Comments from anonymous Referee #2:

We would like to thank the reviewer for his/her helpful comments. We hope that we could address all questions and unclear points satisfactorily.

In the course of the revision, we have made the following important changes:

Based on a suggestion from Referee#2 we have looked into the TROPOMI AOT product. We added daily maps of the TROPOMI AOT in the Appendix of the manuscript. The lower branch visible in the TROPOMI PAL versus AirMAP comparison is mainly caused by data from 17 September (Fig. A9) and was discussed to be likely caused by a higher aerosol load which is identified as cloud in the retrieval and not treated adequately in the cloud correction, ending up with too high cloud pressures. This discussion can now be supported by the TROPOMI AOT data, which is showing a high AOT over a large area on 17 September.

During the corrections in the review process we found that the tropospheric NO2 VCD retrieval for the IUP car DOAS used an incorrect AMF of 1.5 instead of 1.3. This was corrected and Fig. 6 was updated. The correlation between the AirMAP and car DOAS measurements remains unchanged at 0.89, but the slope decreased from 0.98 to 0.89.

Referee#3 questioned the use of a NO2 box profile for the AMF calculations for the AirMAP flights, which we have stated in the text. This was an outdated information which we overlooked during the correction phase. The SCIATRAN tropospheric AMF calculations used in the AirMAP tropospheric NO2 VCD retrieval shown in the manuscript are not based on a 1 km box profile but are using a NO2 profile based on an old WRF-chem model run following a more typical urban profile, scaled to the ERA5 boundary layer height, which reached typical values of 1 km around noon. The NO2 profile is added to the Appendix.

Legend: Referee comments in black, author comments in blue

The manuscript by Lange et al. discusses the S5P-VAL-DE-Ruhr validation campaign. It includes a very extensive and well-presented validation of the TROPOMI dataset with aircraft and ground-based measurements, including a good measurement campaign overview. It further includes a comparison of different TROPOMI NO2 datasets and shows the significant improvements of the latest product version PAL over OFFL. It is of interest to readers of AMT. I would recommend publication after addressing some suggestions (see supplement).

General suggestions:

Sect.3: Why are the fitting windows so variable from instrument to instrument (DOAS instruments)? Please comment on this in the instrument description. Looking at the comparison it doesn't seem to make much of a difference, but I would suggest commenting on this.

Thank you for the suggestion, we now included a comment on this in the instrument description. The fitting window is restricted by the spectrometer's wavelength ranges, which are different for the different instruments. The exact fitting window was either chosen to be close to the CINDI-2 fitting window (425-490 nm) or from the groups experience with their instruments. Differences using the different fitting windows were found to be small. Differences using the different fitting windows were found to be small.

Furthermore, we added additional columns with spectrometer wavelength range, fitting window, and information about the VCD calculation and AMF used to Table 2.

Sect. 3, make sure to include uncertainty estimates of all instruments and include references to existing validation papers. The sections 3.3.4, and 3.4.1 to 3.4.4 lack references.

Uncertainty estimates are included for all instruments in the respective figures.

We added references to studies/campaigns in which the instruments participated and some studies with more details about the data retrieval, as far as available.

l. 550-555: I would suggest to check the TROPOMI AOT in the area for that day, available from https://data-portal.s5p-pal.com/; e.g. you could include maps of this in the appendix or at the very least state the average (and std) for the flights. Alternatively, VIIRS AOD or MODIS MAIAC AOD are very good AOD products (but this might be a little too much effort to include here, maybe just something to keep in mind for future studies).

Thank you for your suggestion. We checked the TROPOMI AOT data during the campaign period for the different flight areas. We added daily maps of the TROPOMI AOTs in the Appendix (see Fig.1).

