Machine learning-based NO2 estimation from seagoing ships using TROPOMI/S5P satellite data

Kurchaba, S.

Chapter 5

Ship plume segmentation with supervised machine learning

Based on: 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.

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.

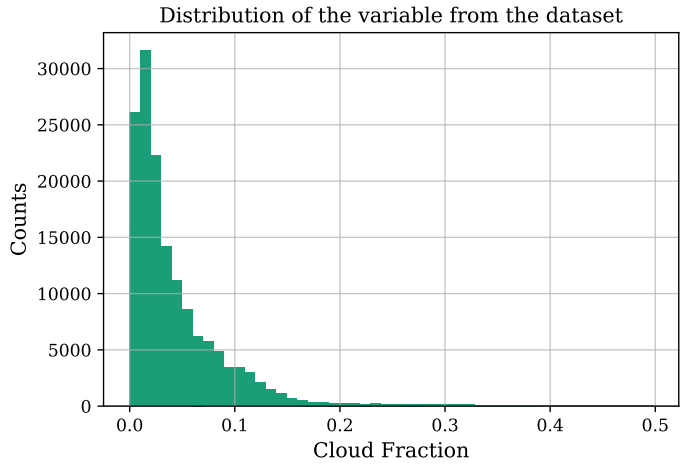

Figure 5.1: Distribution of the variable *cloud fraction* for the dataset used in this study.

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°].

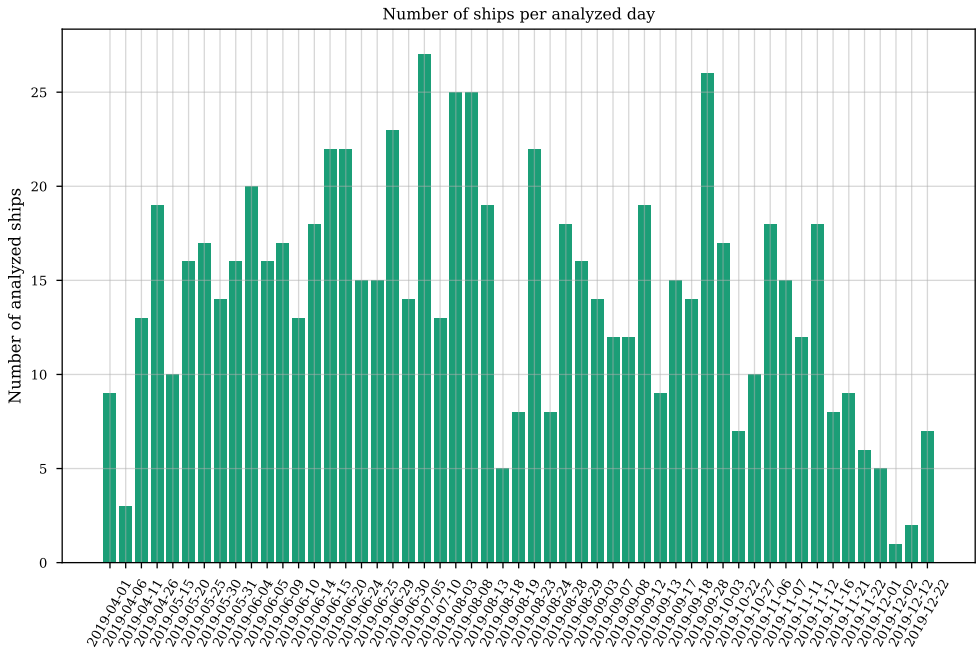

Figure 5.2: A list of days used for the dataset creation and the number of ships per day studied.

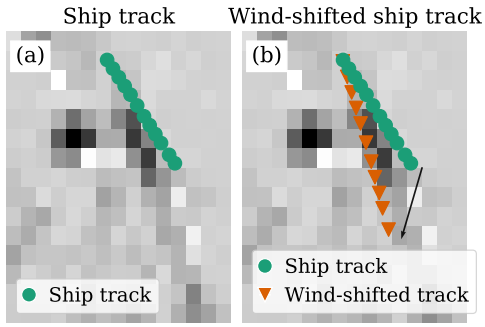

Figure 5.3: Both panels: *ship plume image* with indicated ship tracks. Panel a): *Ship track* – estimated, based on AIS data records. The ship track is shown for the time period starting from 2 hours before the satellite overpass until the moment of the satellite overpass. Panel b): *Wind-shifted ship 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 $0.4° \approx 37.4$ km, in vertical direction $0.4° \approx 44.2$ km.

