that is based on satellite-based observations.

Chapter 7

Conclusions and future work

In this Chapter, we recapitulate the work presented in this thesis. We summarize the research findings that were presented in the preceding chapters and revisit the research questions outlined in the Introduction chapter. Finally, we present an overview of future research opportunities with regard to the application of the TROPOMI instrument or next-generation satellite-based instruments for the task of monitoring of NO_x emission from shipping.

In Chapter 3, we examined the sensitivity limits of TROPOMI data with respect to the detection of NO_2 plumes from individual seagoing ships. In that Chapter, we addressed three research questions:

- **RQ1:** What is the minimum speed and length of a seagoing ship so that the NO_2 plume from it can be detected with the detection system using TROPOMI data?

- **RQ2:** To what extent can the detectability of NO_2 plumes be improved if only the biggest emitters are taken into account?

- **RQ3:** Is there a potential for improvement of detectability of NO_2 plumes from the slow/small ships if more data were used to train the used classification model?

We addressed the above-mentioned questions with a classification model trained to separate image patches into those with and without ship plumes. We proposed to estimate the detection capabilities of the detection system using TROPOMI data based on parameters such as speed and length of the ship, as those are known to be

good estimates of ship emission potential. The study was performed for four regions of interest: the Mediterranean Sea, Arabian Sea, Biscay Bay, and Bengal Bay.

Addressing **RQ1**, we first demonstrated that the smallest ships in our dataset fall below the detection limit. However, once a certain level of ship speed/size is reached, the ship plume becomes detectable with our method. Subsequently, we calculated that for the Mediterranean Sea and the Arabian Sea, the sensitivity limits of the studied detection system are approximately $1 \times 10^7 m^5/s^3$. For the Biscay Bay, the sensitivity limit is lower and is around $3.8 \times 10^6 m^5/s^3$ (c.f. Figure 3.10). Translating this into ship speed and length, we infer that, for the Mediterranean and Arabian Seas, ships slower than 10 knots or shorter than 150 meters are below the sensitivity limit of the detection system using TROPOMI data. For the Biscay Bay, the limit lies around 8 kt and 100 m. For Bengal Bay, we were not able to estimate the sensitivity limit due to the insufficient amount of data.

With respect to **RQ2**, our results indicate that restricting the analysis to faster/larger ships leads to enhanced detectability of ship plumes. This suggests that focusing on the larger emitters could potentially increase the efficiency of the application and accuracy of ship emission monitoring using the TROPOMI instrument. The analysis also shows differences in model performance between studied regions. We concluded that these variations could be partially attributed to variations in ship traffic density between the regions. Additional factors that potentially can influence the performances of the models are measurement conditions (e.g., number of cloudy days), differences in data quality between regions, and different scales of temperature fluctuations or concentration of ozone in the background.

When addressing **RQ3**, we again encountered the variability of the results across the regions. For the Mediterranean Sea and Biscay Bay, an increase in data volume led to a notable enhancement in model performance. While, for the Arabian Sea and Bengal Bay, the impact of increasing the amount of data, even though present, was less pronounced. One of the reasons was the fact that for European regions we had a higher ratio of data points with a high value of emission proxy in the dataset than for the Bengal Bay and Arabian Sea. Nonetheless, the obtained results indicate that the accuracy of currently determined detection limits is perhaps constrained not only by the methodology or the sensor but also by data availability.

In Chapter 4, we addressed the research questions:

- **RQ4:** How to assign a TROPOMI signal associated with a certain plume to a potential emitting ship?

- **RQ5:** To what extent can the NO_2 plumes be segmented in the TROPOMI data using a simple thresholding method?

To address **RQ4**, we proposed a method for automated assignment of the region of interest (RoI) to the analyzed ship. We called the proposed RoI a *ship sector*. The *ship sector* is defined based on AIS data, information about the speed and direction of the prevailing wind, as well as assumed uncertainties in the speed and direction of the wind data. The *ship sector* is defined such that the plume produced by an analyzed ship will always be located within it. Moreover, the proposed approach provides a possibility for large-scale automatic processing of satellite data for the quantification of emissions from individual seagoing ships. Addressing **RQ5**, we concluded that ship plume segmentation within the assigned *ship sector* requires a more complex method than the linear threshold. Indeed, when comparing the estimated values of NO_2 with the ship emission proxy, we obtained low linear correlation scores, calling for a more complex methodology for ship NO_2 estimation.

Therefore, in Chapter 5, we focused on the development of a supervised machine learning methodology for ship plume segmentation within a *ship sector*. The presented methodology is based on task-specific feature engineering that allows to address ship plume segmentation with a multivariate model. Using the developed methodology, the following research questions were studied:

- **RQ6:** Can we improve the segmentation quality of NO_2 plumes from individual ships using supervised machine learning?

- **RQ7:** Does the machine learning-based segmentation allow for the detection of NO_2 plumes that cannot be recognized visually?

To address **RQ6**, we studied a list of multivariate classifiers of increasing complexity and compared them with the threshold-based benchmarks. With the best (XGBoost) model, we achieved more than 20% increase in the segmentation average precision in comparison to the best benchmark model. We can state that the application of supervised machine learning can indeed improve the quality of ship plume segmentation.

In Chapter 5, we also showed that with an XGBoost model, we can segment more plumes while achieving 6.8% higher correlation with the emission proxy than when the plumes were segmented manually. This result suggests that the machine learning-based segmentation allows for the detection of NO_2 plumes that were not recognized visually (**RQ7**).

Finally, in Chapter 6, we focused on the last research questions:

- **RQ8**: How to identify ships that are potential anomalous emitters using TROPOMI data?

