Review of 'Detecting ship-produced NO_2 plumes and shipping routes in TROPOMI data with a deep learning model' by Tianle Yuan et al. <u>(Reviewer #2)</u>

The manuscript by Yuan et al. presents a method that uses machine learning to identify NO2 plumes from ships in individual TROPOMI orbits. The main result of the application of the machine learning method is a climatological map of 'NO2 pixel frequencies' obtained from one year of TROPOMI NO2 data. The map itself is convincing, and in line with what has previously been observed in satellite NO2 data: the main shipping routes between Europe and Asia can be well-identified, but ship plumes between Asia and North America, and between Europe and North America are much less detectable. The authors claim that this is mostly due to persistent cloud cover over the Atlantic and Pacific Ocean. In MODIS cloud data they have previously detected ship tracks (as bright clouds) right over these oceans. In my opinion, this complementarity of MODIS-derived ship tracks and TROPOMI-derived ship NO2 plumes is a nice idea that has the potential to push ship emissions research from space further in the long run. This manuscript does not do that, but is mostly an AI-effort to find NO2 plumes from ships, and derives a climatology from the result.

For a non-expert like me, I found section 3.1 where the human labelled training of NO2 ship plume detection is explained with useful imagery (Figure 3) quite useful. However, a bit more details about the capabilities and limitations of the method are needed and strengthen the paper, as detailed below. Should this manuscript be published I encourage the authors to make available their results of positive plume identifications in open access, digital format, as electronic asset to this paper.

Major comments

1. The authors show some convincing examples of massive 50-100 km long NO2 plumes. These are easily identified by eye. But the authors should also give some examples of where their method is just at the level of positive detection. How elongated are such plumes in terms of pixels – a one pixel enhancement should never qualify as a plume. What are the NO2 column values, and to what extent are such 'just positive' plumes above the background NO2 values?

As the first reviewer notices, there is inherently ambiguity here. In practice, we do not set rigid thresholds when we label the data. The strongest signal is the contrast to the background. Then a plume has to shape like a plume, a continuous and more or less semi-linear feature.

2. The main metric shown is 'NO2 pixel frequency', but it is not entirely clear what this means. Is it the accumulated number of (gridded?) pixels identified by the machine learning from all qa_value-filtered over one year? Would it not be more appropriate to express this frequency as a normalized frequency, to account for differences in the available cloud-free scenes?

Thanks for this comment. We now add more clarification on the definition. The reviewer is right that the plotted frequency is accumulated number of plume pixels in a grid.

3. It does not become clear how the machine learning approach handles plumes originating from land that flow out over sea. If unaccounted for, this leads to frequent misclassifications of plumes over coastal seas.

This is handled in the manual label process. A few factors help to discriminate such scenarios. Plumes overflow from the land usually have different shapes and their column amount is higher too. We also check for location of the plumes to see if they are located along shipping routes.

4. The claim that the methodology of the authors could be used for 'verification of compliance and effect of emission control policies' is not substantiated in the manuscript. The study deals with recognizing NO2 plumes in noisy data, which could be a first step in a system that estimates ship NOx emissions. Beyond attributing NO2 plumes to individual or multiple nearby ships, much more is needed in terms of inverse modelling, considering atmospheric chemistry and transport which influence the relationship between NOx emissions and NO2 columns.

Thanks for this comment. The reviewer is right that this is not reality yet. We modify the text to reflect what the reviewer points out here.

5. That tropospheric NO2 is not well detectable from space for pixels (partially) covered by clouds is obviously not a new finding and has been well-documented since the early 2000s (Martin et al. (2002); Palmer et al. (2001); Boersma et al. (2004)-papers). In lines 222-223 the impression should be avoided that this is a new result from this manuscript. Furthermore, there are new insights that in situations of sun-glint, screened out by the authors now given their qa_value threshold, the NO2 retrieval is performing better (Riess et al., 2022). Including these scenes rather than excluding them could improve the performance of the Machine Learning model.