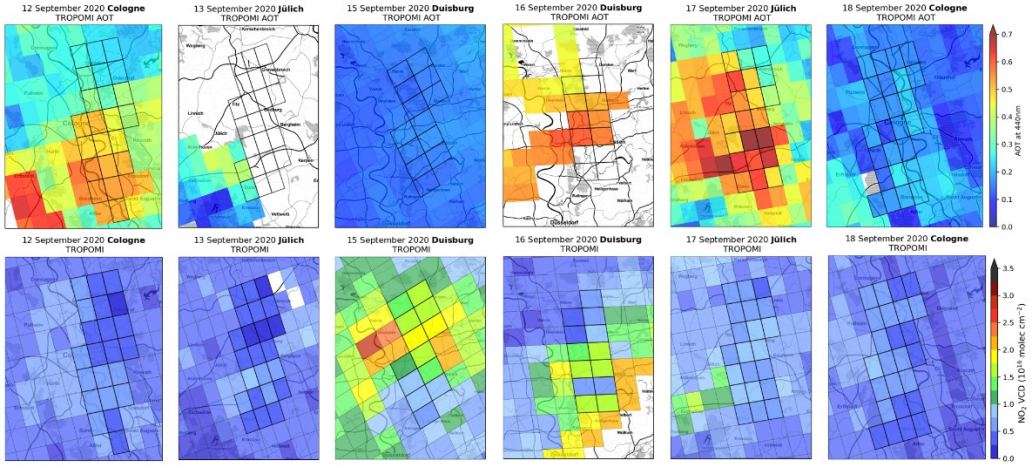

Figure 1: Daily maps of TROPOMI AOT with qa_value > 0.5 (top row) and TROPOMI PAL V02.03.01 tropospheric NO2 VCDs with qa_value > 0.75.

The observed AOT is quite variable, but it is good to see that there is no correlation between the TROPOMI AOT and NO2 data product. The lower branch visible in the TROPOMI PAL vs AirMAP comparison is mainly caused by data from 17 September and was discussed to be likely caused by a higher aerosol load which is treated as clouds in the retrieval and not accounted for adequately in the cloud correction, ending up with too high cloud pressures. This discussion can now be supported by the TROPOMI AOT data, which is showing a high AOT over a large area on the 17 September.

l. 619: just a thought: there was no snow during the campaign; there could be a larger difference between the DLER and LER TROPOMI product for snow covered surfaces with high reflectivity. It would be nice to include a little comparison of the DLER and LER product for snow covered surfaces. Nothing extensive, just a sentence (near l. 619) and a scatter plot (TROPOMI DLER vs TROPOMI LER) in the appendix (if time permits).

We added a sentence to clarify that differences between the DLER and LER TROPOMI product can be larger about snow-covered surfaces and that the found differences are only valid for the campaign area and time. We only have the data available for the campaign period and region. Nevertheless, we did some additional comparisons with the available data. Figure 2 shows the TROPOMI tropospheric NO2 VCD retrieved with TROPOMI LER respectively TROPOMI DLER for (a) the 117 TROPOMI pixels

used throughout the study in comparison with AirMAP, showing a slope of 1.04, with slightly higher NO2 VCDs for the TROPOMI NO2 product using the TROPOMI LER in the AMF calculations instead of the DLER. Figure 2b shows the same comparison for a larger orbit segment over western Europe on 13 September 2020 and Fig. 2c the complete orbit, both showing a correlation coefficient of 1 and a slope of 1.03. All data are quality and cloud filtered using a qa_value of 0.75. Since the observations are made in September, no larger snow covered areas are expected and a more detailed analysis including a larger period and area would be needed. We included Figure 2 in the Appendix of the manuscript.

