Anomalous NO2 emitting ship detection with TROPOMI satellite data and machine learning

Kurchaba, S.; Vliet, J. van; Verbeek, F.J.; Veenman, C.J.

Remote Sensing of Environment 297 (2023) 113761

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Anomalous NO_2 emitting ship detection with TROPOMI satellite data and machine learning

Solomiia Kurchaba [a,*], Jasper van Vliet [b], Fons J. Verbeek [a], Cor J. Veenman [a,c]

[a] Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Niels Bohrweg 1, Leiden, 2333 CA, The Netherlands
[b] Human Environment and Transport Inspectorate (ILT), Graadt van Roggenweg 500, Utrecht, 3531 AH, The Netherlands
[c] Data Science Department, TNO, Anna van Buerenplein 1, The Hague, 2595 DA, The Netherlands

ARTICLE INFO

Edited by Menghua Wang

Dataset link: https://s5phub.copernicus.eu/

Keywords:
TROPOMI
Machine learning
IMO 2020
Seagoing ships
NO2

ABSTRACT

Starting from 2021, more demanding NO_x emission restrictions were introduced for ships operating in the North and Baltic Sea waters. Since all methods currently used for ship compliance monitoring are financially and time demanding, it is important to prioritize the inspection of ships that have high chances of being non-compliant. The current state-of-the-art approach for a large-scale ship NO_2 estimation is a supervised machine learning-based segmentation of ship plumes on TROPOMI/S5P images. However, 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 study, we present a methodology towards the automated and scalable selection of potentially non-compliant ships using a combination of machine learning models on TROPOMI satellite 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.

1. Introduction

The industry of international shipping is one of the strongest sources of anthropogenic emission of nitrogen oxides (NO_x) - a substance harmful both to ecology and human health. The contribution of the shipping industry to the global emission of NO_x is estimated to vary between 15%–35% (Crippa et al., 2018; Johansson et al., 2017), causing approximately 60,000 premature deaths annually (Corbett et al., 2007). To mitigate the negative impact of this industry, the International Maritime Organization (IMO) stepwisely tightens the restrictions put on emission factors of marine engines (IMO, 1997). The latest step is an 80% reduction of NO_x emission for ships operating in the North and Baltic Sea (IMO, 2020).

The monitoring of the compliance of ships with the IMO regulations is being performed by manual onboard inspections. However, due to the high costs, a selection of ships that will undergo inspection is needed. Among the sources of information currently used for the selection of ships are in-situ emission measurement stations (Beecken et al., 2014; McLaren et al., 2012; Kattner et al., 2015) usually located at the entrance of the harbors, or airborne platform-based measurements such as planes, drones or helicopters (Van Roy and Scheldeman, 2016). The data collected with such methods give limited information on how much the selected ships emit outside of a port and are usually done near-shore. Additionally, the above-mentioned methods are spot checks that usually only happen once. This does not give a possibility of having a wider perspective on ship performance. As a result, the decisions regarding the worthiness of a ship inspection do not have sufficient justification.

Remote sensing is a well-established technique for the measurement of emission levels. In particular, there is an extensive list of studies using satellite-based instruments for the quantification of NO_2 emission levels produced by the shipping industry (Burrows et al., 1999; Beirle et al., 2004; Bovensmann et al., 1999; Richter et al., 2004; Levelt et al., 2006; Vinken et al., 2014). However, until recently the low spatial

* Corresponding author.
E-mail address: s.kurchaba@liacs.leidenuniv.nl (S. Kurchaba).

https://doi.org/10.1016/j.rse.2023.113761
Received 9 March 2023; Received in revised form 2 June 2023; Accepted 11 August 2023
Available online 21 August 2023

resolution of instruments did not allow to access plumes produced by individual ships.

The game changer is the TROPOspheric Monitoring Instrument onboard the Sentinel 5 Precursor (TROPOMI/S5P) satellite launched in 2018 (Veefkind et al., 2012). It is the first remote sensing instrument that is able to distinguish nitrogen dioxide (NO_2) plumes from individual ships (Georgoulias et al., 2020). This technical improvement allows considering remote sensing as a potential solution for ship compliance monitoring (SCIPPER, 2020). In particular, the data from the TROPOMI instrument could be used for the development of a data-driven inspection recommendation.

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 images to attribute the measured NO_2 to individual ships (Kurchaba et al., 2021, 2022). 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 (Fan et al., 2016; Georgoulias et al., 2020). For instance, in Georgoulias et al. (2020), the authors propose to estimate the expected amount of NO_x emission from ships as $Ship\_length^2 \times Ship\_speed^3$. However, such proxies often 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 study, we propose a robust method for automated selection of anomalously NO_2 emitting seagoing ships. 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 sensor 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.

We train the regression model with an automatically delineated Region of Interest (RoI) based on ship, wind speed, and direction. We apply Automated Machine Learning (AutoML) 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 previously developed (Kurchaba et al., 2022) method for ship plume segmentation. 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 the candidates for further inspection.

The rest of this paper is organized as follows: In Section 2, we describe the data sources used in this study. In Section 3, we introduce the developed methodology, which is followed by the results presented in Section 4. In Sections 5 and 6, the reader can find the discussion and final conclusions respectively.

2. Data

To prepare the dataset, we combine several sources of data. We use the TROPOMI file.[1] to retrieve an NO_2 tropospheric vertical column density (VCD_{trop}) variable; wind data that is used to define the RoI of a ship, and as a feature of the segmentation and regression models; albedo data, as well as two VCD_{trop} priors (slant column density (SCD) and air mass factor (AMS)) that are used as features of the regression model. We use Automatical Identification System (AIS) data for the position of ships at the moment of the satellite overpass. Finally, official ship registries are used to retrieve information about the dimensions of the studied ships. In the following section, we provide a detailed description of all used data sources.

2.1. TROPOMI data

TROPOMI/S5P (Veefkind et al., 2012) – TROPOspheric Monitoring Instrument onboard the Sentinel 5 Precursor (S5P) satellite is a UV–Vis-NIR-SWIR (UV, visible, near-infrared, short-wave infrared) spectrometer operating from May 2018. It is a sun-synchronous satellite that achieves global coverage with approximately 14 orbits in 24 h. The local equatorial overpass time of the satellite is 13:30. The TROPOMI instrument measures spectra of multiple trace gasses including NO_2. The NO_2 gas is an outcome of photochemical reactions of NO_x emitted by ships and, therefore, is suitable for ships' compliance monitoring (Kurchaba et al., 2022). In this study, the variable of interest is NO_2 tropospheric vertical column density – VCD_{trop} (Eskes et al., 2022).