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

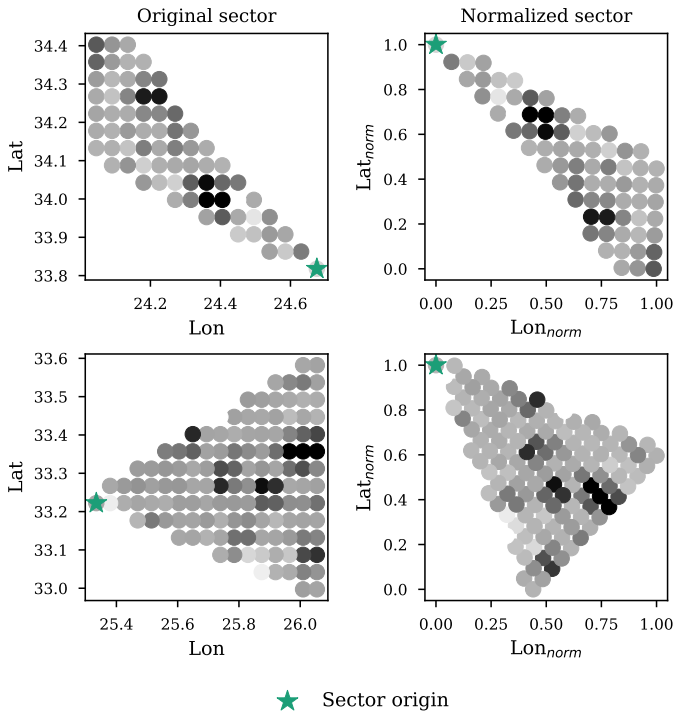

Figure 5.4: Sector normalization. We rotate the *ship sectors* so that all resulting sectors have the same orientation equal to 320° independently of the original direction of the ship's heading. We then rescale the image so that the range of both coordinates is between 0 and 1. The gray area in each 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:

$$lon\_norm = \frac{lon\_rot - min(lon\_rot)}{max(lon\_rot) - min(lon\_rot)}, \qquad (5.2)$$
$$lat\_norm = \frac{lat\_rot - min(lat\_rot)}{max(lat\_rot) - min(lat\_rot)}$$

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

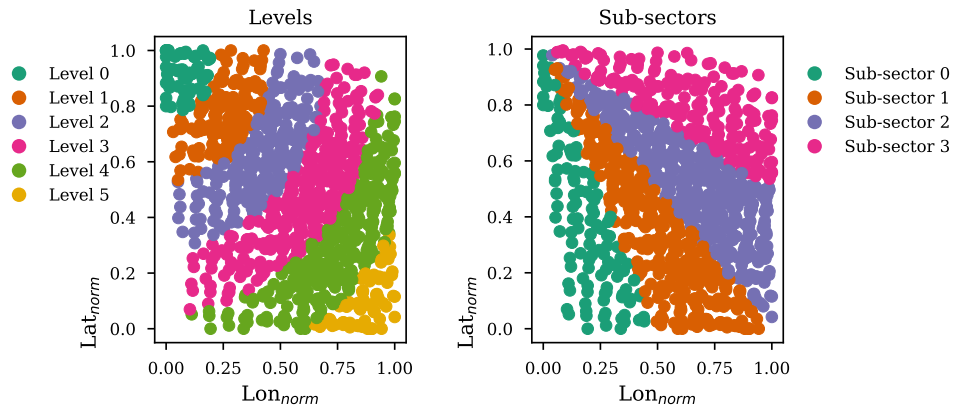

Figure 5.5: Levels and sub-sectors. We perform a feature construction by dividing the *normalized sector* into sub-regions: *levels* and *sub-sectors*. For the convenience of visualization, data points from one day of analysis were used for the preparation of the figure.