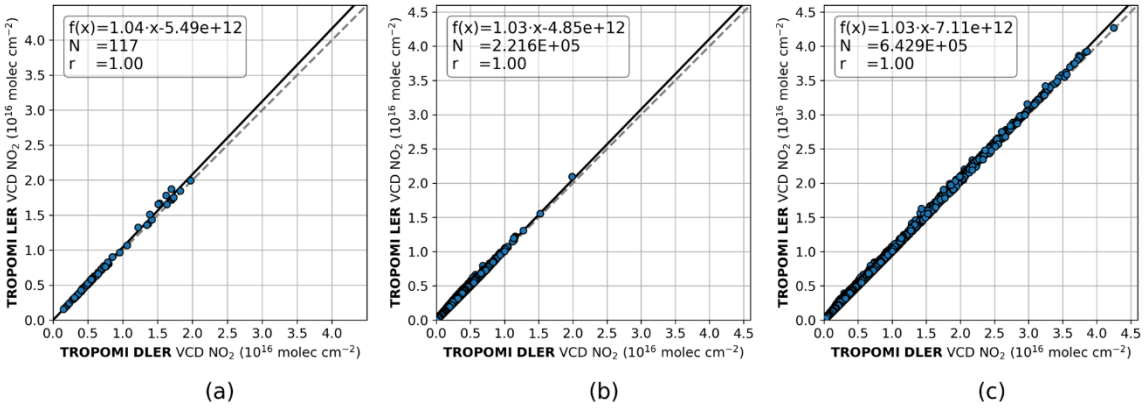

Figure 2: Scatter plot of TROPOMI IUP V02.03.01 tropospheric NO2 VCDs with TROPOMI LER respectively TROPOMI DLER for: (a) the 117 TROPOMI pixels coinciding with the AirMAP measurements used throughout the study, (b) a larger orbit segment over western Europe on 13 September 2020 and (c) one full orbit including the campaign area on 13 September 2020.

Comment added in the manuscript: "Larger differences could for example be expected for snow-covered surfaces with high reflectivity. Figure A10 in the Appendix shows scatter plots of the TROPOMI tropospheric NO2 VCD retrieved with TROPOMI LER and TROPOMI DLER for the 117 TROPOMI pixels used throughout the study but also for larger areas up to one full orbit. All comparisons show only minor influences by the directional component. Since only TROPOMI observations made in September are compared, no larger snow-covered areas are expected and a more detailed analysis including a different period and area would be needed to investigate possible larger differences."

Technical/minor suggestions:

l. 1: suggesting to change to "Airborne, ground-based stationary and car imaging differential optical absorption spectroscopy (DOAS) measurements…"

Since not all mentioned measurements are imaging, we did not implement this proposed change.

l. 4: emitters -> sources

Changed

l. 5: "The DOAS measurements…"

Changed

l. 7… suggest using "observations" instead of measurements throughout the text (technically measurements are in situ measurements), and remote-sensing are observations

Thank you for the comment. Since it is common in the community to use the term measurements also for the here described cases we stayed with this usage but tried to change it to observations in some more specific descriptions.

l.13: data create

Changed

l. 13-31: This paragraph can be shortened, I think the most important points are: 1) The PAL version improves the bias significantly, 2) cloud height and NO2 profile have a major impact, 3) surface reflectivity has a minor impact (in this region and time of year – this is likely different when there is snow on the ground). These points get a little lost in the lengthy paragraph, maybe include the correlations and biases inside the sentence in brackets rather than writing whole sentences about it (a little repetitive).

Thank you for the suggestion. We slightly shortened this paragraph to highlight the major findings.

l.36: change to "…combustion processes, such as power plants and engines, as well as anthropogenic biomass burning." (How much does anthropogenic biomass burning impact the Ruhr area? Consider removing the last half of the sentence.)

It is right that anthropogenic biomass burning has not a large impact in the Ruhr area, but since in this section of the introduction the sources of NOx are described more in general, we left it like this.

l. 37: "NOx is primarily emitted as NO, the reaction…"

Changed to "NOx is primarily emitted as NO, which is reacting with ozone (O_3) and is rapidly forming NO_2."

l.37: "The NOx sources are …" (remove the characteristics)

Done

l. 38: chemically active -> reactive and short lived

Changed

l. 38/39: "..., there is a high spatial and temporal variability of NO2 near emission sources." (there is not much variation in background areas)

Changed to "As a result, the spatial and temporal variability of NO2 is large, especially in regions characterized by a variety of NOx emission sources."

l. 39: remove "on"

Done

l.44: "is remotely observed from different platforms" -> " can be observed remotely on a variety of platforms"

Changed to "is remotely observed on a variety of platforms"

l.46f: is identified -> can be identified; are -> can be

Done

l. 53: remove “on board the European…satellite.”

Removed

l. 55 TEMPO is planned for launch in March 2023

Changed it to 2023.

l. 56-58: consider re-wording this sentence, maybe change it to 2 shorter sentences

Rewritten into two sentences.

l. 86: include a sentence about the new TROPOMI version and what changed/improved in comparison to the previous version. Maybe include studies that validated this new version, I know of Zhao et al. (2022) see reference list, there might be others too. E.g. Riess et al (2022) also talks about the improvements of the new TROPOMI version.