To identify anomalously emitting ships, we presented a method that combines two independently trained machine-learning models. Based on the models' responses, we identify ships that emit more than can be expected based on the operating conditions and characteristics of the ship. To minimize the effects of disruptive events (weather, neighboring another ship, or land outflow), we observed each analyzed ship for an extensive period of time, creating respective ship profiles. Ships that are ranked as highly deviating throughout the time of observation according to both machine learning models, are considered to be potential anomalous emitters and require further attention. If no other explanations can be found, the ships are advised to be the candidates for inspection.

7.1 Future directions

Referring back to the overarching research question of the thesis, we can state that indeed the TROPOMI instrument has the potential to be used for monitoring NO_2 emissions from individual ships on a global scale. Nevertheless, to make such an application industrially operational, several challenges still need to be addressed:

- Attribution of a signal to individual ships in case of overlapping plumes. In this thesis, we reduce this issue by means of the application of multi-day data averaging. This is suitable when drawing conclusions on the basis of long-term observations. However, in order to observe day-to-day changes in the level of emissions produced by a ship, the solution should be adapted.

- The study of other chemical compounds that can be used by machine-learning models as an additional signal of ship plumes. Throughout the thesis, we witnessed that NO_2 plumes from such comparable sources of emissions as ships are often difficult to detect in TROPOMI signals. However, it is known from the literature that the presence of gases such as ozone (O_3) or formaldehyde (HCHO) may be coupled with the presence of ship plumes [27, 44, 53, 2]. The current quality of satellite-based measurements does not allow to distinguish ship traces of these components on daily data. Yet, multi-day averaging confirms that the HCHO signal from ships is present in the data [72, 44], while the concentration of O_3 in the background affects the speed of $NO_x \longrightarrow NO_2$ transformation [19].

- The investigation of causes for regional differences observed in TROPOMI sensitivity with respect to the detection of NO_2 plumes from individual ships. In Chapter 3, we showed that the sensitivity limits of the instrument vary significantly depending on the region under study. We see that it can be explained partially by the traffic density in the region. However, we do not have evidence to state that the latter is the only cause of the differences observed. Other potential sources of regional differences can be varying ozone concentrations, wind speed, or the presence of emitters that are not registered with AIS (e.g. oil rigs or military ships).

We believe that by tackling the aforementioned directions, we can improve the overall understanding of the potential of the application of the TROPOMI instrument with respect to ship emission monitoring.

7.2   Final remarks

Summarizing our conclusions, we can state that with the work presented in this thesis, we notably moved forward the state-of-the-art with respect to the application of satellite observations for the task of continuous and global monitoring of emissions from individual ships. When starting the work of this thesis, all we knew was that some of the plumes from individual ships could be distinguished with the TROPOMI instrument. When reaching the end, we know the sensitivity limits of the detection system of ship NO_2 plumes using TROPOMI data, how to process the TROPOMI signal to extract information about the ship emissions automatically, and how to automatically select ships that are potential anomalous emitters.

The aim of this thesis has been to improve the quality of monitoring of emissions coming from the shipping industry, toward mitigating its adverse environmental effects. A collaborative and continuous effort of professionals across various domains and fields of expertise is essential to ensure that both monitoring of ship emissions and related policies yield the intended outcomes. Indeed, it was noted recently that the implementation of the nitrogen emission control area (NECA) in the North and Baltic Seas in 2021 had little impact on remotely measured NO_x concentrations in European waters [104]. The work presented here is just a piece of a complicated web of required actions, reactions, and decisions necessary to achieve desired environmental outcomes. Nonetheless, we believe that it lays a solid foundation for the future applications of satellite observations for the monitoring of emissions (including but

not limited to NO_2) produced by individual ships. In this manner, it will contribute to mitigating the environmental impact of the shipping industry as a whole.

Bibliography

[1] Tropomi. https://www.tropomi.eu/. Accessed: 2023-10-06.

[2] Juanito Jerrold Mariano Acdan, Robert Bradley Pierce, Angela F Dickens, Zachariah Adelman, and Tsengel Nergui. Examining tropomi formaldehyde to nitrogen dioxide ratios in the lake michigan region: implications for ozone exceedances. *Atmospheric Chemistry and Physics*, 23(14):7867–7885, 2023.

[3] JA Acosta, M Gabarrón, A Faz, S Martínez-Martínez, R Zornoza, and JM Arocena. Influence of population density on the concentration and speciation of metals in the soil and street dust from urban areas. *Chemosphere*, 134:328–337, 2015.

[4] Harshit Agrawal, Quentin GJ Malloy, William A Welch, J Wayne Miller, and David R Cocker III. In-use gaseous and particulate matter emissions from a modern ocean going container vessel. *Atmospheric Environment*, 42(21):5504–5510, 2008.

[5] Luc Anselin. Local indicators of spatial association—lisa. *Geographical analysis*, 27(2):93–115, 1995.

[6] Gopi Battineni, Getu Gamo Sagaro, Nalini Chinatalapudi, and Francesco Amenta. Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of personalized medicine*, 10(2):21, 2020.

[7] Jörg Beecken, Johan Mellqvist, Kent Salo, Johan Ekholm, and J-P Jalkanen. Airborne emission measurements of so 2, no x and particles from individual ships using a sniffer technique. *Atmospheric Measurement Techniques*, 7(7):1957–1968, 2014.

[8] S Beirle, U Platt, R Von Glasow, M Wenig, and T Wagner. Estimate of nitrogen oxide emissions from shipping by satellite remote sensing. *Geophysical Research Letters*, 31(18), 2004.

[9] Nicholas Berg, Johan Mellqvist, J-P Jalkanen, and J Balzani. Ship emissions of so 2 and no 2: Doas measurements from airborne platforms. *Atmospheric Measurement Techniques*, 5(5):1085–1098, 2012.

[10] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, 13(null):281–305, feb 2012.