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

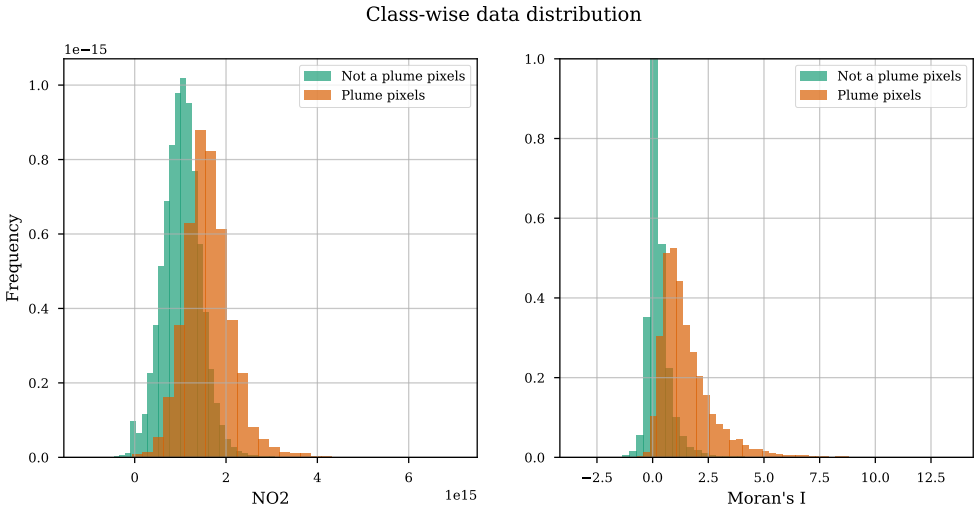

Figure 5.6: Class-wise distribution of the two main features of the dataset: NO_2 and Moran's I.

	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.

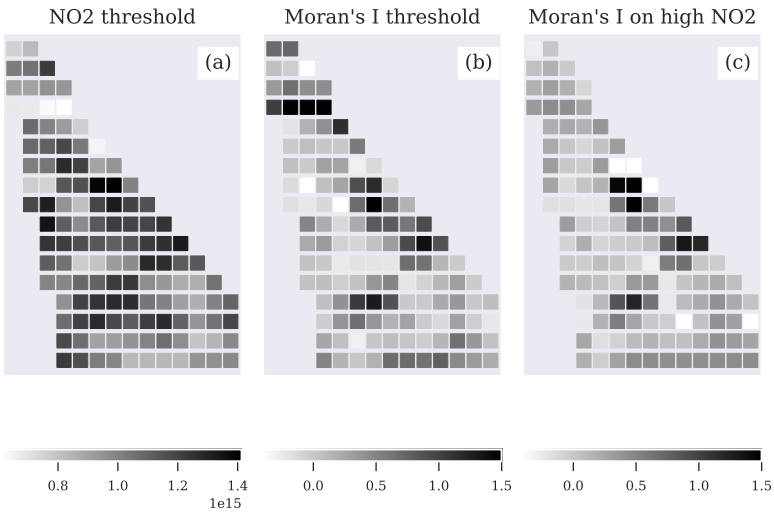

Figure 5.7: Input data example for univariate threshold-based benchmarks. a) Input data for a benchmark method *NO_2 threshold*. b) Input data for a benchmark method *Moran's I threshold*. At the top of the *ship sector* the reader can find an example when a cluster of low value NO_2 was mistakenly enhanced 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 benchmarking 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 enhancement. 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