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 (Eskes et al., 2022). This results in the fact that the plumes located in the regions of historical shipping lanes will be enhanced by the retrieval algorithm (Douros et al., 2023). 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.

This study we based on the analysis of the same region[2] in the eastern Mediterranean Sea as in Kurchaba et al. (2022). An outline of the studied area can be found in Fig. 1. The study period is 20 months, starting from 1 April 2019 until 31 December 2020. The spatial resolution of the TROPOMI spectrometer equals 3.5×5.5 km^2 at nadir, with the real size of the pixel varies depending on the distance between the satellite and the captured part of the earth's surface. To obtain the images of regular size, we perform regridding[3] 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 (Kurchaba et al., 2022). The following quality filters were applied on the satellite data: only measurements flagged with $qa\_value > 0.5$ (Sneep, 2021) are taken into consideration. In addition, since the TROPOMI measurements 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.

2.2. Meteorological data

For the study, the wind information is taken from wind speed data from the European Center for Medium-range Weather Forecast (ECMWF) at 10 m height, available with 0.25° resolution at a 6-hourly time step. The surface albedo data is the OMI minimum Lambertian equivalent reflectance (LER) at a resolution of 0.5°. Both ECMWF wind data and OMI surface albedo data are available as support products in the TROPOMI/S5P data file.

[1] Open access under https://s5phub.copernicus.eu/. TROPOMI/S5P data version: 2.3.1.

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

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

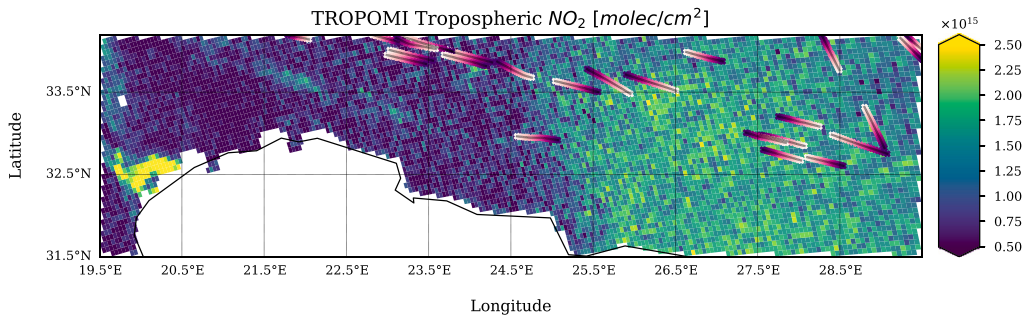

Fig. 1. The NO_2 tropospheric column. Visualized day: June 14th, 2019. Study area: part of the Mediterranean Sea, bound by the Northern coasts of Libya and Egypt in the South and South of Crete in the North. Magenta lines indicate tracks of ships based on AIS data. The elevated background concentrations (green) in the east correspond to outflow from a variety of land-based sources. To visualize the TROPOMI data, a native local pixel size is plotted before regridding.
Source: (Kurchaba et al., 2022).

2.3. Ship-related data

Another data source used in this study is relayed through Automatic Identification System (AIS) transponders.[4] The data include the position, speed, heading, and unique identifier (MMSI) of each ship carrying an active transponder. Due to the fact that at the moment there is no open-access AIS data available, for the scope of this study, the AIS data as well as information about the dimensions of the ships (such as length, type, and gross tonnage) were provided by the Netherlands Human Environment and Transport Inspectorate (ILT). This is the Dutch national designated authority for shipping inspections, has access to commercial databases for the AIS data set used in this study, and is participating in this research.

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 and faster than 12 kt. 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 h 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.

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 (Kurchaba et al., 2022) 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 measurements. Fig. 2 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.

3.1. Regression model

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

3.1.1. 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:

$$\mathrm{NO}_{2;d,s} = \sum_{i \in RoI_{d,s}} VCD_{\mathrm{NO}_{2;i}} \tag{1}$$

where $\mathrm{VCD}_{\mathrm{NO}_2}$ is the value of the retrieved TROPOMI pixel within the RoI of the analyzed ship (see Section 3.1.2 for more details of RoI definition). 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: $\hat{\mathrm{NO}}_{2;d,s} \in \mathbb{R}$.

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

The list of features X can be found in Table 1. In Appendix A, we provide histograms of the features, as well as other dataset details. As a next step, we calculate $diff_{d,s}[\%]$ – a percentage difference between the predicted and observed amount of NO_2. Finally, assuming $|D_s|$ is the number of days when the ship s was observed, min_obs_nb is the minimal 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 assure that our ship profile is representative to make the decision about being anomalously emitting and taking into consideration data availability (see Fig. 3), 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}.

[4] Since 2002 all commercial sea-going vessels are obliged to carry on board an AIS transponder (Mou et al., 2010).

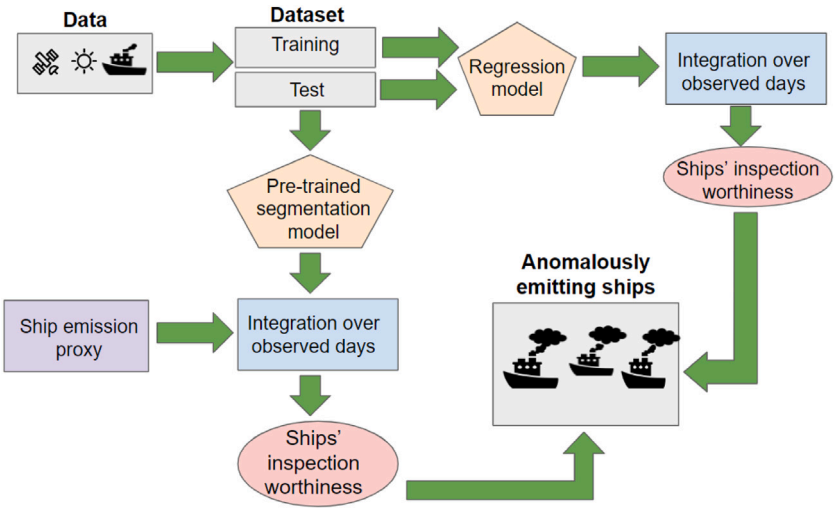

Fig. 2. High-level diagram of the proposed methodology.