[11] K Folkert Boersma, Henk J Eskes, Andreas Richter, Isabelle De Smedt, Alba Lorente, Steffen Beirle, Jos HGM Van Geffen, Marina Zara, Enno Peters, Michel Van Roozendael, et al. Improving algorithms and uncertainty estimates for satellite no 2 retrievals: results from the quality assurance for the essential climate variables (qa4ecv) project. *Atmospheric Measurement Techniques*, 11(12):6651–6678, 2018.

[12] K Folkert Boersma, Geert CM Vinken, and Jean Tournadre. Ships going slow in reducing their nox emissions: changes in 2005–2012 ship exhaust inferred from satellite measurements over europe. *Environmental Research Letters*, 10(7):074007, 2015.

[13] Heinrich Bovensmann, JP Burrows, M Buchwitz, Johannes Frerick, Suresh Noel, VV Rozanov, KV Chance, and APH Goede. Sciamachy: Mission objectives and measurement modes. *Journal of the atmospheric sciences*, 56(2):127–150, 1999.

[14] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[15] Molly E Brown, David J Lary, Anton Vrieling, Demetris Stathakis, and Hamse Mussa. Neural networks as a tool for constructing continuous ndvi time series from avhrr and modis. *International Journal of Remote Sensing*, 29(24):7141–7158, 2008.

[16] Bert Brunekreef and Stephen T Holgate. Air pollution and health. *The lancet*, 360(9341):1233–1242, 2002.

[17] John P Burrows, Mark Weber, Michael Buchwitz, Vladimir Rozanov, Annette Ladstätter-Weißenmayer, Andreas Richter, Rüdiger DeBeek, Ricarda Hoogen, Klaus Bramstedt, Kai-Uwe Eichmann, et al. The global ozone monitoring experiment (gome): Mission concept and first scientific results. *Journal of the Atmospheric Sciences*, 56(2):151–175, 1999.

[18] Gavin C Cawley and Nicola LC Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, 11:2079–2107, 2010.

[19] Siddika Celik, Frank Drewnick, Friederike Fachinger, James Brooks, Eoghan Darbyshire, Hugh Coe, Jean-Daniel Paris, Philipp G Eger, Jan Schuladen, Ivan Tadic, et al. Influence of vessel characteristics and atmospheric processes on the gas and particle phase of ship emission plumes: in situ measurements in the mediterranean sea and around the arabian peninsula. *Atmospheric Chemistry and Physics*, 20(8):4713–4734, 2020.

[20] Ka Lok Chan, Ehsan Khorsandi, Song Liu, Frank Baier, and Pieter Valks. Estimation of surface no2 concentrations over germany from tropomi satellite observations using a machine learning method. *Remote Sensing*, 13(5):969, 2021.

[21] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27, 2011.

[22] 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*, pages 785–794, 2016.

[23] James J Corbett, James J Winebrake, Erin H Green, Prasad Kasibhatla, Veronika Eyring, and Axel Lauer. Mortality from ship emissions: a global assessment. *Environmental science & technology*, 41(24):8512–8518, 2007.

[24] Monica Crippa, Diego Guizzardi, Marilena Muntean, Edwin Schaaf, Frank Dentener, John A Van Aardenne, Suvi Monni, Ulrike Doering, Jos GJ Olivier, Valerio Pagliari, et al. Gridded emissions of air pollutants for the period 1970–2012 within edgar v4. 3.2. *Earth Syst. Sci. Data*, 10(4):1987–2013, 2018.

[25] Paul J Crutzen. The influence of nitrogen oxides on atmospheric ozone content. *Paul J. Crutzen: A Pioneer on Atmospheric Chemistry and Climate Change in the Anthropocene*, pages 108–116, 2016.

[26] Martijn de Ruyter de Wildt, Henk Eskes, and K Folkert Boersma. The global economic cycle and satellite-derived no2 trends over shipping lanes. *Geophysical research letters*, 39(1), 2012.

[27] Isabelle De Smedt, Gaia Pinardi, Corinne Vigouroux, Steven Compernolle, Alkis Bais, Nuria Benavent, Folkert Boersma, Ka-Lok Chan, Sebastian Donner, Kai-Uwe Eichmann, et al. 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.

[28] Galina Deeva, Daria Bogdanova, Estefanía Serral, Monique Snoeck, and Jochen De Weerdt. A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education*, 162:104094, 2021.

[29] Ruud J Dirksen, K Folkert Boersma, Henk J Eskes, Dmitry V Ionov, Eric J Bucsela, Pieternel F Levelt, and Hennie M Kelder. Evaluation of stratospheric no2 retrieved from the ozone monitoring instrument: Intercomparison, diurnal cycle, and trending. *Journal of Geophysical Research: Atmospheres*, 116(D8), 2011.

[30] John Douros, Henk Eskes, Jos van Geffen, K Folkert Boersma, Steven Compernolle, Gaia Pinardi, Anne-Marlene Blechschmidt, Vincent-Henri Peuch, Augustin Colette, and Pepijn Veefkind. Comparing sentinel-5p tropomi no 2 column observations with the cams regional air quality ensemble. *Geoscientific Model Development*, 16(2):509–534, 2023.

Bibliography

[31] 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 nitrogendioxide. Technical Report S5P-KNMI-L2-0021-MA, July 2022.

[32] Pedregosa Fabian. Scikit-learn: Machine learning in python. *Journal of machine learning research 12*, page 2825, 2011.

[33] Qianzhu Fan, Yan Zhang, Weichun Ma, Huixin Ma, Junlan Feng, Qi Yu, Xin Yang, Simon KW Ng, Qingyan Fu, and Limin Chen. Spatial and seasonal dynamics of ship emissions over the yangtze river delta and east china sea and their potential environmental influence. *Environmental science & technology*, 50(3):1322–1329, 2016.

[34] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *the Journal of machine Learning research*, 9:1871–1874, 2008.

[35] Tom Fawcett. Roc graphs: Notes and practical considerations for researchers. *Machine learning*, 31(1):1–38, 2004.