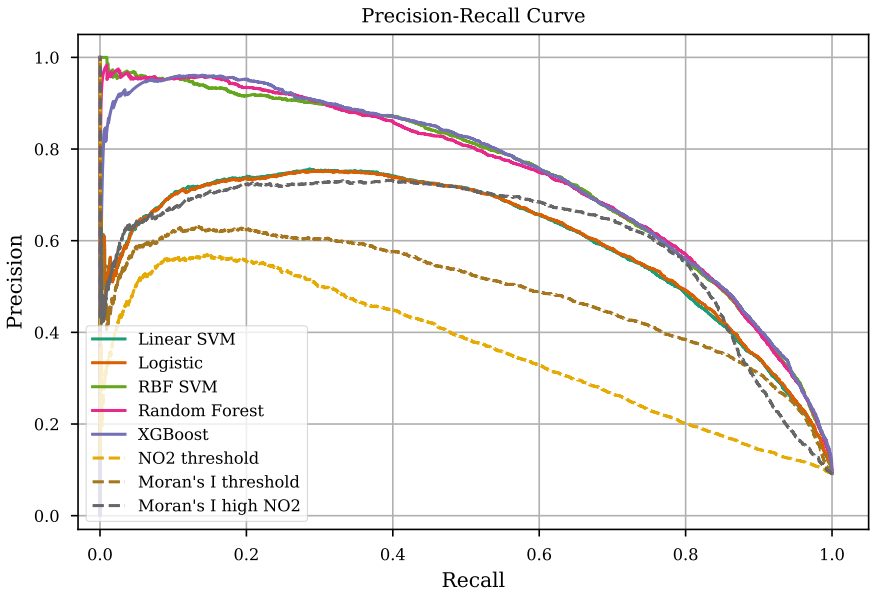

Figure 5.8: Precision-recall curve based on 5-fold cross-validation. Dashed lines indicate the results obtained from univariate threshold-based methods that, in this study, we considered as benchmarks.

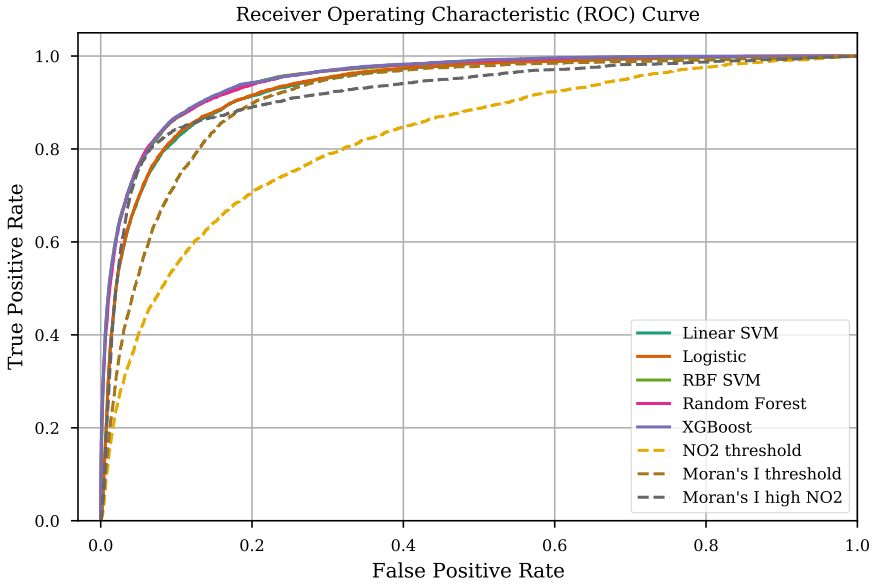

Figure 5.9: Receiver Operating Characteristics (ROC) curve based on 5-fold cross-validation. Dashed lines indicate the results obtained from univariate threshold-based methods that, in this study, we considered as benchmarks.

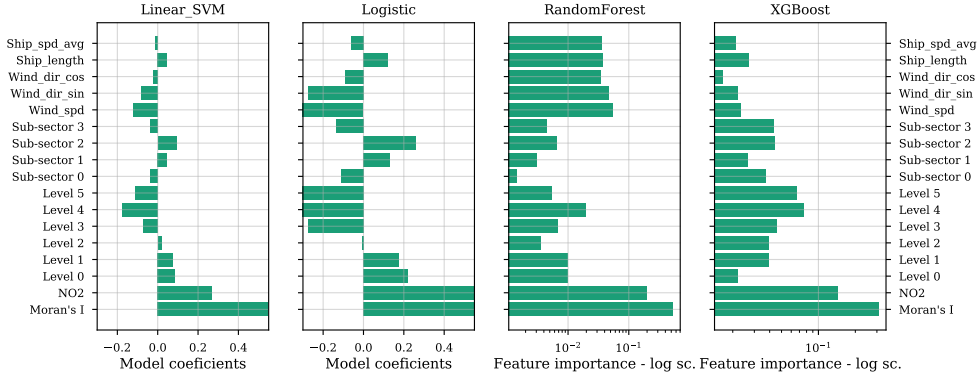

Figure 5.10: Coefficients of the features in the decision function of the linear models and the impurity-based feature importance values for tree-based models.