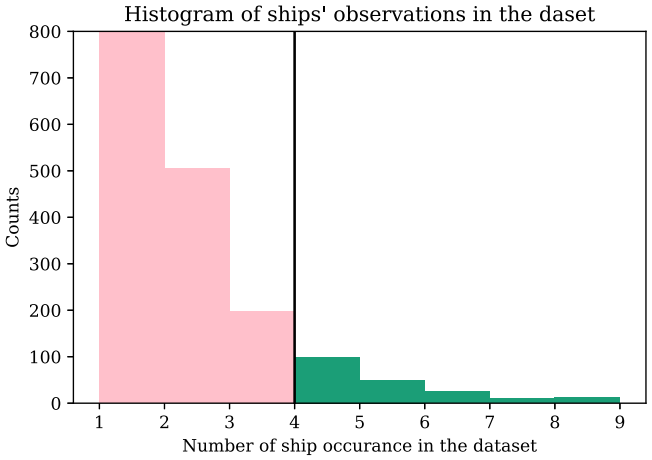

Fig. 3. Histogram of occurrences of the same ship in the created dataset. The black line indicates the set level of *min_obs_nb*. Only ships that have been observed more than *min_obs_nb* = 4 days are taken into account for the detection of anomalously emitting ships.

Table 1
List of features used for the regression model. The area outside the RoI is restricted to the ship neighborhood defined as the *ship plume image* in accordance to Kurchaba et al. (2022).

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 RoI Average NO_2 SCD outside RoI AMF outside RoI Sensor zenith angle

The value $\sigma(diff_{d,s})$ is a measure of the consistency of the obtained results. Since the satellite 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 $\sigma(diff_{d,s})$.

3.1.2. Defining RoI

The RoI of a ship defines the region within which the emitted NO_2 during the last two hours is expected based on ship speed, wind speed, direction, and uncertainties. We use the method of RoI assignation presented in Kurchaba et al. (2021). First, we estimate the trajectory of the ship – a *ship track* – using AIS ship data, starting from two hours before, until the moment of the satellite overpass (c.f. Fig. 4a). The observation duration of two hours was selected considering an average lifetime of NO_x (de Foy et al., 2015). Secondly, 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 Fig. 4b.

Both wind speed and wind direction are assumed to be constant for the whole time during which we study the plume. Such an assumption 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. Fig. 4c. 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 maximal distance from the position of the ship at the moment of the satellite overpass to the position of the ship 2 h before the satellite overpass 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 Fig. 4d an example of a resulting RoI that we call a *ship sector* is presented.

3.1.3. Model optimization

We use a nested scheme of cross-validation (see Fig. 5). Within the outer 5-fold loop of cross-validation we create 5 "hold out" non-overlapping test sets and 5 training sets. The test sets are used for:

1. Performance evaluation of the regression model.

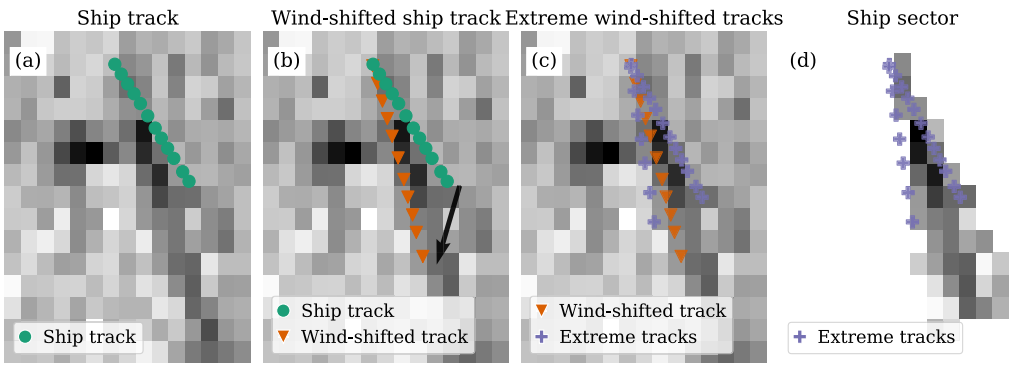

Fig. 4. Ship sector definition pipeline. Background – the TROPOMI NO_2 signal for the area around the analyzed ship. Two ship plumes can be distinguished at the central part of the image, and one less intense ship plume is located in the lower right of the image. Only one is of interest here. (a) *Ship track* – estimated, based on AIS data records. The ship track is shown for the time period starting from 2 h before until the moment of the satellite overpass. (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. For all presented images, the size of the pixel is equal to 4.2 × 5 km^2.
Source: (Kurchaba et al., 2022).

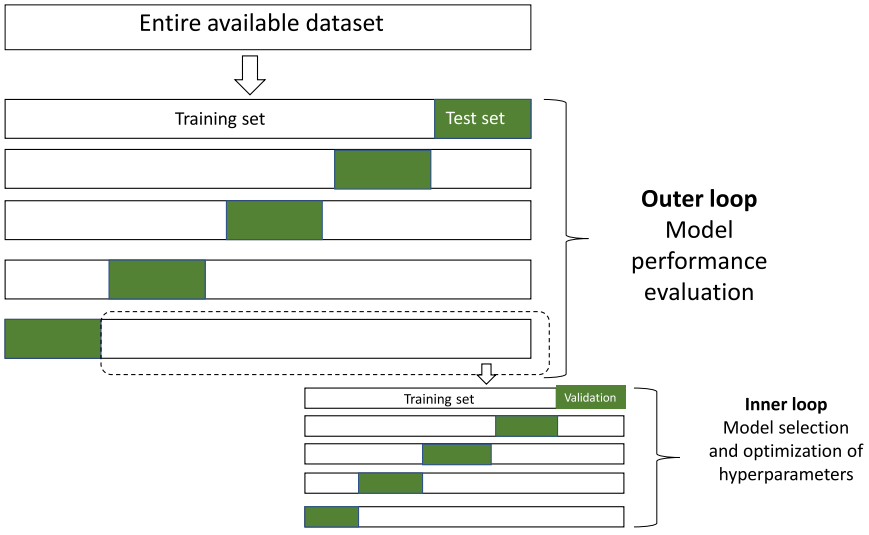

Fig. 5. Applied scheme of cross-validation. In the outer loop, we generated five test sets that were used for the regression model performance evaluation, as well as for the detection of anomalously emitting ships. In the inner loop, the generated training and validation sets were used by AutoML algorithms to optimize machine learning pipelines for regression models.

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 using automated machine learning (AutoML) (Hutter et al., 2019). With AutoML, we aim to solve a so-called CASH problem, which stands for Combined Algorithm Selection and Hyperparameter optimization (Kotthoff et al., 2019). Given the absence of available 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 without preliminary model selection experiments. In this study, we address the CASH problem using TPOT (Tree-based Pipeline Optimization Tool) (Olson et al., 2016) – a Python package for automatic selection of machine learning pipelines based on genetic programming (GP) (Koza, 1994).

The results obtained using the TPOT AutoML library are benchmarked towards the results obtained using the eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016) regression model with the default hyperparameters settings. The XGBoost model is considered to be a good choice when it comes to tabular data (Grinsztajn et al., 2022), as well as showed the best performance on the same type of data in our previous study (Kurchaba et al., 2022).

