Reviewer # 3

In this manuscript, Yuan and co-authors describe a machine learning approach to identifying NO2 from ship emissions in daily images of TROPOMI tropospheric NO2 data. The model is trained by human classification of test data over known shipping lanes and then applied to a full year of TROPOMI data. The resulting map of frequencies of pixels identified as being affected by ship emissions is consistent with TROPOMI NO2 observations and earlier work on NO2 from ships.

The topic of the manuscript (a machine learning tool for the identification of NO2 from ships) is relevant for atmospheric applications and fits into the scope of AMT. The manuscript is clearly written but would benefit from another round of English proof reading. The level of details is low, and I recommend to provide more details on the method, the training and the performance of the model to make the manuscript more useful for the readers. It would also be good if the authors could give an example of a possible application of their method, as this is not clear to me.

Major comments:

1. *In the title and throughout the manuscript, the authors write about NO2 plumes. At first, I understood this to be plumes from individual ships. However, after reading the full manuscript, I think the algorithm looks for large elongated regions of enhanced NO2 as they are present over certain shipping routes. These NO2 enhancements are the accumulated result of many individual plumes, and in my opinion, they should not be called "plumes".*

   We understand the point that the reviewer raised here. We also changed the text to make it clear that we are not dealing exclusively with individual ship plumes. However, we think these are plumes and just need to note that many of them may not be result of a single ship emission.

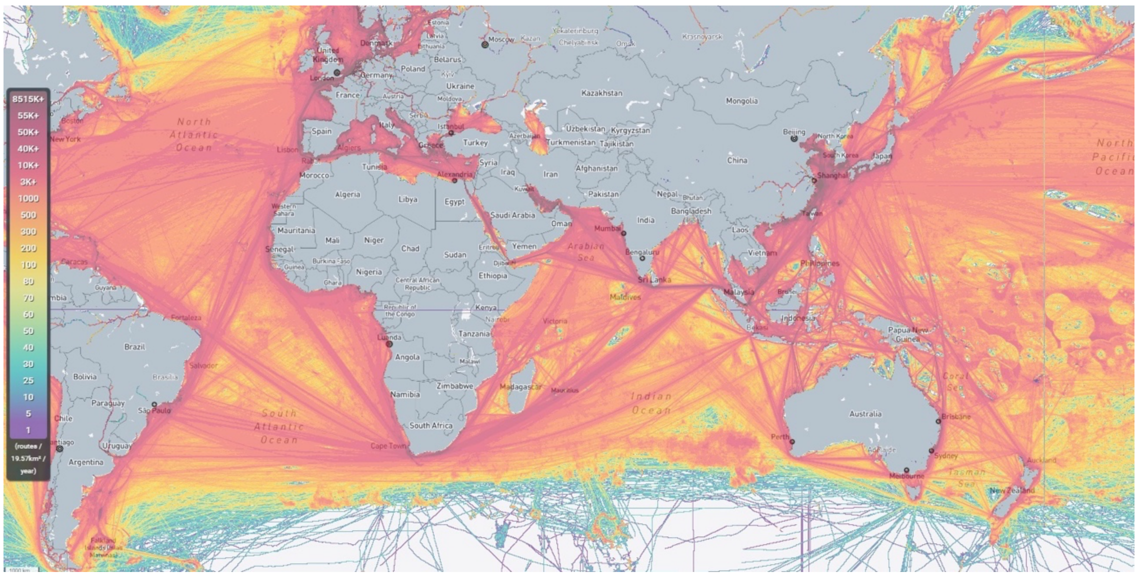
2. *As in all satellite NO2 maps, only some of the shipping lanes can be identified in this study. The authors explain this by clouds, which certainly are a factor, but probably not the main one. In my opinion, NO2 signals from ships can only be detected if a large enough number of ships operates on a narrow track at not too high wind speeds, as dilution can reduce the signal below the detection limit of the satellite measurements. These aspects need to be discussed, and the statement "we show that cloudiness in these shipping routes is the culprit …" needs to be changed.*

   Thank you for raising this valid point. We have changed the abstract and added more discussion on this point in the discussion section to reflect this point.