[36] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[37] Bruno Ferreira, Muriel Iten, and Rui G Silva. Monitoring sustainable development by means of earth observation data and machine learning: A review. *Environmental Sciences Europe*, 32(1):1–17, 2020.

[38] Douglas P Finch, Paul I Palmer, and Tianran Zhang. Automated detection of atmospheric no 2 plumes from satellite data: a tool to help infer anthropogenic combustion emissions. *Atmospheric Measurement Techniques*, 15(3):721–733, 2022.

[39] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.

[40] Jerome H Friedman and Bogdan E Popescu. Predictive learning via rule ensembles. 2008.

[41] Aristeidis K Georgoulias, K Folkert Boersma, Jasper van Vliet, Xiumei Zhang, Prodromos Zanis, Jos de Laat, et al. Detection of no2 pollution plumes from individual ships with the tropomi/s5p satellite sensor. *Environmental Research Letters*, 15(12):124037, 2020.

[42] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow.* " O'Reilly Media, Inc.", 2022.

[43] Arthur Getis and Jared Aldstadt. Constructing the spatial weights matrix using a local statistic. *Geographical analysis*, 36(2):90–104, 2004.

[44] GS Gopikrishnan and Jayanarayanan Kuttippurath. A decade of satellite observations reveal significant increase in atmospheric formaldehyde from shipping in indian ocean. *Atmospheric Environment*, 246:118095, 2021.

[45] Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on tabular data?, 2022.

[46] Robert Guicherit and Dick van den Hout. The global nox cycle. In *Air Pollution by Nitrogen Oxides*, pages 15–19. Elsevier Scientific Publishing Company Amsterdam, 1982.

[47] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[48] R Hoogerbrugge. Grootschalige concentratie- en depositiekaarten nederland. rapportage 2022, 2022.

[49] Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren. *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.

[50] IMO. Amendments to the annex of the protocol of 1978 relating to the international convention for the prevention of pollution from ships, 1997.

[51] IMO. Marpol annex vi - regulation 13, 2020.

[52] Lasse Johansson, Jukka-Pekka Jalkanen, and Jaakko Kukkonen. Global assessment of shipping emissions in 2015 on a high spatial and temporal resolution. *Atmospheric Environment*, 167:403–415, 2017.

[53] Matthew S Johnson, Sajeev Philip, Rajesh Kumar, Aaron Naeger, Amir H Souri, Jeffrey Geddes, Laura Judd, Scott Janz, and John Sullivan. Satellite remote-sensing capability to assess tropospheric column ratios of formaldehyde and nitrogen dioxide: case study during the listos 2018 field campaign. *Atmospheric Measurement Techniques Discussions*, 2022:1–41, 2022.

[54] Marilena Kampa and Elias Castanas. Human health effects of air pollution. *Environmental pollution*, 151(2):362–367, 2008.

[55] Marios Kefalas. *Data-driven predictive maintenance and time-series applications*. PhD thesis, Leiden University, 2023.

[56] Shristi Shakya Khanal, PWC Prasad, Abeer Alsadoon, and Angelika Maag. A systematic review: machine learning based recommendation systems for e-learning. *Education and Information Technologies*, 25:2635–2664, 2020.

[57] Lars Kotthoff, Chris Thornton, Holger H Hoos, Frank Hutter, and Kevin Leyton-Brown. Auto-weka: Automatic model selection and hyperparameter optimization in weka. In *Automated machine learning*, pages 81–95. Springer, Cham, 2019.

[58] John R Koza. Genetic programming as a means for programming computers by natural selection. *Statistics and computing*, 4(2):87–112, 1994.

[59] K. Krause, F. Wittrock, A. Richter, S. Schmitt, D. Pöhler, A. Weigelt, and J. P. Burrows. Estimation of ship emission rates at a major shipping lane by long-path doas measurements. *Atmospheric Measurement Techniques*, 14(8):5791–5807, 2021.

[60] Solomiia Kurchaba, Artur Sokolovsky, Jasper Van Vliet, Fons J. Verbeek, and Cor J. Veenman. Sensitivity analysis for the detection no2 plumes from seagoing ships using tropomi data. reproducibility package, 2024.

[61] 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.

[62] 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 *29th International Conference on Advances in Geographic Information Systems*, pages 454–457, 2021.

[63] 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.

[64] 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.

[65] Sampo Kuutti, Richard Bowden, Yaochu Jin, Phil Barber, and Saber Fallah. A survey of deep learning applications to autonomous vehicle control. *IEEE Transactions on Intelligent Transportation Systems*, 22(2):712–733, 2020.

[66] Daniel A Lack, James J Corbett, Timothy Onasch, Brian Lerner, Paola Massoli, Patricia K Quinn, Timothy S Bates, David S Covert, Derek Coffman, Berko Sierau, et al. Particulate emissions from commercial shipping: Chemical, physical, and optical properties. *Journal of Geophysical Research: Atmospheres*, 114(D7), 2009.

[67] Jochen Landgraf, Stephanie Rusli, Ryan Cooney, Pepijn Veefkind, Tim Vemmix, Zeger de Groot, Andrew Bell, James Day, Anton Leemhuis, and Bernd Sierk. The tango mission: A satellite tandem to measure major sources of anthropogenic greenhouse gas emissions. In *EGU General Assembly Conference Abstracts*, page 19643, 2020.

[68] David J Lary, LA Remer, Devon MacNeill, Bryan Roscoe, and Susan Paradise. Machine learning and bias correction of modis aerosol optical depth. *IEEE Geoscience and Remote Sensing Letters*, 6(4):694–698, 2009.

[69] Pieternel F Levelt, Ernest Hilsenrath, Gilbert W Leppelmeier, Gijsbertus HJ van den Oord, Pawan K Bhartia, Johanna Tamminen, Johan F de Haan, and J Pepijn Veefkind. Science objectives of the ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing*, 44(5):1199–1208, 2006.