We added a few sentences about the new TROPOMI version and included the study of van Geffen et al. (2022) which is also comparing the new tropospheric NO2 VCD version with MAX-DOAS data. We have not included the suggested studies since Zhao et al. (2022) is doing validation only with total columns and Riess et al. (2022) is using the new version for monitoring shipping NOx emissions but is not including validation.

l. 89: industrial estates -> industrial facilities, arterial highways -> busy highways (or large highways)

Included the first suggestion, second left like it is.

l. 89: “Back-ground areas with low pollution, as well as moderately polluted regions are also observed…”

Changed

l. 96: remove “In the following,”

Done

l. 107: 5 million inhabitants -> has a population of 5 million

“The Ruhr area itself has over 5 million inhabitants.” changed to “The Ruhr area itself has a population of 5 million.”

l. 108f: “The region, including nearby metropolitan centres along the Rhine and populated surroundings is called Metro… and is comprised of a population of over 10 m, large power plants [can you include a number here], … industrial facilities and several large highways.”

Thank you for your suggestions. We changed it to:

"Together with the populated surroundings and metropolitan centers along the Rhine, the region is called Metropolitan area Rhine-Ruhr (MRR). It comprises a population of more than 10 million inhabitants, large power plants, energy intensive industrial facilities and several large highways."

l. 110: above the campaign location -> in the MRR

Changed

l.120:"… dominated by the emissions of three lignite fire power plants in the area (see European Pollutant Release and Transfer Register (E-PRTR)[include reference or url here]).

Changed to "The research flight area around Jülich is expected to be dominated by the emissions of three large lignite fired power plants located in the area (see European Pollutant Release and Transfer Register, https://industry.eea.europa.eu/, last access: 18 November 2022)"

l. 121: "…around Cologne and Duisburg…" , remove latter part of the sentence "and the flight area around Duisburg has a similar character to that of the Cologne area with a mixture of urban and industrial emitters but includes the central metropolitan Ruhr area, which has a large variety of pollution sources. "

Shortened but kept the information about the difference between the two areas.

l. 137: here and any other occurrences: don't shorten Table to Tab. as per AMT guidelines, "are given in Table 1"

Done

l. 143: "…comparison of the aircraft and TROPOMI NO2…"
l. 143: remove "prior to the dedicated evaluation … satellite pixel area", I'm not sure what you mean here, but I think it's not necessary

Rewritten this passage. We hope it is clearer now.

l. 144f: remove: "In this manner", " on the one hand", "on the other hand"

Removed

l. 145: local -> ground-based , with restricted -> that have restricted, with satellite -> to satellite

Changed

l. 151: Table 2

Changed

l. 157: of aerosol -> aerosol

Changed

l. 160: thus by far -> currently

Changed

l. 167: AMFs. The AMFs are generated using the OMI…

Not changed, as we think it would change the meaning in a wrong way.

l. 167: , and cloud fraction

Changed

l. 171: (e.g. Verholst et al, 2021). (there are many others)

Changed

l. 173f: very long sentence, consider numbering the reasons and removing unnecessary details: " other factors that could contribute are: (1)…"

Thank you for your suggestion we restructured a bit and numbered the reasons.

l. 186: recipe provided-> approach as detailed

Changed to "approach described".

l. 231: Tab.->Table

Done

l. 255: remove "used"

Done

l. 261: 438-490nm; is this the same or different to TROPOMI?, and other DOAS fits used in this study

The AirMAP fitting window is different from the TROPOMI fitting window. Due to different spectrometer wavelength ranges it is not possible to use the same fitting window.

TROPOMI: 310 - 500 nm wavelength range, fitting window: 405 - 465 nm

AirMAP: 429 - 492 nm wavelength range, fitting window: 438 - 490 nm

We included a comment on the different fitting windows in the instrument description. See comment in general suggestions.

l. 265ff: it's unusual to trop as a superscript, subscripts would be more common, e.g. subscript of trop,ref could be used

Changed

l. 271: during -> near

Changed to "close to".

l.272:change to: " There is a maximum difference of 3h between the time of the reference background and the actual measurements."

We changed the sentence.

l. 276: how small, do you have a reference that it is negligible?

We have written "We assume that the effect of the changing solar zenith angle (SZA) and the diurnal variation of the stratospheric NO2 concentration are small, and a stratospheric correction of the measurement data is therefore not necessary." We added that this is only true during the measurement time around noon and not during twilight and included a reference to Schreier et al. 2019, where this was analyzed for car DOAS measurements in Vienna (see Fig. 3).