3.2. Detection of anomalously emitting ships

In order to assure 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 introduce the ship plume segmentation model (Kurchaba et al., 2022) that 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.

3.2.1. 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 Kurchaba et al. (2022). This method uses manually annotated data to train a supervised model for the segmentation of a

ship plume within the ship RoI as defined in Section 3.1.2. 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 amount of NO_2 can be expressed as:

$$\hat{NO}_{2;d,s} = \sum_{i \in RoI_{d,s}} \hat{y}_i \cdot NO_{2,i}, \tag{3}$$

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[5] $E_{d,s}$. 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:

$$\mu(\hat{NO}_{2;d,s}) = \alpha \cdot \mu(E_{d,s}) + \beta + \epsilon_s, \tag{4}$$

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 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.

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 RoI, and ship plume segmentation model developed in previous study (Kurchaba et al., 2022). 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:

$$norm\_IW_s^{regr} = \frac{IW_s^{regr} - min(IW_s^{regr})}{max(IW_s^{regr}) - min(IW_s^{regr})} \tag{5}$$

$$norm\_IW_s^{segm} = \frac{IW_s^{segm} - min(IW_s^{segm})}{max(IW_s^{segm}) - min(IW_s^{segm})} \tag{6}$$

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

$$norm\_IW_s^{regr} > t \wedge norm\_IW_s^{segm} > t \iff s \in Anomalous\_emitters, \tag{7}$$

such that:

$$Anomalous\_emitters = \{s_1, \ldots, s_n\} :$$

$$norm\_IW_{s_i}^{regr} \cdot norm\_IW_{s_i}^{segm} < norm\_IW_{s_{i+1}}^{regr} \cdot norm\_IW_{s_{i+1}}^{segm} \tag{8}$$

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 $t = 0.55$.

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.

[5] For details on model training and emission proxy definition see Appendix B.

Table 2
Regression model results. Hyperparameters applied for AutoML optimization: Maximal evaluation time: 10 min; Population size: 50; Number of generations: 50; Early stopping criteria: 10.

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

Table 3
A model and a feature pre-processor selected by TPOT as optimal at a given iteration of cross-validation.

Feature processor	Model
MaxAbs Scaler	Gradient Boosting (Friedman, 2002)
MaxAbs Scaler	Gradient Boosting
Polynomial Features (2_{nd} deg.)	XGBoost (Chen and Guestrin, 2016)
Standard Scaler	Gradient Boosting
Standard Scaler	XGBoost

4.1. Regression model optimization

In Table 2, we present the results of the regression model optimization. The application of the TPOT pipeline optimization algorithm allowed us to improve the results of both used quality metrics over our benchmark – default XGBoost. In Table 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.

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. Fig. 6 presents the consistency measures for regression and segmentation models along with the applied cut-off thresholds.

In Fig. 7, 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. Fig. 8 presents the 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. Fig. 9 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

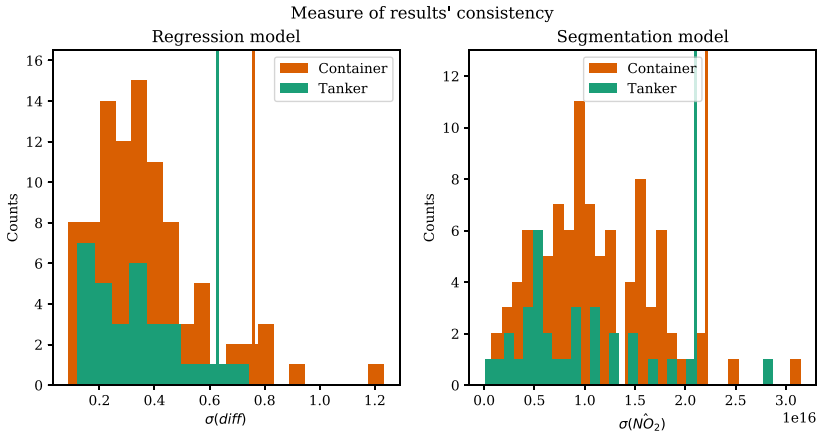

Fig. 6. A measure of the consistency of results of the regression model $\sigma(diff_s)$ and segmentation model $\sigma(\hat{NO}_{2;d,s})$. The threshold $\mu(\hat{NO}_{2;d,s}) + 2\sigma(\hat{NO}_{2;d,s})$ is indicated with vertical lines.

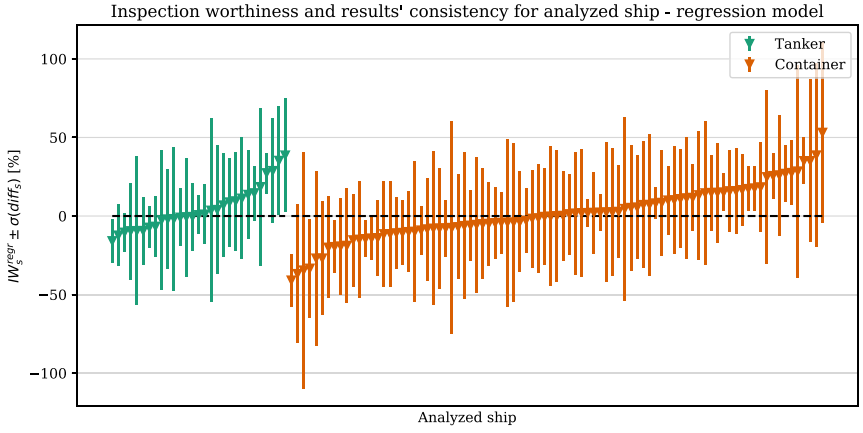

Fig. 7. The triangle-shaped markers indicate the measure of ship inspection worthiness in accordance with the regression model IW_s^{regr}. The vertical lines indicate $\sigma(diff_s)$ - the measure of the consistency of results for a given ship.

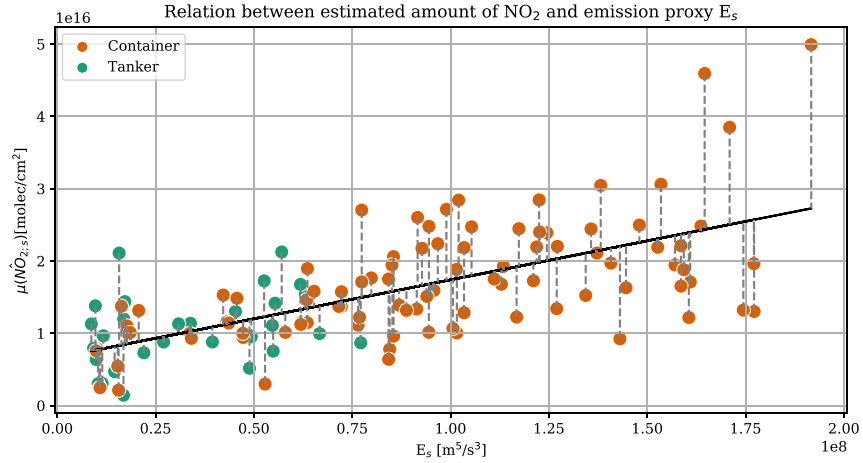

Fig. 8. Relation between the estimated amount of NO_2 using the segmentation model and ship emission proxy with a fitted linear trend. Gray dashed lines indicate the measure ship inspection worthiness IW_s^{segm} according to the plume segmentation model.