[70] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[71] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z Sheng, Hui Xiong, and Leman Akoglu. A comprehensive survey on graph anomaly detection with deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[72] T Marbach, S Beirle, U Platt, P Hoor, F Wittrock, A Richter, M Vrekoussis, M Grzegorski, JP Burrows, and T Wagner. Satellite measurements of formaldehyde linked to shipping emissions. *Atmospheric Chemistry and Physics*, 9(21):8223–8234, 2009.

[73] Robert McLaren, Patryk Wojtal, Jamie D Halla, Cris Mihele, and Jeffrey R Brook. A survey of no2: So2 emission ratios measured in marine vessel plumes in the strait of georgia. *Atmospheric environment*, 46:655–658, 2012.

[74] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., USA, 1 edition, 1997.

[75] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

[76] Jun Min Mou, Cees Van der Tak, and Han Ligteringen. Study on collision avoidance in busy waterways by using ais data. *Ocean Engineering*, 37(5-6):483–490, 2010.

[77] Randal S Olson, Nathan Bartley, Ryan J Urbanowicz, and Jason H Moore. Evaluation of a tree-based pipeline optimization tool for automating data science. In *Genetic and evolutionary computation conference*, pages 485–492, 2016.

[78] J Keith Ord and Arthur Getis. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis*, 27(4):286–306, 1995.

[79] Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer. Deep learning for financial applications: A survey. *Applied Soft Computing*, 93:106384, 2020.

[80] 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.

[81] Liisa Pirjola, Aki Pajunoja, J Walden, J-P Jalkanen, T Rönkkö, A Kousa, and T Koskentalo. Mobile measurements of ship emissions in two harbour areas in finland. *Atmospheric Measurement Techniques*, 7(1):149–161, 2014.

[82] FT Princiotta et al. Stationary source nox control technology overview. *Studies in Environmental Science*, 21:737–745, 1982.

[83] Prabhat Kumar Rai. Biodiversity of roadside plants and their response to air pollution in an indo-burma hotspot region: implications for urban ecosystem restoration. *Journal of Asia-Pacific Biodiversity*, 9(1):47–55, 2016.

[84] Veerabhadran Ramanathan and Yan Feng. Air pollution, greenhouse gases and climate change: Global and regional perspectives. *Atmospheric environment*, 43(1):37–50, 2009.

[85] Andreas Richter, Veronika Eyring, John P Burrows, Heinrich Bovensmann, Axel Lauer, Bernd Sierk, and Paul J Crutzen. Satellite measurements of no2 from international shipping emissions. *Geophysical Research Letters*, 31(23), 2004.

[86] Tobias Christoph Valentin Werner Riess, Klaas Folkert Boersma, Jasper Van Vliet, Wouter Peters, Maarten Sneep, Henk Eskes, and Jos Van Geffen. Improved monitoring of shipping no 2 with tropomi: decreasing no x emissions in european seas during the covid-19 pandemic. *Atmospheric Measurement Techniques*, 15(5):1415–1438, 2022.

[87] M Salim Akhter and Ismail M Madany. Heavy metals in street and house dust in bahrain. *Water, Air, and Soil Pollution*, 66:111–119, 1993.

[88] SF Schreier, E Peters, A Richter, J Lampel, F Wittrock, and JP Burrows. Ship-based max-doas measurements of tropospheric no2 and so2 in the south china and sulu sea. *Atmospheric Environment*, 102:331–343, 2015.

[89] Berend J Schuit, Joannes D Maasakkers, Pieter Bijl, Gourav Mahapatra, Anne-Wil Van den Berg, Sudhanshu Pandey, Alba Lorente, Tobias Borsdorff, Sander Houweling, Daniel J Varon, et al. Automated detection and monitoring of methane super-emitters using satellite data. *Atmospheric Chemistry and Physics*, 23(16):9071–9098, 2023.

[90] SCIPPER. Shipping contributions to inland pollution push for the enforcement of regulations, 2020.

[91] Himan Shahabi, Ataollah Shirzadi, Kayvan Ghaderi, Ebrahim Omidvar, Nadhir Al-Ansari, John J Clague, Marten Geertsema, Khabat Khosravi, Ata Amini, Sepideh Bahrami, et al. Flood detection and susceptibility mapping using sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on k-nearest neighbor classifier. *Remote Sensing*, 12(2):266, 2020.

[92] Tej Bahadur Shahi, Cheng-Yuan Xu, Arjun Neupane, and William Guo. Recent advances in crop disease detection using uav and deep learning techniques. *Remote Sensing*, 15(9):2450, 2023.

[93] Maarten Sneep. Sentinel 5 precursor/tropomi knmi and sron level 2 input output data definition. Technical Report S5P-KNMI-L2-0009-SD, March 2021.

[94] CH Song, G Chen, SR Hanna, J Crawford, and DD Davis. Dispersion and chemical evolution of ship plumes in the marine boundary layer: Investigation of o3/noy/hox chemistry. *Journal of Geophysical Research: Atmospheres*, 108(D4), 2003.

[95] Jianing Song, Duarte Rondao, and Nabil Aouf. Deep learning-based spacecraft relative navigation methods: A survey. *Acta Astronautica*, 191:22–40, 2022.

[96] Mervyn Stone. Cross-validation and multinomial prediction. *Biometrika*, 61(3):509–515, 1974.

[97] Swapan Talukdar, Pankaj Singha, Susanta Mahato, Swades Pal, Yuei-An Liou, and Atiqur Rahman. Land-use land-cover classification by machine learning classifiers for satellite observations—a review. *Remote Sensing*, 12(7):1135, 2020.

[98] United Nations: Department of Economic and Social Affairs: Population Division. *World urbanization prospects.* United Nations, New York, NY, November 2019.

[99] Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, Anant Madabhushi, Parantu Shah, Michaela Spitzer, et al. Applications of machine learning in drug discovery and development. *Nature reviews Drug discovery*, 18(6):463–477, 2019.

