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

The current spatial resolution is very good, but even if it was 40 m, it would still be good enough to resolve a lot of details. It might be checked if the dSCD are similar for low resolution data and the error is reduced, if so the high resolution data have a better basis.

We agree that the SNR could be improved by binning the spectra to a coarser spatial resolution. However, the goal of this work was to determine whether Leicester's NO_2 emission sources could be adequately mapped at this resolution, so we feel that this would be beyond the scope of this work. Furthermore, a coarser spatial resolution may adversely affect our ability to interpolate over the coverage gaps, which may result in additional artefacts. Future refinements to the ANDI design will consider this trade-off between spatial resolution and SNR.

Concerning the DOAS fit: to omit the offset correction in the DOAS analysis (P 5686) cannot be recommended, especially not when working with an imaging spectrograph in combination with a CCD … please add few comments about the influence of the fitted offset.

Fits with and without offset were completed as part of the analysis. The following text has been added to the paper.

"removal of the offset within the DOAS fit did not substantially change the spatial structure in the vertical column data, but marginally improved RMS-derived error estimates. An increase of approximately 20% in the fitted differential slant column density of NO2 was observed. Future analysis should include a detailed examination of the impact of the offset fitting, particularly if extending analysis through to retrieval of surface volume mixing ratios."

… what is the resolution of the instrument (20 m or 5 m)? Both numbers are mentioned in the manuscript and dividing 600 m by 128 pixels results in 5 m. Here a binning of 4 CCD lines is already included.

The spatial resolution of the instrument is 5 m. However, because of the temporal coverage issue we interpolated the raw data to produce an 80 x 20 m data product. We have clarified this in the manuscript.

The general problem of all the airborne DOAS instruments is that one measures relative to an unknown background. There were some attempts to overcome this problem, comparison with satellite instruments or modeled data, but the general uncertainty remains … So please do not set the results for the respective regions in relation, even if the error might be small. The total numbers speak for themselves, add a comment form time to time to remind the reader of the background problem.

We agree with the reviewer that the reference NO_2 column will bias comparisons made between different regions in the final dataset. We have added additional lines in the manuscript to remind the reader that any such comparison is done in the assumption that there is no significant difference in background NO_2 between the observations.

The spectral resolution of the instrument is 1.5 nm. However, in the paper by Whyte et al. (2009) it seems better spectral resolutions are possible with the instrument. Did the authors use a wider entrance slit to improve the light through put and hence spectral intensity?

A different grating has been used for the visible window optimised for NO_2 retrieval. The UV optimised grating was 2350 grooves mm^{-1}. In ANDI a grating of 1800 grooves mm^{-1} was used for the window specified in the paper. Additionally, in the Whyte et al (2009) breadboard design a fold mirror was incorporated. This was removed in the airborne iteration to simplify the payload and reduce internal stray light.

It is most likely that the degradation in the instrument line shape has occurred as a result of the build of the airborne version. Tolerances in the manufacture of the telescope mirrors, their mounts and the mounting on to the optical bench were not able to be compensated; therefore the alignment of light incident on the entrance slit is not optimised. We have added these points to the manuscript.

The frame transfer camera worked with 1 Hz, this means the data are recorded for roughly 1 second, (e.g. 0.99 s) transferred to the readout chip within a few milliseconds and read out for about 1 second while the next image is taken. If this is correct this should be considered in the figures and the interpretation of the data, because it means the data are averages of 80 m and not point observations. In figure 4 it looks like point measurements, which is of course easier to plot. Add a comment in the caption that the centre is taken here. After the temporal shift correction and the interpolation this should be very similar. Does the time stamp of the measurements belong to the beginning or to the end of the measurements?

The frame transfer rate quoted in this work is the total time elapsed between subsequent measurements. Once the CCD has been exposed for a single frame (300 ms), charge is transferred to the covered storage areas. When this happens, the vertical clocks are stopped on the exposed part of the CCD, and the data is processed. This processing lasts ~650 ms. This clarification has been added to the manuscript.

The location of the point measurements is only accurate to the temporal resolution of the GPS chip (1 s), so the data plotted in Figure 4 are only approximations of where the pixels actually are. In any case, Figure 4 shows that the interpolation and spatial correction techniques manage to adequately capture the variations in terrain reflectivity over the whole scene, which is why we opted to analyse the interpolated data only.

P 5685 Ls 15-25 Was the temporal correction performed with the interpolated data or with the raw data? Here it seems the interpolated data were used (compared to the previous comment on the camera). Is a constant shift assumed for the complete flight or are some variations allowed, caused by changes in the pitch angle, uncertainties of the time stamps of both image and position data?

