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

Kurchaba, S.

Chapter 6

Automatic detection of anomalously emitting ships

Based on: 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.

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.

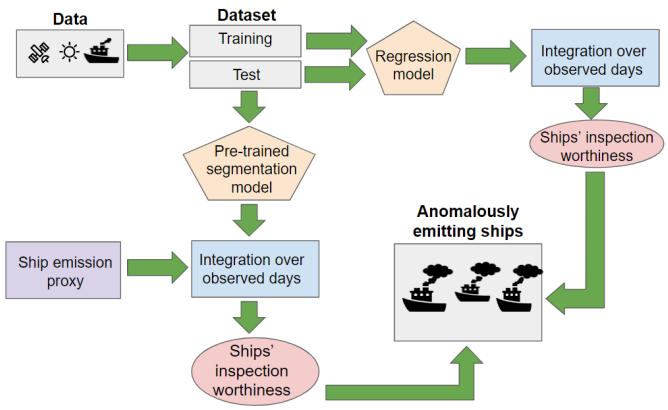

Figure 6.1: High-level diagram of the proposed methodology.

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:

$$NO_{2;d,s} = \sum_{i \in RoI_{d,s}} VCD_{NO_{2;i}} \tag{6.1}$$

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: $\hat{NO}_{2;d,s} \in \mathbb{R}$.

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

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

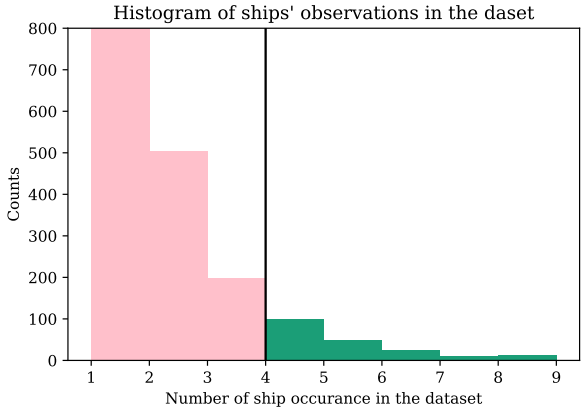

Figure 6.2: Histogram of occurrences of the same ship in the created dataset. The black line indicates the set level of min_obs_nb. Only ships that have been observed more than $min\_obs\_nb = 4$ days are taken into account for the detection of anomalously emitting ships.

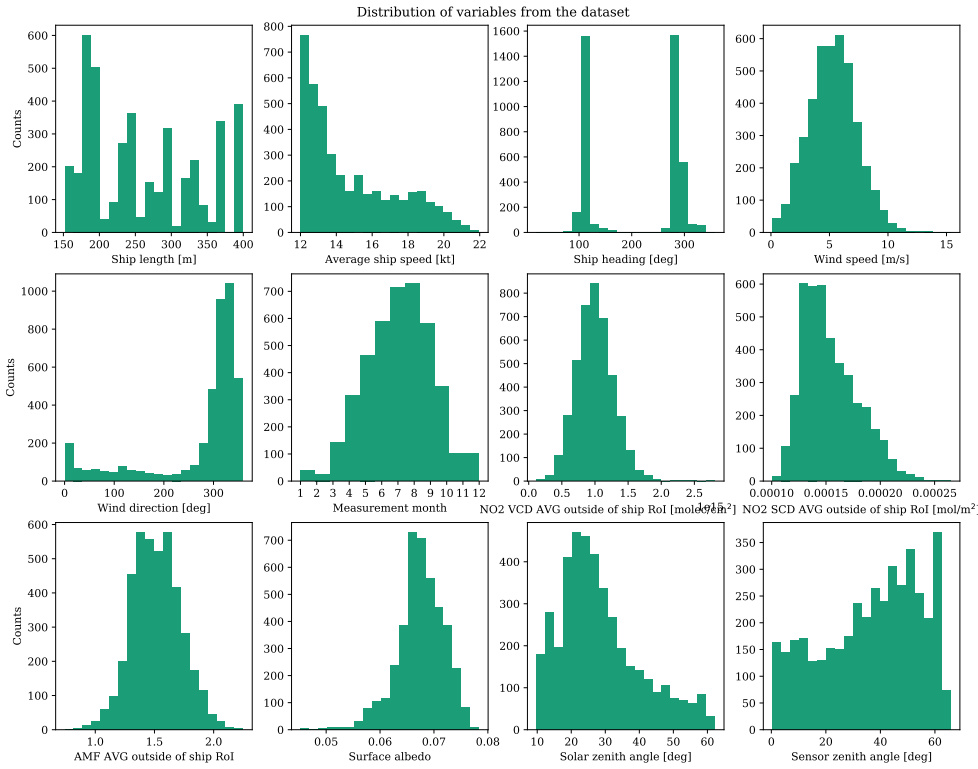

Figure 6.3: Distribution of variables of the dataset.