[100] Peter van der Maas. *Chemically Enhanced Biological NO x Removal from Flue Gases: Nitric Oxide and Ferric EDTA Reduction in BioDeNOx Reactors.* Wageningen University and Research, 2005.

[101] Jos van Geffen, Henk Eskes, Folkert Boersma, and J. Pepijn Veefkind. Tropomi atbd of the total and tropospheric no2 data products. Technical Report S5P-KNMI-L2-0005-RP, July 2022.

[102] W Van Roy and K Scheldeman. Results marpol annex vi monitoring report: Belgian sniffer campaign 2016. 2016.

[103] Ward Van Roy, Kobe Scheldeman, Benjamin Van Roozendael, Annelore Van Nieuwenhove, Ronny Schallier, Laurence Vigin, and Frank Maes. Airborne monitoring of compliance to nox emission regulations from ocean-going vessels in the belgian north sea. *Atmospheric Pollution Research*, 13(9):101518, 2022.

[104] 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.

[105] GCM Vinken. *From ships to soil microbes: NOx emission estimates using OMI satellite retrievals and an improved modelling framework*. PhD thesis, Eindhoven University of Technology, 2014.

[106] GCM Vinken, KF 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.

[107] 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.

[108] Mei Wang and Weihong Deng. Deep face recognition: A survey. *Neurocomputing*, 429:215–244, 2021.

[109] Shuo Wang. *Ensemble diversity for class imbalance learning*. PhD thesis, University of Birmingham, 2011.

[110] Wenhao Wang, Xiong Liu, Jianzhao Bi, and Yang Liu. A machine learning model to estimate ground-level ozone concentrations in california using tropomi data and high-resolution meteorology. *Environment International*, 158:106917, 2022.

[111] Taofeek Akangbe Yekeen, Xijin Xu, Yuling Zhang, Yousheng Wu, Stephani Kim, Tiina Reponen, Kim N Dietrich, Shuk-mei Ho, Aimin Chen, and Xia Huo. Assessment of health risk of trace metal pollution in surface soil and road dust from e-waste recycling area in china. *Environmental Science and Pollution Research*, 23:17511–17524, 2016.

[112] Xin Zhang, Ya'nan Zhou, and Jiancheng Luo. Deep learning for processing and analysis of remote sensing big data: A technical review. *Big Earth Data*, 6(4):527–560, 2022.

[113] Yuzhe Zhao, Yujun Fan, Kjetil Fagerholt, and Jingmiao Zhou. Reducing sulfur and nitrogen emissions in shipping economically. *Transportation Research Part D: Transport and Environment*, 90:102641, 2021.

List of publications

While working towards this thesis, the following contributions were made:

1. Kurchaba, S., Sokolovsky, A., van Vliet, J., Verbeek, F.J., Veenman, C.J., 2024. Sensitivity analysis for the detection of NO_2 plumes from seagoing ships using TROPOMI data. Remote Sensing of Environment 304, 114041. doi:10.1016/j.rse.2024.114041.

2. Kurchaba, S., van Vliet, J., Meulman, J.J., Verbeek, F.J., Veenman, C.J., 2021. Improving evaluation of NO_2 emission from ships using spatial association on TROPOMI satellite data, in: 29th International Conference on Advances in Geographic Information Systems, pp. 454–457. doi:10.1145/3474717.3484213.

3. Kurchaba, S., van Vliet, J., Verbeek, F.J., Meulman, J.J., Veenman, C.J., 2022. Supervised segmentation of NO_2 plumes from individual ships using TROPOMI satellite data. Remote Sensing 14. doi:10.3390/rs14225809.

4. Kurchaba, S., van Vliet, J., Verbeek, F.J., Veenman, C.J., 2023. Anomalous NO_2 emitting ship detection with TROPOMI satellite data and machine learning. Remote Sensing of Environment 297, 113761. doi:10.1016/j.rse.2023.113761.

List of publications

Samenvatting

De zeevaartsector is een van de grootste vervuilers waar het gaat om de uitstoot van stikstof-oxides (NO_x). Dit is een groep van vervuilende stoffen die zeer schadelijk zijn voor ecosystemen en humane gezondheid. De afgelopen 20 jaar is de vervuiling veroorzaakt door electriciteitscentrales, de zware industrie en het autoverkeer constant afgenomen. Deze trend zien we niet terug in de zeevaart, integendeel, we zien in feite dat de bijdrage van de scheepvaart aan de vervuiling continu toeneemt. Deze opwaartse trend zorgt voor een sterke druk vanuit de maatschappij die erin geresulteerd heeft dat er nieuwe regelgeving is voorgesteld vanuit de "International Maritime Organisation" (IMO). Deze voorschriften leggen resticties op aan individuele schepen met betrekking tot het niveau van uitstoot (emissie) dat een schip kan produceren. Er bestaan verschillende methodes voor het monitoren van uitstoot voor kustvaart en voor schepen in havengebieden, terwijl het monitoren van schepen op open zee tot op heden onhaalbaar is gebleken. De grote verandering in dit verband is het TROPOMI instrument dat is ingebouwd in de Sentinel 5 Precursor Satelliet - studies hebben aangetoond at NO_2 rookpluimen van sommige zeegaande schepen op TROPOMI beeldmateriaal kunnen worden onderscheiden. Het doel van het onderzoek dat in dit proefschrift wordt gepresenteerd is de mogelijkheden te laten zien van het TROPOMI instrument voor de naleving de van de regelgeving door IMO opgesteld voor zeegaande schepen. Ons doel wordt bereikt door het inzetten van innovatieve combinaties van geavanceerde methodes uit het machinaal leren, kenmerk extractie en data integratie. Ieder hoofdstuk uit dit proefschrift bouwt voort op de bevindingen uit het voorgaande hoofdstuk; het geheel van de hoofdstukken representeert de nieuwste-van-het-nieuwste kennis met betrekking tot het toepassen van TROPOMI satelliet data voor het monitoren van NO_2 uitstoot van individuele zeegaande schepen. In de eerste twee hoofdstukken introduceren we het werkveld en achtergrondinformatie relevant voor ons onderzoek. Vervolgens ontwikkelen we in Hoofdstuk 3 een methodologie gebaseerd op machinaal-