The temporal correction was performed with the interpolated data, as it was impossible to resolve the ground features without it. Only data taken during level flights were corrected, as during banking manoeuvres the methodology used to obtain the correction term was invalid. Without accurate knowledge of the banking angle the data taken during such manoeuvres cannot be corrected.

P5687 L10 Plumes should not be included in the averages used for the destriping polynomial, therefore it's critical to include the measurements over Leicester city, however omitting the city data will reduce the statistical basis of the destriping.

As we have already mentioned in this section, the destriping procedure was performed using only data from Leicester city and the immediate surroundings. Plumes from heavily polluted areas such as EMA and the power station are not included. The results of this are shown in Figure 7.

P5701 Ls 8-16 I am not sure the authors are right with the assumption that the enhanced NO_2 columns observed in this region are caused by a small farm machinery factory. Can transport be excluded? I suggest to skip the last half of this paragraph L13 " A possible reason" . . . to the end. To me this seems to be highly speculative.

We agree that this is an unsubstantiated claim, and have removed this from the manuscript.

P 5704 L 1-3 Here the authors compare concentrations despite the facts that the background column is unknown and that the profile has a significant influence on the concentration, especially for elevated plumes like the one from the power station in Ratcliffe-on-Soar. (see P5698 Ls 1-4)

While we agree with the reviewer that such comparisons would be biased by reference sector and vertical profile influences, but we are not comparing absolute concentrations. The comparisons made in this work are relative to the reference sector column which we have called the "urban background". Similarly, the goal of the power station comparison was to show that the above urban background difference between two disparate emission sources could be resolved. We have clarified this in the manuscript.

P 5684 Equation 2: In theory the pitch angle of the plane should be included here, but this is unknown (failure of the IMU) and probably constant through out the flight, but a varying pitch angle might cause some small difference in the temporal synchronization. A pitch angle of 2 causes 30 m difference on the ground.

To compensate for the IMU failure we have since analysed the step change in GPS altitude per frame to determine possible changes in the pitch angle. For the main phase of the flight (N-S mapping of Leicester city centre) we observed that the actual measurement location was ±14 m from the raw GPS xy position used for processing and gridding the data. However, the frame rate of ANDI (0.3 s capture + 0.65 s processing) resulted in an average data gap of 90 m. Therefore, the captured frame would have always been positioned within the tolerance of the along-track resolution of the final dataset, due to the interpolation method used in this work.

Figures 4 and 5 can the authors use the same part of the city. Because in Figure 4 the intensity at the stadium (King Power) seems not to match, in Figure 5 (corrected) this can not be checked.

Figure 4 is an example of the interpolation process without temporal correction, while Figure 5 shows the joint effect of both processes. The spatial registration in Figure 5 shows a good agreement between the industrial units and intensity data over a larger, more inhomogeneous area than Figure 4. As such, we feel that such a comparison would not add much to the paper as it is, especially for a small, single site.

Figure 5: There are some weak stripes in the intensity map, which of course can later be found in the AMF and the VCD. Also the intensity data were destriped but it seems some stripes remained e.g north of the "G" of "Google". Does it matter whether the aircraft headed south (towards the sun) or north.

The heightened intensities appear to occur when the aircraft was flying south, towards the sun's position at the time of the flight. It is possible that the enhanced diffuse radiance caused by the high SZA would cause a constant increase in measured intensity during north-to-south swaths. Another possibility is that this effect could have been enhanced by an additional pitch angle introduced by the ANDI instrument being misaligned so that it did not have a totally nadir viewing geometry.

The striping behaviour appears to be repeated in the AMF data (Figure 10), though it does not appear as prevalent in the VCD data (Figure 12). The albedo data was derived from the intensity data, so it is expected that the AMFs would exhibit similar behaviour. Figure 12 suggests that the albedo derivation has helped to at least partially account for this effect, but in future flights a functioning IMU would help to properly account for this effect.

Figure 14 and 15: there are some single "scans" just downwind of the power station. What is the reason for these individual scans? It looks the instrument failed was restarted and failed again after one scan. The data look realistic or are they highly uncertain due to instrumental failure?

We believe the single scans are due to the instrument temporarily stopping and restarting due to an issue with the control software. The single scans are too far spaced apart to interpolate successfully, so we cannot accurately verify if the data is realistic. As a result, these single scans were not included in any analysis in this work.