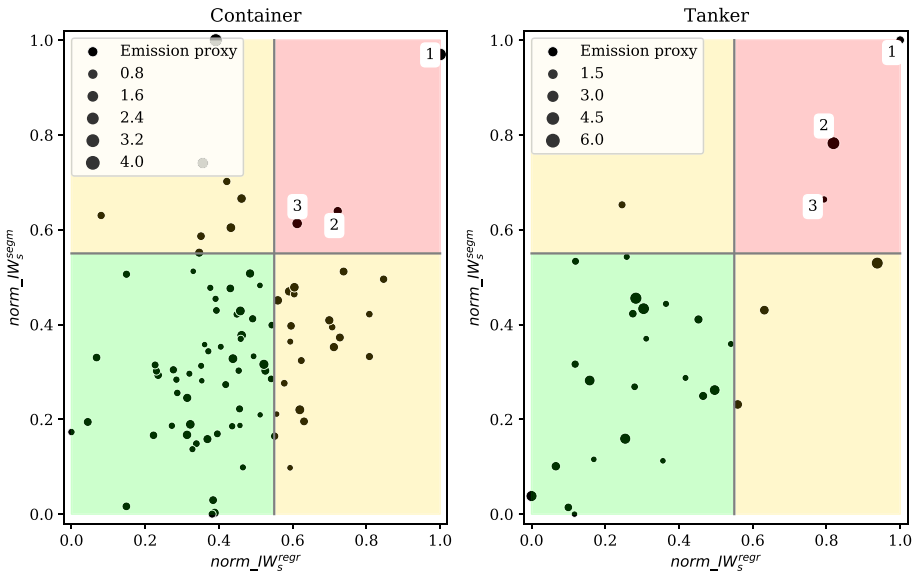

Fig. 9. Combination of results of segmentation and regression models. Values of the inspection worthiness obtained from each model were normalized using a min–max scaler.

Table 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 Fig. 9 for containers and tankers respectively.

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 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	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

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.

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 measurements related to those ships. Fig. 10a–c and Fig. 11a–c provide the TROPOMI images for the red-zone containers and tankers respectively. On the images from the corresponding dates of TROPOMI observations, we indicate the trajectory of the ship of interest, the other ships in the ship image, 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 images 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 4) for ships selected as anomalous emitters with the data distribution for the whole set of studied ships (Fig. 6), 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. Fig. 11a), we can see that for two (2019-03-13, 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 Fig. 11b, 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 images, 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.

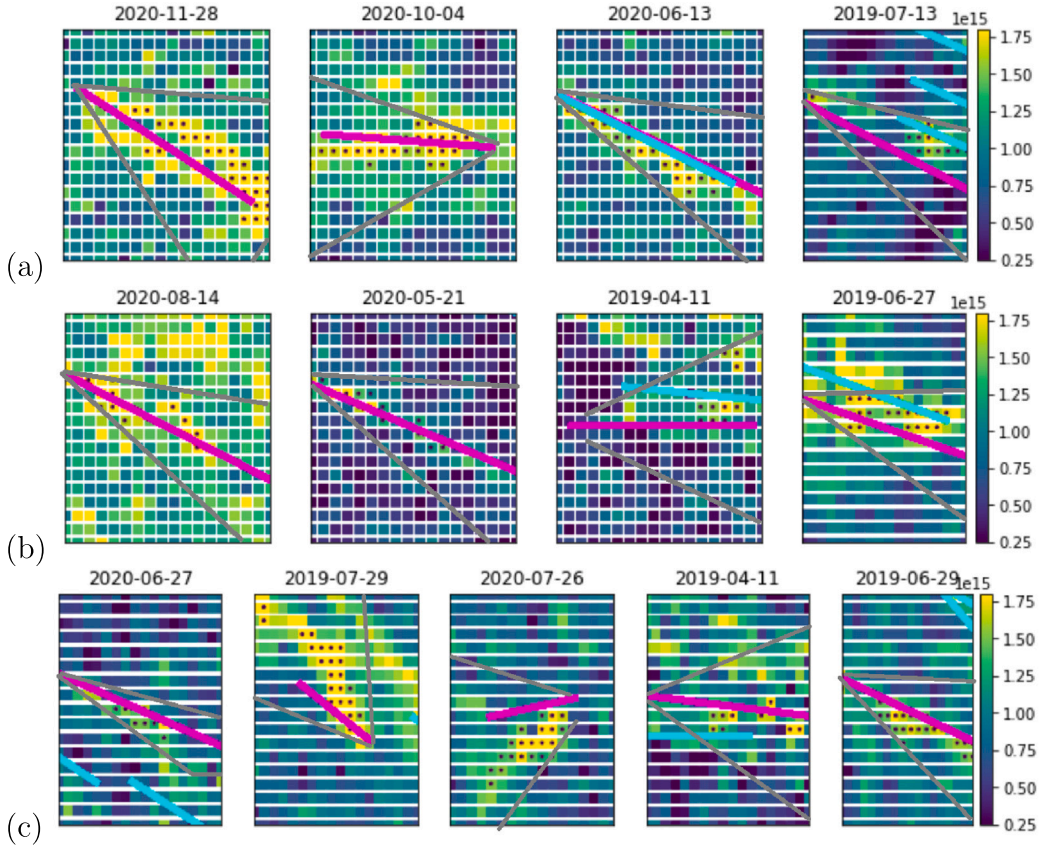

Fig. 10. Ship type: Container. Lines represent shifted ship tracks. Magenta line – ship of interest. Cyan line – other ships in the area. Gray lines – borders of the RoI of the analyzed ship. Dots indicate pixels classified by the segmentation model as a plume. (a) Outlying ship 1. Ship length: 398 m. 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 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.

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

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 ships RoI with the RoI of other ships[6]) were not a part of any model[7] 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 5) and performed a univariate one-way ANOVA test (Table 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.