leren en we gebruiken deze methodologie om de grenzen van de gevoeligheid van het detectiesysteem uit TROPOMI data te onderzoeken met betrekking tot de het detecteren van NO_2-rookpluimen van schepen. De inzichten uit dit hoofdstuk bepalen de reikwijdte van het verdere onderzoek in dit proefschrift. In Hoofdstuk 4 presenteren we een methode voor het automatische afbakening van een TROPOMI-beeld dat correspondeert met een schip onder aandacht. Door hiervan gebruik te maken kunnen delen waarin geen informatie van belang voorkomt worden uitgesloten van de analyse, terwijl de specifieke aandacht kan gaan naar dat deel van van het TROPOMI-beeld waar de lokatie van de rookpluim van een schip kan worden verwacht. Daaropvolgend presenteren we in Hoofdstuk 5 een methode gebaseerd op machinaal-leren voor de automatische segmentatie van NO_2-rookpluimen zoals die zijn geproduceerd door individuele schepen. De resultaten die hiermee worden verkregen suggereren dat het gebruik van machinaal leren voor de taak van rookpluim detectie (segmentatie) ons in staat stelt de rookpluim correct te detecteren; anderszins zijn deze rookpluimen moeilijk tot niet detecteerbaar voor het menselijk oog. Tenslotte, presenteren we in Hoofdstuk 6 een methode voor de automatische detectie van schepen die potentieel abnormale uitstoot produceren. Onze aanpak maakt het mogelijk grote hoeveelheden "remote sensing" satelliet data geautomatiseerd verwerken teneinde schepen te kunnen selecteren die consistent meer uitstoot produceren dan op basis van eigenschappen van een schip en vaarcondities kan worden afgeleid. De voorgestelde methode voorziet in een mogelijke benadering voor het ontwikkelen van een schaalbaar systeem voor het verstrekken van aanbevelingen voor scheep-inspecteurs gebaseerd op satelliet observaties. Concluderend, kunnen we stellen dat met het onderzoek en de resultaten gepresenteerd in dit proefschrift, we opmerkelijke stappen voorwaarts gezet hebben betreffende vernieuwingen in het inzetten van satelliet-observaties voor de taak van het continue en wereldwijd monitoren van de NO_2-uitstoot van individuele schepen op open zee. Bij aanvang van het onderzoek voor dit proefschrift wisten we niet meer dan dat sommige rookpluimen met het TROPOMI instrument zouden kunnen worden onderscheiden. Maar bij de afsluiting van dit onderzoek hebben we inzicht verkregen in de beperkingen van de gevoeligheid van het detectiesysteem van het TROPOMI-instrument met betrekking tot NO_2-pluimen van schepen, weten we hoe het TROPOMI-signaal moet worden verwerkt teneinde automatisch informatie te verkrijgen over de uitstoot van schepen, en hoe we automatisch schepen kunnen selecteren met een potentieel abnormale uitstoot. De door ons beschreven vooruitgang legt een solide basis voor toekomstige toepassing van satelliet-gebaseerde technologie in het wereldwijd en continue monitoren van antropogene uitstoot.

Samenvatting

Summary

The marine shipping industry is one of the strongest emitters of nitrogen oxides (NO_x), a pollutant detrimental to both ecology and human health. Over the last 20 years, the pollution produced by power plants, the industry sector, and automobile vehicles has been constantly decreasing. In contrast, the pollution impact of maritime transport persists and continues to increase. This generates a big societal pressure, resulting in regulations proposed by the International Maritime Organization. These regulations impose restrictions on emission levels that can be produced by individual ships. While various methods are used to assess the emission from ships in ports and off-coastal areas, monitoring over the open sea has been infeasible until now. The game-changer is the TROPOMI instrument on board the Sentinel 5 Precursor Satellite – studies show that NO_2 plumes from some individual seagoing ships can be distinguished on TROPOMI images. The objective of this thesis is to pave the way toward the application of the TROPOMI instrument for the monitoring of compliance of seagoing ships with the regulations of the International Maritime Organization. This is being achieved through the innovative fusion of the methods of advanced machine learning, feature engineering, and data integration. Each chapter of this thesis builds upon the findings of its predecessor, as a whole, representing state-of-the-art knowledge in the application of TROPOMI satellite data for the monitoring of NO_2 emissions from individual seagoing ships.

Following the introductory chapters, in Chapter 3, using a developed machine learning-based methodology, we examine the sensitivity limits of the detection system using TROPOMI data concerning the detection of NO_2 plumes from ships. The insights gained in this chapter establish the scope for the rest of this study. In Chapter 4, we present a method for automated delineation of a part of a TROPOMI image that corresponds to a ship under study. This way, one can exclude unnecessary pieces of information from the analysis and focus attention on the parts of a TROPOMI

image, where the ship plume is expected to be located. Such an advancement opens a possibility for a large-scale processing of the ship plume data. In Chapter 5, we presented a machine-learning-based method for automated segmentation of NO_2 plumes produced by individual ships. The results presented in the Chapter suggest that by using machine learning for the task of ship plume segmentation we are able to correctly segment plumes that are hardly or not-at-all detectable by the human eye. Lastly, in Chapter 6, we present a methodology for the automated detection of potentially anomalously emitting ships. The presented approach allows the automatic processing of a huge amount of satellite remote sensing data in order to select for the inspection ships that consistently emit more than can be inferred based on their properties and sailing conditions. The proposed methodology provides a potential path toward the development of a scalable recommendation system for ship inspectors that is rooted in satellite-based observations.