We agree with the reviewer that this is not a new finding. We have added references and modified the text to avoid this potential confusion. In regard to the sun-glint retrievals, it would be interesting to test our model on them in the next iteration. We added discussion on this point in the discussion section.

6. The authors have chosen to use version 1.2.2 of the official TROPOMI NO2 data product. This is a pity since there have been a few important retrieval improvements and reprocessing efforts since then. These have been discussed in van Geffen et al. (2022) and Riess et al. (2022), and especially the more accurate FRESCO+ cloud heights from version 1.3 onwards are important for the author's purpose. In version 1.2 the too low cloud heights led to a substantial low bias in the TROPOMI NO2 columns. This low bias has been partly resolved in the later versions, and results in better detectability of ship NO2 plumes. I think it would be a sorely missed opportunity to not apply the method on version 1.4 or later.

Thank you for this information. It sounds like a great update to the data. We added in the discussion section that such new data may provide opportunity to expand and improve the results in the future.

7. The most pressing concern I have with the study is the lack of uncertainty analysis and the lack of validation. Although the (climatological) results presented appear plausible, a success metric is missing – what fraction of the 'NO2 pixel frequency' could have been misclassified? Besides retrieval uncertainties, also one/two-pixel size plumes, mis-training or mislabeling by the human, and the influence of plumes from land likely play a role. The authors should do more to address this issue, for example by comparing their (relative) frequencies to the locations of actual ships as stored in AIS-data or inventories on ship emissions,

We agree that more can be done to characterize such uncertainties. We have added discussions in the final section on this topic that outlines the need and challenges involved.

Some relevant literature is not cited or discussed:

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 Sensing, 14(22), 5809.

Kurchaba, S., van Vliet, J., Verbeek, F. J., & Veenman, C. J. (2023). Anomalous NO2 emitting ship detection with TROPOMI satellite data and machine learning. Remote Sensing of Environment, 297, 113761.

Riess, T. C. V. W., Boersma, K. F., Van Vliet, J., Peters, W., Sneep, M., Eskes, H., & Van Geffen, J. (2022). Improved monitoring of shipping NO< sub> 2</sub> with TROPOMI: decreasing NO< sub>< i> x</i></sub> emissions in European seas during the COVID-19 pandemic. Atmospheric Measurement Techniques, 15(5), 1415-1438.

Riess, T. C. V. W., Boersma, K. F., Van Roy, W., de Laat, J., Dammers, E., and van Vliet, J.: To new heights by flying low: comparison of aircraft vertical NO_2 profiles to model simulations and implications for TROPOMI NO_2 retrievals, Atmos. Meas. Tech., 16, 5287–5304, https://doi.org/10.5194/amt-16-5287-2023, 2023.

van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J.-C., Sneep, M., ter Linden, M., Ludewig, A., Boersma, K. F., and Veefkind, J. P.: Sentinel-5P TROPOMI NO_2 retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data, Atmos. Meas. Tech., 15, 2037–2060, https://doi.org/10.5194/amt-15-2037-2022, 2022.

Technical corrections

L17: "Straight" – Strait

Fixed.

L35: "over the global" – throughout the world?

Fixed.

L36: "the aggregated plumes … miss other routes" – plumes cannot 'miss' routes. The absence of plumes could be an indication that the method is incapable of detecting plumes in some routes,.

Fixed.

L36: "block the signals" – make clear what signals are meant here

Fixed.

L45: "show up as a long linear and bright" … as a long linear and bright cloud?

Long linear and bright features.

L63: here a discussion of NO2 retrieval characteristics for ship plumes that were the in-depth topic of investigation in the papers by Riess et al. (2022; 2023) should be discussed.

These are indeed very important new development. We added discussions here and in the discussion section.

Figure 4: top middle panel states 'NO2 emission' which should be (annual mean) NO2 column.

Fixed.

L280-285: a discussion of the findings in Riess et al. (2022; 2023) is missing here.

L302-305 and L324-326: first names of authors should be replaced by their last names for these references.

Fixed.

L392-395: this paper has already been published in AMT and should be cited accordingly.

Fixed.