Comment on 'Intercomparison of four airborne imaging DOAS systems for tropospheric NO2 mapping – The AROMAPEX campaign' by Frederik Tack et al.

General Comments

Tack et al. present the results of the 2016 AROMAPEX campaign. Four different imaging instruments simultaneously recorded reflected/scattered sunlight spectra over Berlin. The retrieved NO2 VCD maps are compared and found to show good agreement.

Similar data of the individual instruments have been published before. However, to the best of my knowledge, the intercomparison of airborne imaging DOAS datasets simultaneously recorded by different instruments hasn't been done before. This might be useful for calibration/validation campaigns for future satellite instruments. The paper also represents a state of the art of airborne NO2 imaging measurements combined in one study. The paper fits the scope of AMT and is well-structured. However, before publication, substantial points need to be clarified and added to the study.

Specific Comments

1) VCD spatial resolution

In Sect. 3, the individual imaging instruments are introduced with their respective spatial resolution given by e.g. 80m x 60m for an APEX pixel. It is the surface that is covered by the pixels FOV at ground level. This spatial resolution is then, if I understand correctly, also assigned to the retrieved VCDs. However, in the AMF retrieval 3D effects of the radiative transport are not taken into account. On this high spatial resolution, the assumed NO2 layer of 1km, the large SZAs and the inhomogeneous NO2 distributions, 3D effects will dominate the uncertainties and reduce the effective spatial resolution of the VCD maps (of e.g. APEX and AirMAP) by up to 2 orders of magnitude.

This should be included into the error budget and indicated in the captions of the VCD maps (Fig. 12, 13).

2) Validation with ground based DOAS

The study aims at validating the VCDs of satellite measurements. However, retrievals similar to satellite retrievals are used.

Ground based DOAS measurements can deliver tropospheric VCDs with strongly reduced uncertainty due to their much simpler geometry (e.g. Tack et al., 2015; Brinksma et al., 2008). The reference can be taken at the same location as the airborne reference. The validation of the presented airborne VCDs maps with e.g. zenith mobile DOAS data would drastically increase the scientific quality and significance of the study.

In 'Inter-comparison of airborne atmospheric imagers during the AROMAPEX campaign' (http://www.eufar.net/weblog/2016/06/15/inter-comparison-airborne-atmospheric-imagers-during-aromapex-campaign//, last access: 30.04.2018), Magdalena Ardelean and Alexis Merlaud state that both mobile DOAS and stationary MAX DOAS measurements have been performed during the AROMAPEX campaign.

3) What does the acronym AROMAPEX stand for? If the 'RO' still stands for Romania, why did the campaign take place in Berlin? This might be interesting regarding the submission to the 'AROMAT' special issue.

4) p.2, l.30: Sentinel 5p is already in operation.

5) In Sect. 3, please indicate an approximate detection limit for the dSCDs for the individual instruments in the setup used during the campaign. The VCD maps strongly differ by structures of weaker but still

seemingly significant NO2 VCDs. I think a map with the NO2 fit error or the RMS of the DOAS fit residuals of the individual instruments would be revealing, especially because of the large differences in the DOAS retrieval parameters.

6) Table 3: Why is there no water vapour and O3 absorption cross section used for the DOAS analysis of APEX and no water vapour and no O4 for SBI? Especially when fitting above 500nm the water vapour and O3 absorption cross sections strongly increase. And the SZA differences are significant during a single flight with only one reference.

The results of the DOAS evaluations of e,g, AirMAP could be used to motivate leaving out these species in the evaluations of the other instruments.

7) In Sect 4.1 p.7, l24: 'The differential approach (1) largely reduces systematic instabilities…' compared to what?

8) For the SBI the dataset used in the intercomparison for the morning flight is reduced (only 10 overpasses), while the other instruments deliver data for 14 overpasses. The reason for that should be given.

9) Section 4.2.2.1: The retrieved surface reflectances are compared. AirMAP's surface reflectances are retrieved for the DOAS fit wavelength range and the spatial resolution used in the discussed measurement. They are however compared to two 'APEX surface reflectance products', both having a much higher spatial resolution ('4 by 3 m^2'). As far as I understand, the APEX AMFs are calculated with an 80 by 60 m^2 resolution. Is the high resolution of the surface reflectances taken into account in the retrieval? If not, I would suggest to compare surface reflectances with the spatial resolution of the respective AMF retrieval.

Also the choice of 490-500nm for the surface reflectance retrieval for APEX seems arbitrary and should be motivated (why not 470-510nm?).

10) 4.4 Error budget

a) The argument that a larger FOV per pixel results in more collected photons is only true if all optics use the same effective aperture. The light throughput is determined by the etendue (beam solid angle x effective aperture) of the optics.

b) sigma_scd_ref is included in the error budget as a statistical error. Howerver, it is, as I understand it, an unknown offset. An offset shouldn't be treated as a statistical error.

c) The error analysis should include a discussion of the error introduced by the 3D radiative transfer effects (see Comment 1).

11) p.17, l.12: The artefact in the south of the map is assigned to an eventual spectral structure in the reflection of a specific crop type. This would be interesting. Is there a specific residual structure observed in all affected spectra?

There are significant differences in the DOAS fits used for APEX and AirMAP. Particularly, the APEX fit does not include water vapour, even though the water vapour absorption is much stronger in the APEX evaluation interval compared to the fit interval used for AirMAP. A map of the RMS of the DOAS fit residuals (see Comment 5) would be instructive here.