To conclude, we can state that with the work presented in this thesis, we notably moved forward with the state-of-the-art concerning the application of satellite observations for the task of continuous and global monitoring of emissions from individual ships. At the onset of the work for this thesis research, all we knew was that some of the plumes from individual ships could be distinguished with the TROPOMI instrument. When reaching the end, we know the sensitivity limits of the detection system of ship NO_2 plumes using TROPOMI data, how to process the TROPOMI signal to extract information about the ship emissions automatically, and how to automatically select ships that are potential anomalous emitters. These advancements lay a solid foundation for the future of the application of satellite-based technologies for continuous monitoring of anthropogenic emissions on a global scale.

Summary

Acknowledgements

"Differences of habit and language are nothing at all if our aims are identical and our hearts are open."

Albus Dumbledore, Harry Potter and the Goblet of Fire

Here, I would like to say thank you to everyone without whom this PhD would never be possible. First of all, I would like to say thank you to my supervision team: Cor Veenman, Jasper van Vliet, Fons Verbeek, and Jacqueline Meulman. Thank you Cor, for your always timely feedback, critical opinions, but also patience and willingness to change and adjust when needed. Thanks, Jasper, for always being there, ready to help. Your open-mindedness and adventurous spirit alongside your willingness to understand every detail of the project made it a pleasure to work with you. Thanks, Fons, for your support and great care that I always have everything I need. Thanks for always being there with a solution and understanding, when yet another "it's a disaster" situation needed to be resolved. I would like to say thank you to Jacqueline for your support and faith in me from the first day we met. It is rare and extremely valuable when a supervisor becomes more like a friend.

It was a great pleasure to spend four years of my PhD in such a friendly and welcoming environment as LIACS. I would like to say thank you to my colleagues, who made me feel that together we are a part of something bigger. I would like to thank Antonio and Gerrit-Jan – you were the first PhDs I met when I first came to Leiden. You made me feel a part of a LIACS family even before I was officially accepted. I would like to thank Marie, Daniela, Sander, Koen, Lieuwe, and Theodoros – thanks to you, even starting during the lockdown, I did not feel lonely and socially isolated. My big thanks to Lea for all our adventures and unforgettable moments. I wish Innsbruck was not that far and we could see each other more often. I also would like to send my words of gratitude to Thomas, Jan, Rohola, and Lu - I always enjoyed our conversations and it's difficult to underestimate how much your wise bits

Acknowledgements

of advice helped me to navigate the way towards the PhD. I would like to say thank you to Gerrit-Jan, Arina, and Qi for your enthusiasm, help, and the fun we had when together organizing LIACS PhD seminars. Finally, I would like to say thank you to my colleagues from office 115 - Erick, Jia, Irene, Danyi, Mike, and Yi. Thanks to you, work there always felt cozy and fun.

I would also like to say thank you to my colleagues outside LIACS. First of all, thank you, Laura - I was so lucky to work with you on that Data Science course. I am very grateful the work on a course grew into a friendship. I would like to say thank you to Christoph for being a teammate on this journey toward understanding TROPOMI data and ship plumes. So many times your knowledge of atmospheric chemistry helped me to move this research forward. I would like to thank Artur for your willingness to help in a difficult moment and for our endless conversations that started way before I began my PhD, nevertheless, had a great influence on leading me towards this point. Finally, I would like to say thank you to the whole team of IDlab, ILT for your support throughout my PhD journey.

A special place in this Chapter belongs to Charles. I can't imagine how my PhD would turn out if I hadn't met you. Thank you for always being there for me, for all the support, and encouragement, and for not letting me give up at the moment of weakness. I am so lucky to have you in my life.

Last, but most I would like to say thank you to my mom and dad. You are my example of strength and perseverance even in the most difficult times. I am eternally grateful for all the opportunities you gave me throughout my life and all the sacrifices it cost you. Thanks for believing in me, sometimes even more than I believe in myself.

This thesis I dedicate to my grandpa. Unfortunately, you won't be able to read it, but without you, it would never happened. It was you, who saw in your little girl a future scientist, showing how fascinating and beautiful the world of problem-solving, experiments, and searching for answers could be. Thank you for that, you are always in my heart.

Curriculum Vitae

"- Who are you? - said the Caterpillar.
- I hardly know, Sir, just at present—at least I know who I was when I got up this
morning, but I think I must have been changed several times since then."
Alice's Adventures in Wonderland

Solomiia Kurchaba was born on November 24, 1995, in Lviv, Ukraine. In 2013, she graduated from the Lviv Physics and Mathematics Lyceum. She then started her studies at the University of Silesia in Katowice, Poland. During her bachelor's studies, she was a part of the Multidisciplinary College for Individual Studies combining econophysics as a major and financial mathematics as a minor. She also participated in Erasmus+ student exchange program and spent one semester of her studies at Universidade do Minho, in Braga, Portugal. In 2016, she obtained her bachelor's degree cum laude in Econophysics. In 2018, she completed her Master's studies in theoretical physics, cum laude. During her studies, she was a recipient of the Stephan Banach scholarship - a Polish government scholarship for international students in STEM. During her Master's she also enrolled in a summer research internship at the Machine Learning Group of the Universite Libre de Bruxelles in Brussels, Belgium. She then started her job as a data scientist in Storkjet – a startup, where her job was to develop machine-learning-based solutions for aircraft performance optimization. During her work in the company, the developed solutions were implemented into the production systems of several big European and American airlines. In 2019, she started her PhD studies at Warsaw University of Technology. However, after one semester she decided to change her PhD program. In May 2020, despite COVID restrictions, she moved to Leiden, The Netherlands. Starting from May 2024, she is a postdoctoral researcher at SRON Netherlands Institute for Space Research.