$|D_s|$ 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:

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

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

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

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

such that:

$$\begin{aligned} &Anomalous\_emitters = \{s_1, ..., s_n\} : \\ &norm\_IW_{s_i}^{regr} \cdot norm\_IW_{s_i}^{segm} < norm\_IW_{s_{i+1}}^{regr} \cdot norm\_IW_{s_{i+1}}^{segm} \end{aligned} \tag{6.8}$$

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

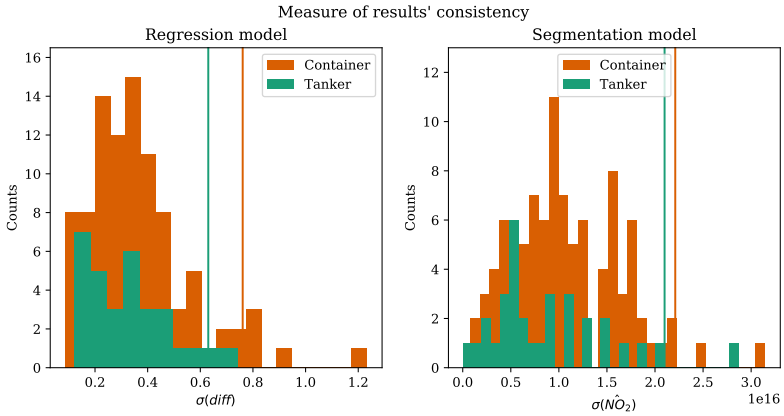

Figure 6.4: A measure of the consistency of results of the regression model $\sigma(diff_s)$ and segmentation model $\sigma(\hat{NO}_{2;d,s})$. The threshold $\mu(\hat{NO}_{2;d,s}) + 2\sigma(\hat{NO}_{2;d,s})$ is indicated with vertical lines.

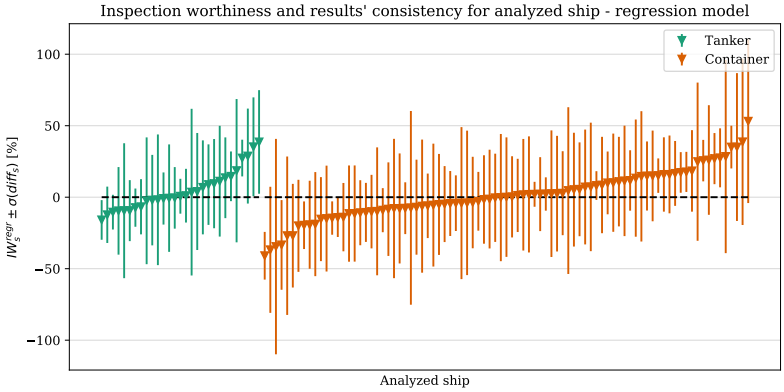

Figure 6.5: The triangle-shaped markers indicate the measure of ship inspection worthiness in accordance with the regression model IW_s^{regr}. The vertical lines indicate $\sigma(diff_s)$ - the measure of the consistency of results for a given ship.

