We are grateful to Christoph for his very helpful comments. We have revised the manuscript to take his suggestions into account, and will discuss the most significant of these herein.

P. 5685, L.16 and Fig. 4: is surface intensity really the correct term here? Or should it rather be just “measured intensity”? Or are those raw DNs? Are the data atmospherically corrected? Please clarify. Also, maybe a color bar should be added to the Fig.

We agree that the term is incorrect. As we did not perform any radiometric calibration on the instrument, the measured spectra are in the form of raw digital numbers from the CCD. We have since corrected the manuscript to refer to this. As these are raw CCD output we do not see the need to include a colour bar to Figure 4, as the purpose of this image is to demonstrate that a representative intensity map can be produced from the raw data using the interpolation method described in the manuscript.

P. 5686: Please justify how the fitting window was chosen. Have you also tested different ones? The fitting window given in the text and the one in Tab. 2 slightly differ, please correct or explain. A little more information about the spectroscopy here would be nice in general. For example, what’s the outcome of the wavelength calibration? Have you also used a solar reference spectrum to determine spectral shifts and spectral resolution?

The different fitting windows in the manuscript were the result of a typing error. We have corrected this mistake. The fitting window used was 432-493 nm

The fitting window was selected to be reflective of similar studies, permit fitting over a broad wavelength range, and to minimise the fit RMS. A number of wavelength windows (approximately 20) were tested, with the final configuration selected on the criteria above. We did not perform an iterative selection, or a quantitative optimisation in this study.

The wavelength calibration was performed by fitting the Fraunhofer lines of the spectra with the aid of a solar reference spectrum (Kurucz et al, 1984), using a NLLS fit as discussed in the QDOAS literature (Fayt et al, 2015). From this, a wavelength dependent shift and Gaussian ILS was calculated over the 428-496 nm spectral window. This has been added to the manuscript.

P. 5687, L.6: please specify how exactly the fitting errors are calculated (or what is the “standard mechanism within QDOAS”?).

QDOAS determines the SCD errors using the reduced chi-squared statistic of the DOAS fit. Further information about this can be found in the QDOAS manual. An additional citation for this has been included on this line in the revised manuscript.

P. 5687, L.21: I think this approach is fine. However, I wonder what might cause this large spike around rows 80-90? Is that really just due to CCD artefacts?

To investigate these spikes we performed this analysis on varying temporal ranges. We found that rows 80-90 always featured some enhancements, which makes it likely that these spikes are at least partially due to CCD artefacts. It is possible that reference sector contamination may have also contributed to these features. We have these comments to the manuscript.

P. 5690, L.16ff: I'm a little confused about the nomenclature used here and in other parts of the study. To my understanding, radiometric calibration is the process that converts the initial digital numbers to (at-sensor) (spectral) radiances (e.g. in units of W $m^{-2}sr^{-1}nm^{-1}$). So when you use the term radiances here, do you mean the raw DNs instead to infer surface albedo? Please clarify.

As mentioned previously, we have erroneously referred to the raw digital numbers as radiance data. This has been clarified in the manuscript.

P. 5692, L.6: I don't quite understand the 98% here. Is this 98% of the average SCD? The calculation of the uncertainty should be defined somewhere. It would in addition be informative to also have the average SCD here.

This was supposed to be the dSCD error as a percentage of the average dSCD. However, due to a typing error (see below) this statistic was erroneously calculated. In reality, the average dSCD error is 37% of the average dSCD, and has been added to the manuscript.

P. 5697, L. 15: I cannot clearly see if those four stripes are really that temporally consistent, for example in region A the said junction reveals high NO_2 in the first and third flights (~4.8 x 10^{16} molec cm^{-2}) whereas the second and fourth flights reveal rather low NO_2 (~3.8 x 10^{16} molec cm^{-2}) It's also hard to detect the low NO_2 region within the city center. I suggest the authors add some arrows to that Figure to better highlight the discussed areas. It might also help to plot the NO_2 values along a common transect for the four flight lines. Further, if available ground-based in-situ NO_2 measurement should be integrated in the discussion here.

We have updated the Figure to show which regions we were referring to. However, owing to the differing flight paths it is difficult to determine a common transect over all four flight lines that quantitatively demonstrates the features we are referring to.

At the time of the flight the only ground measurements available was from a single chemiluminescence instrument situated in the city centre, which recorded hourly background NO_2 concentrations. Only a single overpass of this region was made by the flight, and the coarse temporal resolution of this instrument severely restricted any meaningful comparisons between background measurements made elsewhere by ANDI. As such, we are unable to offer comparisons with in-situ data with our measurements at this time. Future flights will be accompanied by a more extensive in-situ monitoring network to validate our VCDs.

P. 5701, L.20-23: Are the 0.7 x 10^{15} molec cm^{-2} the error on the SCD (like written on p5692; L.5 and 6) or on the VCD or dSCD like written here. Please clarify.

This was the result of another typing error. The section is meant to say that the mean dSCD over the flight is 1.9 x 10^{16} molec cm^{-2}, while the 7.0 x 10^{15} molec cm^{-2} error comes from the DOAS fit itself (see above). We have corrected the manuscript.

Fig. 14: What is the region with enhanced NO_2 to the north-east of Leicester, where the aircraft turned around? Is that an artifact or real? The retrieved NO_2 VCDs there are distinctively higher than in the city center but are to my knowledge not mentioned in the text at all. Please discuss and provide some details somewhere in the text.

The enhancement seen over that region is similar to those observed over the other banking manoeuvres during the flight, so it is possible that this is due to a large path length enhancement caused by a change in the roll angle. This would explain why the largest VCD appears to be at the outer edge of the swath. As the IMU data was corrupted it was not possible to adequately include these effects in the AMF computation, which may lead to features such as this appearing in the final dataset. We believe that such features will not have appeared if the IMU was correctly working, and envision that subsequent flights will not be subject to these effects.

References

Kurucz, R. L., Furenlid, I., Brault, J., and Testerman, L.: Solar flux atlas from 296 nm to 1300 nm, National Solar Observatory Atlas No. 1, 1984

Fayt, C., De Smedt, I., Letocart, V., Merlaud, A., Pinardi, G., and Van Roozendael, M.: QDOAS Software user manual, BIRA-IASB, http://uv-vis.aeronomie.be/software/QDOAS/index.php, 2015