Automated detection of atmospheric NO$_2$ plumes from satellite data: a tool to help infer anthropogenic combustion emissions
AMT - Reviewer Response

Major comments:
Line 40. It would be helpful to clarify how and why higher resolution emission estimates in time and space are necessary.

We have now clarified this statement:

The importance of accurate emission estimates becomes even more prevalent at smaller geographical and temporal scales for which, for example, hotspots and diurnal variations will play a larger role in driving observed atmospheric concentrations.

Line 132 and Figure 2, the example plum images presented here seem to be subject to winds. How does a plum looks like when there are no winds? downwind, upwind etc.. Are these considered when the authors created the training data ('the ground truth')? How the meteorological conditions and 'the ground truth' introduced to the model are related? Would the authors like to claim that such conditions related to meteorology would not be a problem when averaging across the study period (time-averaging for detecting plum locations only)?

This is a good question. We have clarified this point in the text after Figure 2:

For an emission source to create a plume detectable by TROPOMI, the source must be subject to winds strong enough to disperse the emissions across multiple pixels within the lifetime of NO2. We anticipate that the number of occurrences where these conditions are not met will be relatively small compared to the entire dataset and therefore should not have an adverse effect on our results.

Line 144. In addition to the point above, a discussion regarding the authors' judgement should be presented in the manuscript.

We have added the following statement:

Subjective judgement of the images could lead to small variations in repeated experiments and therefore a more rigorous approach may be needed for future applications.

Line 184. If the seasonal dependency is expected but not presented, can you elaborate a bit more about the temporal aspect of the presented approach?

Although we present all detections over the two year study period, we attribute plumes to biomass burning based on daily VIIRS biomass burning data - corresponding to the date of the plume. This means seasonality of the location of the biomass burning plumes will be present in the final dataset. We have added the following statement to clarify:

We account for the seasonal variation of fire activity by using daily VIIRS data, and remove fire-influenced scenes from those identified by our CNN.

Line 278-281. I appreciate the authors included the raised concern from the last review. However, it would be helpful to mention general issues for false detections, which can be caused by TROPOMI retrieval or the model (authors' judgement) itself

We have included the following statement:

As well as errors in the TROPOMI retrieval leading to false detections in the final dataset, errors may also occur during the creation of the model (e.g. mislabelled training data). A single plume data point may not represent a real-life plume and should be considered in context of other data (e.g. frequent recurrence, land use, proximity to other sources). Further refinement of the training dataset, model parameters and data analysis stages will reduce the number of false detections and feedback to the TROPOMI community could help reduce the number of retrieval errors.

Minor comments:
Line 33. The authors mentioned that the emission inventories rely on assumptions that lead to in-accurate values. Which assumptions can lead to inaccuracy?

We have changed the sentence to read:

Compiled inventories, which rely on self-reporting, provide estimates on these emissions but rely on assumptions such as fuel consumption, combustion efficiencies and emission rates that will be imperfect.

Line 36-37. The authors claim that China and India are going to build new coal-fired power plants. Please add references.

Given the timelines associated with the peer-review process, these references are in the "reputable" mainstream press and grey literature. For example,
https://time.com/6090732/china-coal-power-plants-emissions/
https://www.theguardian.com/environment/2021/oct/12/china-coal-fired-plants-uk-cop26-climate-summit-global-phase-out
https://carbontracker.org/reports/do-not-revive-coal/
https://www.reuters.com/business/energy/cop26-aims-banish-coal-asia-is-building-hundreds-power-plants-burn-it-2021-10-29/

With this in mind, we think it is appropriate not to include any one reference to accompany our statement.

Line 92. The authors mentioned that regardless of changes in TROPOMI resolution from August 2019, the original resolution was used. Is it possible to use the same resolution when using

Level2 data? Or Does it mean that you re-gridded the TROPOMI data with the same resolution to make each pixel the same distance?

This statement contained an error in that the 7x3.5 km resolution is the higher resolution product that became available from August 2019 but the updated processing was retroactively applied to the data collected before that date so all data used is of the higher resolution.

We have removed the line to avoid confusion:
'*Higher resolution data are available from August 2019, but for consistency we have used the original resolution throughout our study period.*'

Line 120. references missing

There does not appear to be a reference missing on this line. We reference Srivastava et al. (2014) in regards to dropping convolutional layers in the model.

Line 121. dropping another 50% of the 'layers'

Changed.

Line 129. For the normalization of images, the authors mentioned it was to improve model efficiency. Is it correct? Isn't it a 'necessary' step for training the model without diverging?

We have updated this line to read:

Images...were individually normalised to remove the influence of the magnitude of NO2 features, a step that also ensures the model parameters have a similar data distribution and therefore improves the model efficiency and accuracy.

Line 152. 'our quality threshold' -- does it mean the TROPOMI quality flag?

We do mean the TROPOMI threshold and the text has been amended.

Line 242. 'analysis of of'

Changed.

Line 250. references missing

As this is our assumption based on the work referenced in the introduction we do not think a reference is needed here.