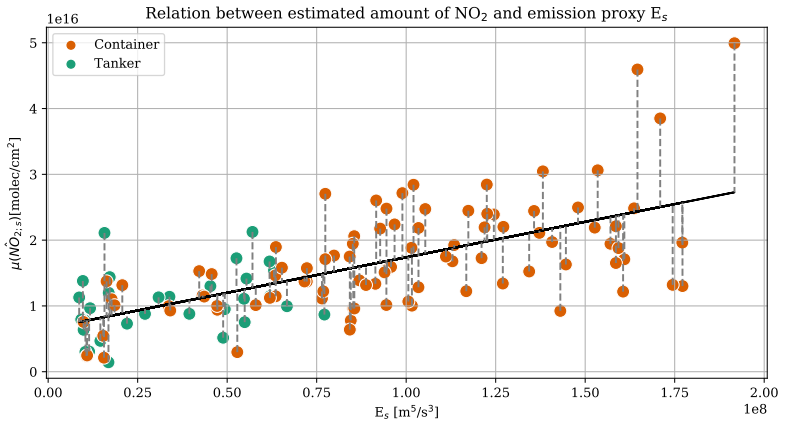

Figure 6.6: Relation between the estimated amount of NO_2 using the segmentation model and ship emission proxy with a fitted linear trend. Gray dashed lines indicate the measure of the ship inspection worthiness IW_s^{segm} according to the plume segmentation model.

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.

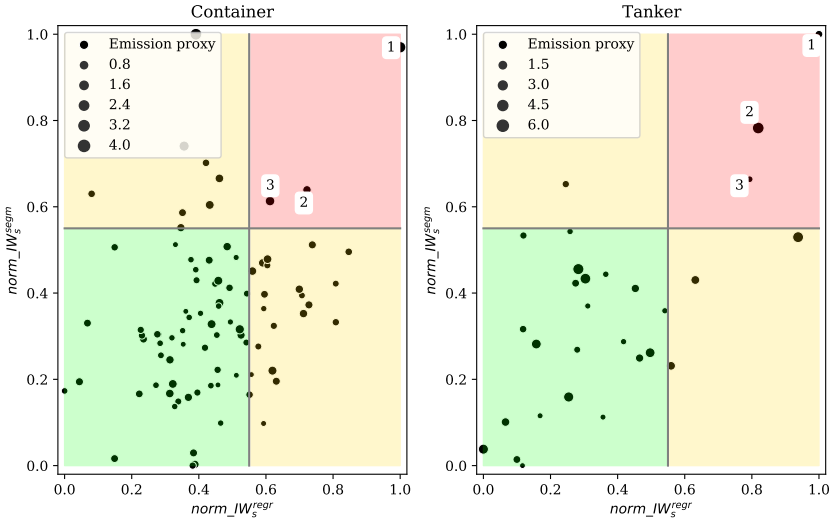

Figure 6.7: Combination of results of segmentation and regression models. Values of the inspection worthiness obtained from each model were normalized using a min-max scaler.