Schreier et al. 2019: "The uncertainty of the diurnal variation is large at twilight but small during the day as changes in stratospheric NO2 are small when compared to tropospheric NO2 columns in polluted regions, such as the urban area of Vienna. As a rough estimate, the uncertainty of the stratospheric correction is assumed to be less than 10% or typically 1×10^{15} molec/cm^2."

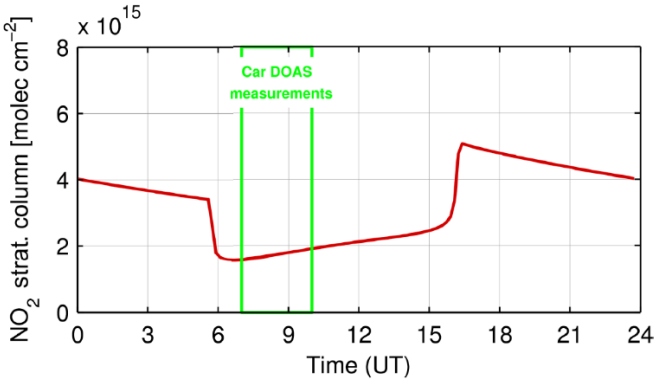

Figure 3: Figure 5 from Schreier et al. (2019): Stratospheric NO2 above Vienna on 19 October 2014 (red line) as obtained from the Bremen 3d chemistry transport model (B3dCTM). The green rectangle indicates the time period of car DOAS measurements performed on that day.

l. 279: (see ERA5 reanalysis; Hersbach et al. (2018). You can include the data source in the data availability section, and data source reference

Changed

l. 298: comprises of

Not included

l. 300: remove "further details can be found therein."

We removed this part.

l. 352: why is the spectral window different to the AirMAP window?

We included a comment on the different fitting windows in the instrument description. See comment in general suggestions.

l. 371/379: again why is the spectral window different?

We included a comment on the different fitting windows in the instrument description. See comment in general suggestions.

l. 397: “different target areas” -> “three target areas”

Changed

l. 411: do you have a reference for this? This paragraph in general could benefit from a couple of reference.

We wrote: “An uncertainty of 30% for the SCD in the reference spectrum is assumed“. Unfortunately we do not have a reference for the given uncertainty.

We inserted a reference for the instrument setup and the Langley-plot method.

l. 413: The same AMF (1.3) is used to convert to a VCD, wouldn’t it depend on the SZA? How much of an impact does the SZA have?

We agree that the AMF depends on the SZA. Since we only used data close to the AirMAP overpass, which was performing measurements around noon, the SZA is not varying much. In the paper we have referred to the work of Merlaud (2013), who analyzed the AMF distribution for a large number of simulations, resulting in a mean of 1.33 ± 0.2 and 2.52 ± 0.32 for the AMF for measurements in 90° respectively 30° viewing zenith angle (see Fig. 4 and Table 1).

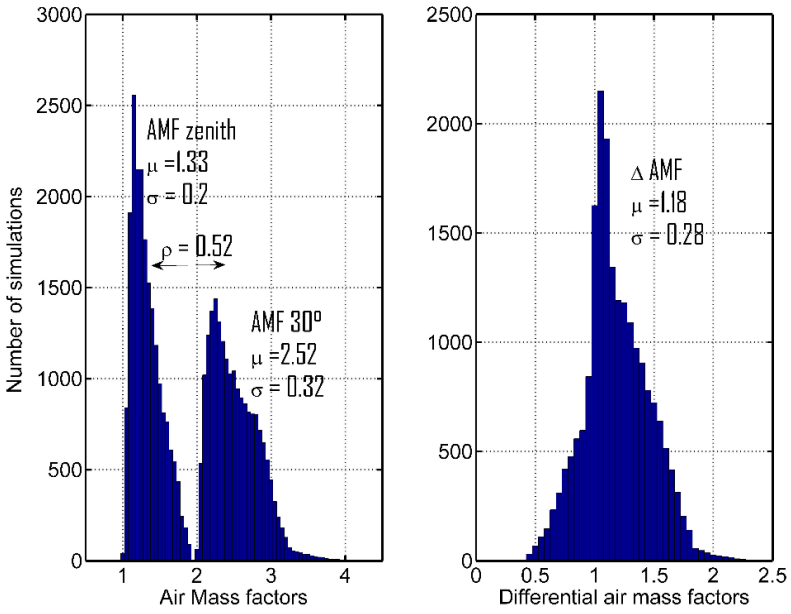

Figure 4: From Fig. 7.3 from Merlaud (2013): Distribution of air mass factors calculated with the parameters of table 7.1(left) and of the resulting differential air mass factor (right).