5. Discussion

In this study, we presented a method for detecting anomalously NO_2 emitting ships by applying a combination of machine-learning-based methods on TROPOMI satellite data. The provided methodology is an important step towards 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 a satellite is the only available measurement instrument that can access ship emissions in the open sea.

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

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

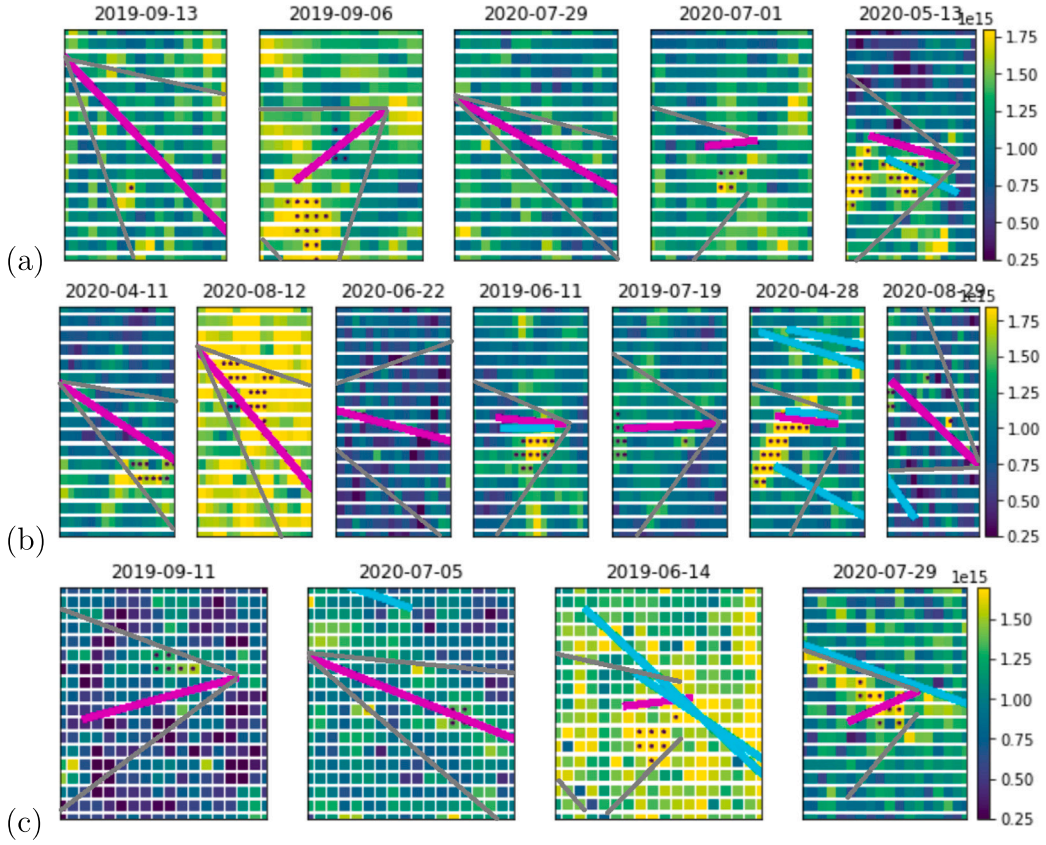

Fig. 11. Ship type: Tanker. Lines represent shifted ship tracks. Magenta line – ship of interest. Cyan line – other ships in the area. Gray lines – borders of the RoI of the analyzed ship. Dots indicate pixels classified by the segmentation model as a plume. (a) Outlying ship 1. Ship length: 180 m. Average 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Another advantage of satellite observations in contrast to all the other methods currently used for ship emission monitoring, is that satellite measurements enable us to observe the emissions over time regularly and remotely. The presented approach exploits this property of a satellite by making multi-days 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 Figs. 10 and 11, 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 measurements with an even higher spatial resolution (for instance, TANGO instrument Landgraf et al., 2020) 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 is 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 satellite observation is 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. Conclusions

In this study, we applied a combination of machine learning-based methods on TROPOMI satellite 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 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 towards the development of a scalable recommendation system for ship inspectors that is based on satellite observations.

CRediT authorship contribution statement

Solomiia Kurchaba: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation,

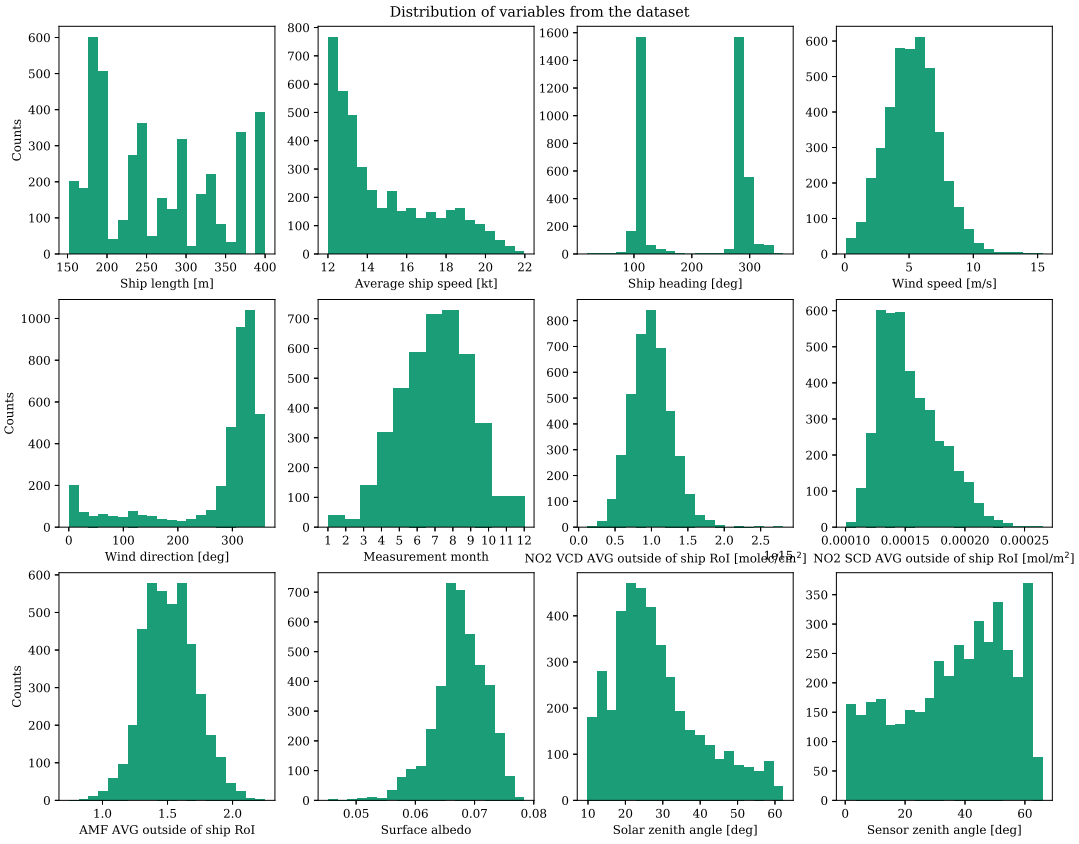

Fig. A.12. Distribution of variables of the dataset used in this study.