3. *The authors discuss a shipping route from the coast of Africa to Madagascar. I'm sceptical about this finding as this is also a well known area of pollution export from*

South African power stations. My guess is, that these have elongated shapes and are therefore misclassified as ship emissions. Please discuss.

We made the speculation based on the known shipping activity and routes around this region (attached figure based on AIS data), but after checking the data, the reviewer is right that there are large sources around the South African coast and continental NO_2 is frequently transported over the ocean, which may indeed affect our results. We added discussions in the text on this possibility accordingly and added text in the discussion section.



4. *The authors claim, that their method is useful for emission inventory and emission compliance studies. These claims need to be substantiated or removed. As it is presented in the manuscript, the counting of pixels classified as containing ship emissions is not quantitative relative to the NOx emissions of ships in general and even less for individual ships.*

   Another reviewer also mentioned this point. We agree that the model in its current form alone does not provide direct information for compliance and inventory. The general method and possibly coupling with inverse modeling could prove useful in this area in the future.

5. *Very little information is given on the model and the training. For example, from which regions are the training data sets? From line 99/100 it appears, as if only data over the Indian Ocean is used? What is the proportion of "empty" training data? What is the performance of the classification?*

   We appreciate the reviewer's comments and now have added additional discussions. Training data are from the Indian Ocean region. If by 'empty training data' the reviewer means samples that do not contain plume pixels, it

occupies about 10% of the data. The model handles these true negatives quite well. This can be seen in the density map where no plume pixels are reported outside of major shipping lanes even though we apply our model on these data. We added some discussion on this.

Minor comments:

Line 32: Satellites do not retrieve NO2 concentrations but NO2 columns

Fixed.

Line 34/35: Language

Fixed.

Line 47/48: These references are more than 20 years old and no longer representative

Added a new reference.

Line 76/77: See major point about "plumes". I don't think that your algorithm is actually identifying the plumes from individual ships.

We made changes to remove the potential confusion.

Line 99: Which shipping route map?

Clarified.

Figure 1: Was this scene part of the training data set? If so, then please replace by a scene from the test data set.

It is part of the test data.

Line 126: Please provide more details about the input matrix. Why did you choose 400 x 400? Does this mean, that you are skipping the outermost 25 pixels on each side of an orbit?

We oversample the data twice to include the 50 pixels. We first take in the 400x400 and then take a stride of 50 to sample again.

Line 128: lever => level

Fixed.

Figure 2: It would be nice to update this figure to show NO2, not clouds

Fixed.

Line 145: Does that imply two flips and 3 rotations? Are these transformations already included in your counting of training data?

Yes. No, the number of training samples are before data augmentation.

Figure 3: Were these scenes part of the training data set? If so, then please replace by scenes from the test data set.

No.

Figure 4: TROPOMI does not retrieve NO2 emissions as stated in the figure and in the text in line 218. It also does not retrieve NO2 concentrations as stated in the caption. It's NO2 columns.

Fixed.

Figure 4: Some aggregation has been applied in these figures which appear to have low spatial resolution (otherwise, no more than 365 counts per year are possible in a given location, while frequencies here are up to 10000. Please explain what is shown here.

Yes. We do aggregate the data into coarser resolution than the original data.

Line 222: Why low clouds? Aren't high clouds even more of a problem? See also my general comments on clouds.

Yes, high clouds would also affect the detection. However, most the shipping routes are over areas where low clouds dominate. The message would be the same if we use all clouds.

Line 282: I think this sentence is a bit mixed up. It's not the uncertainties which add subpixel variabilities, but sub-pixel variability which leads to uncertainties.

Modified it to clarify.

Line 293: See general comment on clouds

Fixed.

Data availability: It would be good to make the model and the derived masks available on a repository

We upload the masks to a public data repository. https://doi.org/10.7910/DVN/L8J05A