Table 1: Table 7.1 from Merlaud (2013): Parameters and ranges used in the air mass factors calculations.

Wavelength(nm)	460
Direction	Zenith,30°
Surface visibility(km)	5, 10, 15, 20, 25, 30
NO_2 mixing height(m)	100, 300, 500, 700, 900
Relative azimuth(°)	0, 30, 60, 90, 120, 150, 180
Solar zenith angle(°)	20,30, 40, 50, 60, 70, 80, 90
Albedo	0.01, 0.03, 0.05, 0.07, 0.09, 0.11, 0.13, 0.15

l. 423: insert reference here

We included a reference.

l. 440-450: what is the uncertainty of the PANDORA observations. Also please include references in this paragraph.

We included additional references to the Pandora instrument and data product in this paragraph. The uncertainty provided with the tropospheric NO2 VCD is described in Cede et al. (2021), we added this reference to the text.

l. 465: is it really +- 10-90 percentile, I think it is just the 10th and 90th percentile.

Thank you for pointing this out, we have changed it to “the error bars are representing the 10th and 90th percentile” throughout the text.

l. 470: data is -> data are

Changed

l. 473: As a result of having more opportunities to make near simultaneous synchronized measurements, -> Consequently,

Included

l. 487: is it really +- 10-90 percentile, I think it is just the 10th and 90th percentile.

Thank you for pointing this out, we have changed it to “the error bars are representing the 10th and 90th percentile” throughout the text.

Fig. 7: I would suggest moving one of these (I suggest Fig. 7b) to the appendix, as they essentially show the same.

Thank you for the suggestion. We removed Fig. 7b completely since we only used it here to point out the time difference between the AirMAP and car DOAS measurements as a reason for the outlier. We added a reference to Fig. A 7 in the Appendix, in which we kept both plots.

l. 637: areas -> areas,

Included

l.642: scatter -> scatter,

Included

l.642: measurements -> measurements,

Included

l. 687: dataset -> data set (either or but be consistent, I found both dataset and data set used in the study, please fix this)

Changed “dataset” to “data set” throughout the text.

l. 704: Sentinel-4 -> Sentinel-4,

Included

References:

Zhao, X.; Fioletov, V.; Alwarda, R.; Su, Y.; Griffin, D.; Weaver, D.; Strong, K.; Cede, A.; Hanisco, T.; Tiefengraber, M.; McLinden, C.; Eskes, H.; Davies, J.; Ogyu, A.; Sit, R.; Abboud, I.; Lee, S.C. Tropospheric and Surface Nitrogen Dioxide Changes in the Greater Toronto Area during the First Two Years of the COVID-19 Pandemic. Remote Sens. 2022, 14, 1625. https://doi.org/10.3390/rs14071625

Riess, T. C. V. W., Boersma, K. F., van Vliet, J., Peters, W., Sneep, M., Eskes, H., and van Geffen, J.: Improved monitoring of shipping NO2 with TROPOMI: decreasing NOx emissions in European seas during the COVID-19 pandemic, Atmos. Meas. Tech., 15, 1415–1438, https://doi.org/10.5194/amt-15-1415-2022, 2022.

Cede, A., Tiefengraber, M., Gebetsberger, M., and Spinei Lind, E.: Pandonia Global NetworkData Products Readme Document, Tech. rep., PGN-DataProducts-Readme, version 1.8-5, 31 December 2021, last access: 2 December 2022, 2021.

Merlaud, A.: Development and use of compact instruments for tropospheric investigations based on optical spectroscopy from mobile platforms, Presses univ. de Louvain, 2013.

Schreier, S. F., Richter, A., and Burrows, J. P.: Near-surface and path-averaged mixing ratios of NO2 derived from car DOAS zenith-sky and tower DOAS off-axis measurements in Vienna: a case study, Atmospheric Chemistry and Physics, 19, 5853–5879, https://doi.org/10.5194/acp-19-5853-2019, https://acp.copernicus.org/articles/19/5853/2019/, 2019.