Writing – original draft, Visualization. **Jasper van Vliet:** Conceptualization, Methodology, Resources, Data curation, Writing – review & editing, Supervision, Funding acquisition. **Fons J. Verbeek:** Resources, Writing – review & editing, Supervision, Project administration. **Cor J. Veenman:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Data availability

The TROPOMI/S5P data is freely available via https://s5phub.copernicus.eu/. Starting from product version upgrade from 1.2.2 to 1.3.0 that took place on March 27, 2019, the ECMWF operational model analyzes 10 meters wind data for coinciding time is available as a support product in the TROPOMI/S5P data file. For the scope of this study, the AIS data, as well as information about the dimensions of the ships were provided to us by the ILT, which is the Dutch national designated authority for shipping inspections, is participating in this research, and has access to commercial databases of AIS data and official ship registries.

Acknowledgments

This work is funded by the Netherlands Human Environment and Transport Inspectorate, the Dutch Ministry of Infrastructure and Water Management, and the SCIPPER project, which receives funding from the European Union's Horizon 2020 research and innovation program under grant agreement Nr.814893. All authors approved the version of the manuscript to be published.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Regression model

The dataset used for the regression model is composed of 4153 rows (aggregated ship plume images). Fig. A.12 presents the histograms of the distributions of the variables from the regression model dataset.

Appendix B. Segmentation model dataset

To train a segmentation model we used a labeled dataset presented in Kurchaba et al. (2022). The dataset is composed of 68 days of NO_2 TROPOMI measurements taken between 1 April 2019 and 31 December 2019. The dataset covers the same area in the Mediterranean Sea as the regression model dataset (see Section 2.1).

In Kurchaba et al. (2022) it was shown that the highest performance quality of the ship plume segmentation task was achieved with XGBoost classifier. Therefore, in this study, for the task of ship plume segmentation, we use XGBoost model and optimize it using the methodology from the original article. The hyperparameters The obtained cross-validation-averaged average precision score is equal to 0.753. For the extensive reports of the model performance evaluation, we direct the reader to Kurchaba et al. (2022).

Following the methodology from Kurchaba et al. (2022), to validate the results obtained with the segmentation model, we use a theoretical ship emission proxy $E_{d,s}$ (Georgoulias et al., 2020) defined as:

$$E_{d,s} = L_s^2 \cdot u_{d,s}^3 \tag{B.1}$$

where L_s is the length of the ship s in m, and $u_{d,s}$ is its average speed on a day d in m/s, derivation details see Georgoulias et al. (2020).

References

Beecken, J., Mellqvist, J., Salo, K., Ekholm, J., Jalkanen, J.-P., 2014. Airborne emission measurements of SO 2, NO x and particles from individual ships using a sniffer

technique. Atmos. Meas. Tech. 7 (7), 1957–1968. http://dx.doi.org/10.5194/amt-7-1957-2014.

Beirle, S., Platt, U., Von Glasow, R., Wenig, M., Wagner, T., 2004. Estimate of nitrogen oxide emissions from shipping by satellite remote sensing. Geophys. Res. Lett. 31 (18), http://dx.doi.org/10.1029/2004GL020312.

Bovensmann, H., Burrows, J., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V., Chance, K., Goede, A., 1999. SCIAMACHY: Mission objectives and measurement modes. J. Atmos. Sci. 56 (2), 127–150. http://dx.doi.org/10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2.

Burrows, J.P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißenmayer, A., Richter, A., DeBeek, R., Hoogen, R., Bramstedt, K., Eichmann, K.-U., et al., 1999. The global ozone monitoring experiment (GOME): Mission concept and first scientific results. J. Atmos. Sci. 56 (2), 151–175. http://dx.doi.org/10.1175/1520-0469(1999)056<0151:TGOMEG>2.0.CO;2.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. pp. 785–794. http://dx.doi.org/10.1145/2939672.2939785, https://doi.org/10.1145/2939672.2939785.

Corbett, J.J., Winebrake, J.J., Green, E.H., Kasibhatla, P., Eyring, V., Lauer, A., 2007. Mortality from ship emissions: a global assessment. Environ. Sci. Technol. 41 (24), 8512–8518. http://dx.doi.org/10.1021/es071686z, https://doi.org/10.1021/es071686z.

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Van Aardenne, J.A., Monni, S., Doering, U., Olivier, J.G., Pagliari, V., et al., 2018. Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4. 3.2. Earth Syst. Sci. Data 10 (4), 1987–2013. http://dx.doi.org/10.5194/essd-2018-31, https://doi.org/10.5194/essd-2018-31.

de Foy, B., Lu, Z., Streets, D.G., Lamsal, L.N., Duncan, B.N., 2015. Estimates of power plant NOx emissions and lifetimes from OMI NO2 satellite retrievals. Atmos. Environ. (ISSN: 1352-2310) 116, 1–11. http://dx.doi.org/10.1016/j.atmosenv.2015.05.056, URL: https://www.sciencedirect.com/science/article/pii/S1352231015301291.

Douros, J., Eskes, H., van Geffen, J., Boersma, K.F., Compernolle, S., Pinardi, G., Blechschmidt, A.-M., Peuch, V.-H., Colette, A., Veefkind, P., 2023. Comparing sentinel-5P TROPOMI NO 2 column observations with the CAMS regional air quality ensemble. Geosci. Model Dev. 16 (2), 509–534. http://dx.doi.org/10.5194/gmd-16-509-2023.

Eskes, H., van Geffen, J., Boersma, F., Eichmann, K.-U., Apituley, A., Pedergnana, M., Sneep, M., Veefkind, J.P., Loyola, D., 2022. Sentinel-5 Precursor/TROPOMI Level 2 Product User Manual Nitrogendioxide. Technical Report S5P-KNMI-L2-0021-MA.

Fan, Q., Zhang, Y., Ma, W., Ma, H., Feng, J., Yu, Q., Yang, X., Ng, S.K., Fu, Q., Chen, L., 2016. Spatial and seasonal dynamics of ship emissions over the Yangtze River Delta and east China sea and their potential environmental influence. Environ. Sci. Technol. 50 (3), 1322–1329. http://dx.doi.org/10.1021/acs.est.5b03965, https://doi.org/10.1021/acs.est.5b03965.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378. http://dx.doi.org/10.1016/S0167-9473(01)00065-2.