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.

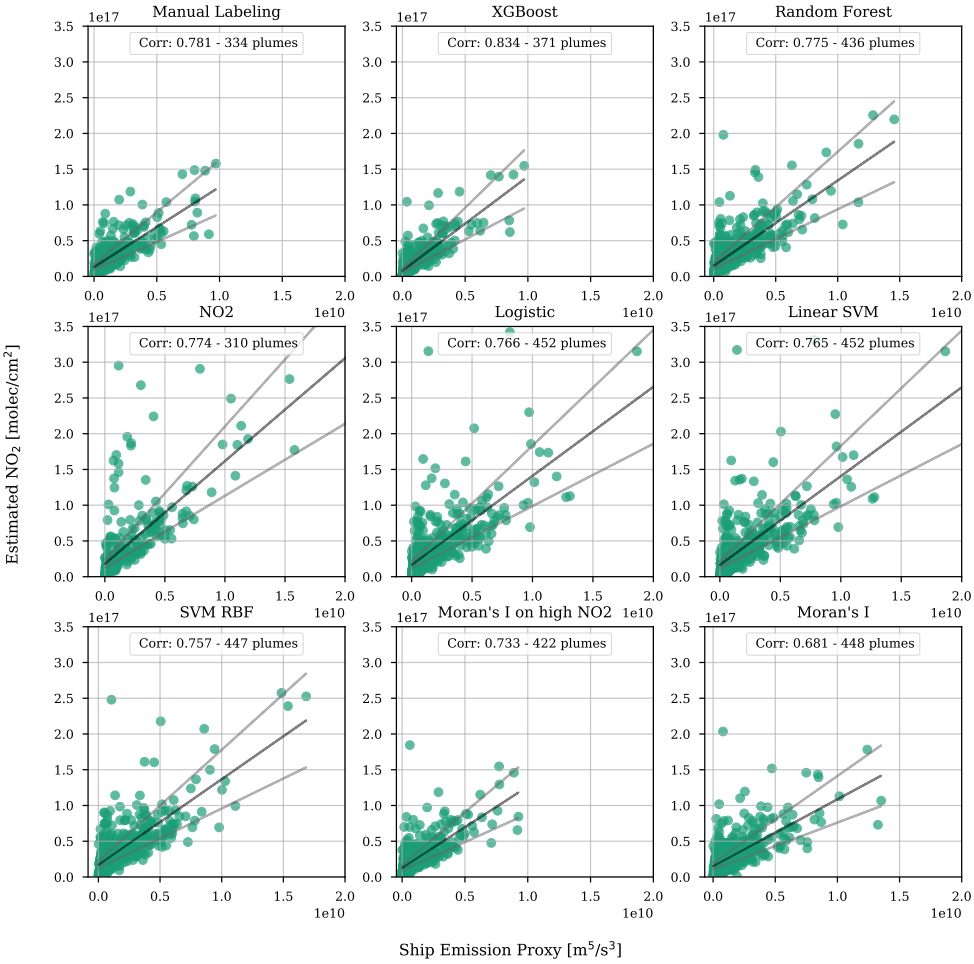

Figure 5.11: Pearson correlations between estimated (based on classification results) values of NO_2 emitted by each ship on a given day and a theoretical ship emission proxy. Black lines indicate a fitted linear trend. Grey lines show 30% deviations from the fitted linear trend.