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

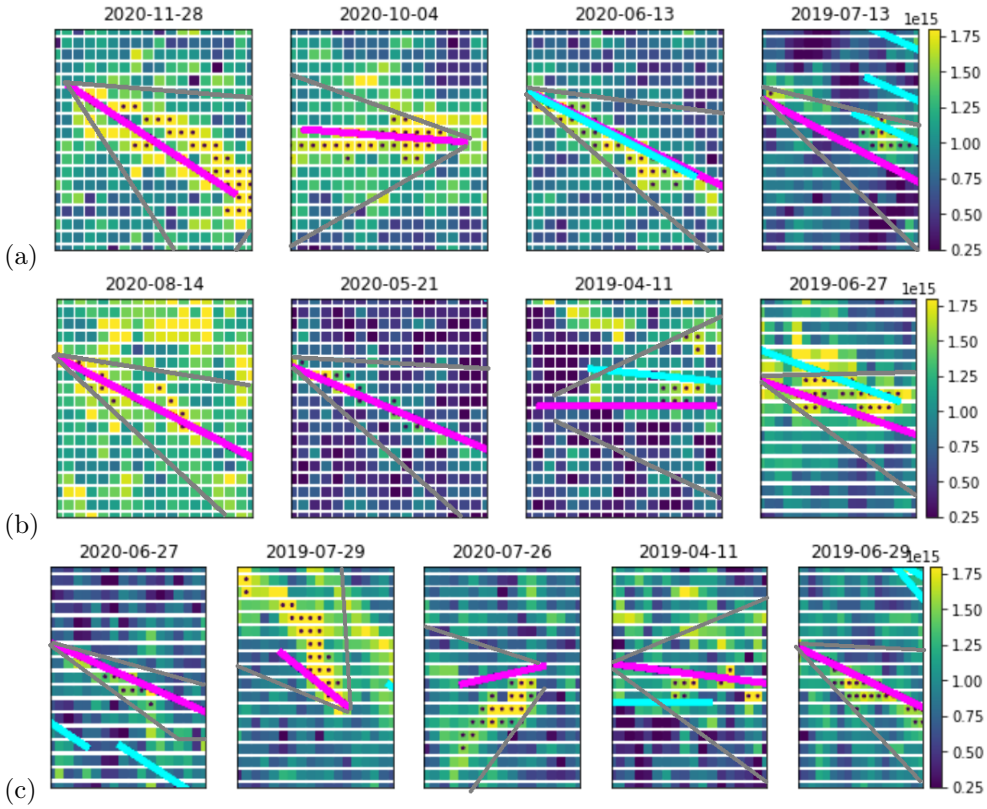

Figure 6.8: Ship type: Container. Lines represent shifted ship tracks. Magenta line – ship of interest. Cyan line – other ships in the area. Grey lines – borders of the *ship sector*. Dots indicate pixels classified by the segmentation model as a plume. a) Outlying ship 1. Ship length: 398 m. Average 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.

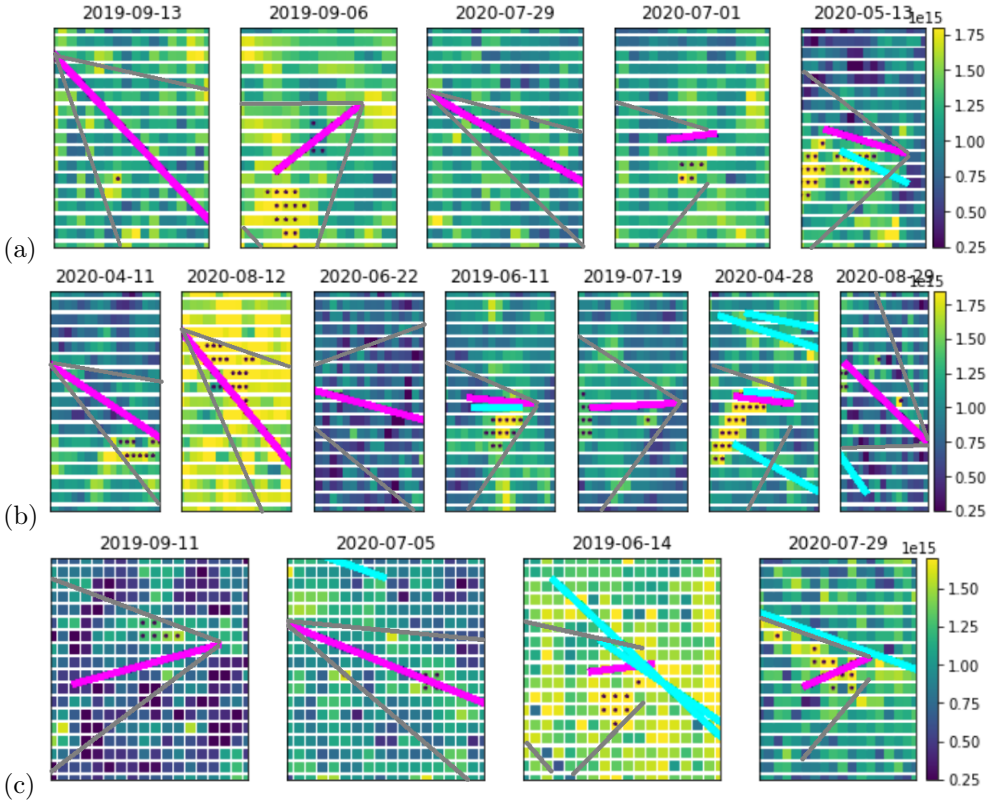

Figure 6.9: Ship type: Tanker. Lines represent shifted ship tracks. Magenta line – ship of interest. Cyan line – other ships in the area. Grey lines – borders of the *ship sector*. Dots indicate pixels classified by the segmentation model as a plume. a) Outlying ship 1. Ship length: 180 m. Average ship 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	$\sigma(diff)$	$\sigma(\hat{NO}_2)$
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 ($\sigma(\hat{NO}_2)$) 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 $\sigma(\hat{NO}_2)$ 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 $\sigma(\hat{NO}_2)$ 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 predetermine 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.