Georgoulias, A.K., Boersma, K.F., van Vliet, J., Zhang, X., Zanis, P., de Laat, J., et al., 2020. Detection of NO2 pollution plumes from individual ships with the TROPOMI/S5P satellite sensor. Environ. Res. Lett. 15 (12), 124037. http://dx.doi.org/10.1088/1748-9326/abc445.

Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why do tree-based models still outperform deep learning on tabular data?. arXiv:arXiv:2207.08815.

Hutter, F., Kotthoff, L., Vanschoren, J., 2019. Automated Machine Learning: Methods, Systems, Challenges. Springer Nature, http://dx.doi.org/10.1007/978-3-030-05318-5, URL: https://doi.org/10.1007/978-3-030-05318-5.

IMO, 1997. Amendments to the annex of the protocol of 1978 relating to the international convention for the prevention of pollution from ships. URL: https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MEPCDocuments/MEPC.75(40).pdf.

IMO, 2020. MARPOL ANNEX VI - regulation 13. URL: https://www.imo.org/en/OurWork/Environment/Pages/Nitrogen-oxides-(NOx)-%E2%80%93-Regulation-13.aspx.

Johansson, L., Jalkanen, J.-P., Kukkonen, J., 2017. Global assessment of shipping emissions in 2015 on a high spatial and temporal resolution. Atmos. Environ. 167, 403–415. http://dx.doi.org/10.1016/j.atmosenv.2017.08.042, https://doi.org/10.1016/j.atmosenv.2017.08.042.

Kattner, L., Mathieu-Üffing, B., Burrows, J.P., Richter, A., Schmolke, S., Seyler, A., Wittrock, F., 2015. Monitoring compliance with sulfur content regulations of shipping fuel by in situ measurements of ship emissions. Atmos. Chem. Phys. 15 (17), 10087–10092. http://dx.doi.org/10.5194/acp-15-10087-2015, URL: https://acp.copernicus.org/articles/15/10087/2015/.

Kotthoff, L., Thornton, C., Hoos, H.H., Hutter, F., Leyton-Brown, K., 2019. Auto-WEKA: Automatic model selection and hyperparameter optimization in WEKA. In: Automated Machine Learning. Springer, Cham, pp. 81–95. http://dx.doi.org/10.1007/978-3-030-05318-5_4, URL: https://doi.org/10.1007/978-3-030-05318-5_4.

Koza, J.R., 1994. Genetic programming as a means for programming computers by natural selection. Statist. Comput. 4 (2), 87–112. http://dx.doi.org/10.1007/BF00175355, https://doi.org/10.1007/BF00175355.

Kurchaba, S., van Vliet, J., Meulman, J.J., Verbeek, F.J., Veenman, C.J., 2021. Improving evaluation of NO2 emission from ships using spatial association on TROPOMI satellite data. In: 29th International Conference on Advances in Geographic Information Systems. pp. 454–457. http://dx.doi.org/10.1145/3474717.3484213, https://doi.org/10.1145/3474717.3484213.

Kurchaba, S., van Vliet, J., Verbeek, F.J., Meulman, J.J., Veenman, C.J., 2022. Supervised segmentation of NO2 plumes from individual ships using TROPOMI satellite data. Remote Sens. (ISSN: 2072-4292) 14 (22), http://dx.doi.org/10.3390/rs14225809, URL: https://www.mdpi.com/2072-4292/14/22/5809.

Landgraf, J., Rusli, S., Cooney, R., Veefkind, P., Vemmix, T., de Groot, Z., Bell, A., Day, J., Leemhuis, A., Sierk, B., 2020. The TANGO mission: A satellite tandem to measure major sources of anthropogenic greenhouse gas emissions. In: EGU General Assembly Conference Abstracts. p. 19643.

Levelt, P.F., Hilsenrath, E., Leppelmeier, G.W., van den Oord, G.H., Bhartia, P.K., Tamminen, J., de Haan, J.F., Veefkind, J.P., 2006. Science objectives of the ozone monitoring instrument. IEEE Trans. Geosci. Remote Sens. 44 (5), 1199–1208. http://dx.doi.org/10.1109/TGRS.2006.872336.

McLaren, R., Wojtal, P., Halla, J.D., Mihele, C., Brook, J.R., 2012. A survey of NO2: SO2 emission ratios measured in marine vessel plumes in the strait of Georgia. Atmos. Environ. 46, 655–658. http://dx.doi.org/10.1016/j.atmosenv.2011.10.044, https://doi.org/10.1016/j.atmosenv.2011.10.044.

Mou, J.M., Van der Tak, C., Ligteringen, H., 2010. Study on collision avoidance in busy waterways by using AIS data. Ocean Eng. 37 (5–6), 483–490. http://dx.doi.org/10.1016/j.oceaneng.2010.01.012, https://doi.org/10.1016/j.oceaneng.2010.01.012.

Olson, R.S., Bartley, N., Urbanowicz, R.J., Moore, J.H., 2016. Evaluation of a tree-based pipeline optimization tool for automating data science. In: Genetic and Evolutionary Computation Conference. pp. 485–492. http://dx.doi.org/10.1145/2908812.2908918, https://doi.org/10.1145/2908812.2908918.

Richter, A., Eyring, V., Burrows, J.P., Bovensmann, H., Lauer, A., Sierk, B., Crutzen, P.J., 2004. Satellite measurements of NO2 from international shipping emissions. Geophys. Res. Lett. 31 (23), http://dx.doi.org/10.1029/2004GL020822.

SCIPPER, 2020. Shipping contributions to inland pollution push for the enforcement of regulations. URL: https://www.scipper-project.eu/.

Sneep, M., 2021. Sentinel 5 precursor/TROPOMI KNMI and SRON level 2 Input Output Data Definition. Technical Report S5P-KNMI-L2-0009-SD.

Van Roy, W., Scheldeman, K., 2016. Results MARPOL annex VI monitoring report: Belgian sniffer campaign 2016.

Veefkind, J., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., Claas, J., Eskes, H., De Haan, J., Kleipool, Q., et al., 2012. TROPOMI on the ESA sentinel-5 precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. Remote Sens. Environ. 120, 70–83. http://dx.doi.org/10.1016/j.rse.2011.09.027, https://doi.org/10.1016/j.rse.2011.09.027.

Vinken, G., Boersma, K., van Donkelaar, A., Zhang, L., 2014. Constraints on ship NO x emissions in Europe using GEOS-Chem and OMI satellite NO 2 observations. Atmos. Chem. Phys. 14 (3), 1353–1369. http://dx.doi.org/10.5194/acp-14-1353-2014.