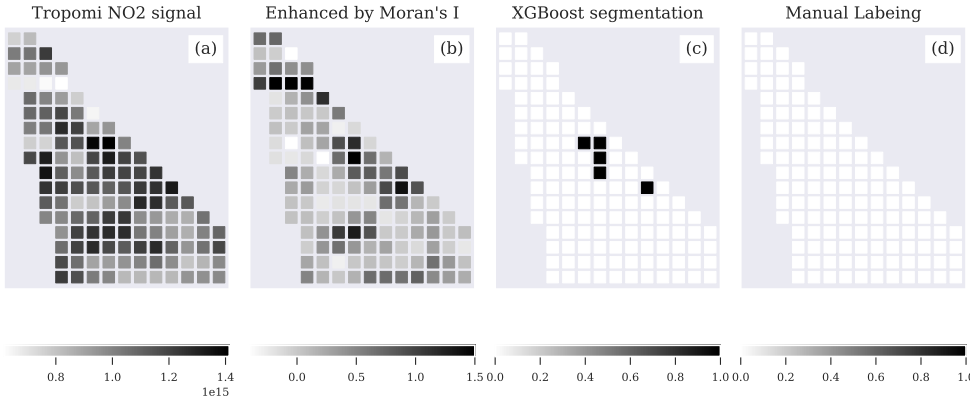

Figure 5.12: XGBoost classifier allows for the segmentation of plumes that were not recognized by the labeler. (a) TROPOMI NO_2 tropospheric vertical column density. Units: mol/m^2. The variable was a part of the input to machine-learning models. Ship plume is difficult to distinguish by the human eye. (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.

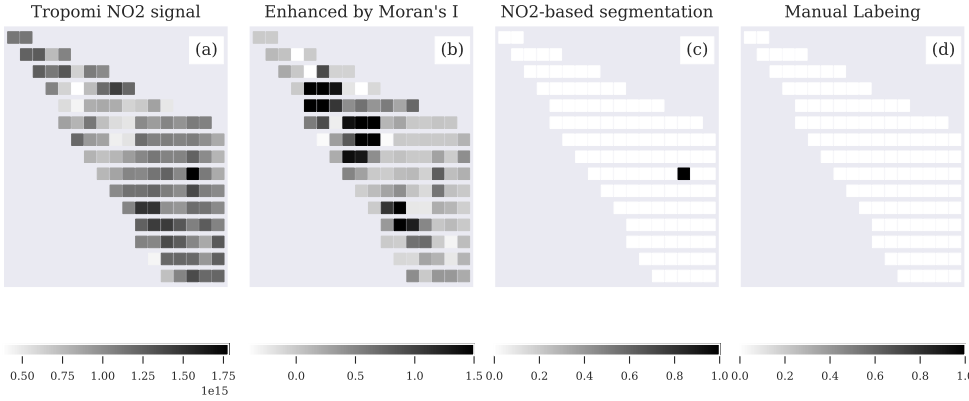

Figure 5.13: NO_2-based thresholding allows for distinguishing plumes cumulated within one pixel of the TROPOMI image. (a) TROPOMI NO_2 tropospheric vertical column density. Units: mol/m^2. (b) TROPOMI NO_2 image enhanced by Moran's I. At the top left of the *ship sector* can be found an example when 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

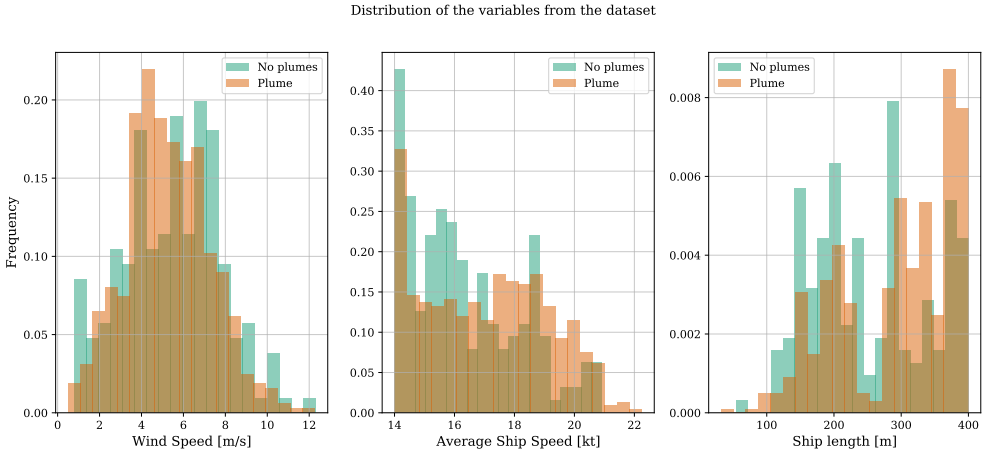

Figure 5.14: Distribution of the dataset features for the images, where there were no visible ship plumes distinguished, and for the images, where there was a visually distinguishable ship plume.

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.