We would like to thank referee #2 for his useful comments and detailed questions.

The authors describe measurements of NO2 above Bucharest by aircraft with an imaging UV/vis spectrometer. The paper focusses on the description of the retrieval chain and error assessment. The manuscript is suitable for publication in AMT after some minor corrections:

Major points:

- **Section 5:** It is not fully clear to me from the text whether only the measured intensity of the centre wavelength of the NO2 fit window was used or the whole spectral range. I assumed so since it is explicitly mentioned for the model intensities. However, the measurements are supposedly a vector. But it is not mentioned what the dimension is. In the first read I assumed it was the wavelength. Please add some more explanations in this section. Also, P. 16, L. 7: LUT of what? P. 19, L. 4: this is the first time that the text refers to the atmospheric correction explicitly.

The recorded spectra are measurement vectors containing the measured intensity (unit: Counts) as a function of wavelength. These spectra are analyzed using the DOAS technique to retrieve trace gas slant column densities. In the surface reflectance retrieval we do not use the full resolution spectra, but the average intensity in the fitting window, normalized to an illumination time of 1 second. The quantity ‘intensity’ used for the surface reflectance retrieval is thus a scalar (for each spectrum), with the unit Counts s⁻¹. In the table-lookup, the center wavelength of the fit window (plus all geometry variables) is used to obtain the modeled intensity from the LUT of modeled intensities.

The measured spectra contain information on the recorded light intensity during the exposure. The scalar quantity ‘intensity’ is computed for each spectrum, which is the average light intensity in the fitting window of the DOAS analysis, normalized to an illumination time of one second.

[...]

Legend:
- Referee comment (bold)
- Author’s answer (italics)
- Modified manuscript text (blue)
The influence of the observation geometry and the surface reflectance on the intensities measured by AirMAP were modeled for a given aerosol scenario (green line in Fig. 7) with the SCIATRAN RTM and compiled into a look-up-table (LUT) of modeled intensities at the aircraft. This LUT of modeled intensities is used to correct for atmospheric effects affecting the intensities measured by AirMAP. In order to isolate the contribution of the surface reflectance to the measured intensities, a 6 step procedure (a-f) is applied.

- **Section 4.1.2:** It’s not explicitly mentioned that stratospheric NO2 varies with the SZA due to photochemical reactions. For this the time step of the model is important. Please also add to Eq. (4) that the VCDs depend on the time. What happens when a model grid cell border falls within the study area?

  In order to obtain B3dCTM model values for an individual measurement, linear interpolation is used in all three dimensions (Latitude, Longitude, Time). The variation of the stratospheric NO2 is characterized by an almost linear increase of the NO2 VCD in the sunlit atmosphere. The temporal resolution of the B3dCTM model is 10 minutes. Due to the high temporal resolution, and the linear behavior of the stratospheric NO2 we think that the temporal evolution of the NO2 VCD is captured well. By using the linear interpolation in latitude and longitude, no abrupt changes are expected when a grid cell border intersects the study area.

  We have changed the text to include the information on the temporal resolution.

The B3dCTM model provides stratospheric VCDs of NO2 on a global grid in a spatial resolution of 2.5°x3.75° and a temporal resolution of 10 minutes.

We have added the time variable to Eq.4. The use of linear interpolation in space and time is now explicitly mentioned in the text.

A model value is obtained for each location (lat,lon) and time (t) of the measurements using linear interpolation in space and time.

**Technical and other small corrections:**

- **P. 1, L. 9:** I suggest rephrasing the part with ‘at the aircraft’.
  As the instrument is not radiometrically calibrated, we have developed a method to derive the surface reflectance from intensities measured by AirMAP.

- **P. 1, L. 10:** apparent instead of seen
  While surface properties are clearly apparent in the NO2 dSCD results, this effect is successfully corrected for in the VCD results.

- **P. 1, L. 16:** Please state coincidence criteria and sample size here.
  We understand that information about coincidence criteria and sample size is needed to interpret the comparison to the car measurements. Providing this information in the abstract may however be too much text for this section. This information is given later in the text (Section 9). Thus we prefer to keep that sentence unchanged.
• P. 2, L. 26-30: The most important property here is the trace gas species and not the location of the measurements. I suggest amending these sentences accordingly.

*That is a good comment in order to support the reader. These sentences now read:* 
In more recent years, imaging DOAS (iDOAS) instruments were developed. Lohberger et al. (2004) demonstrated the applicability for trace gas retrievals in a ground-based setup. Installed on aircrafts, these systems enable the creation of maps of the horizontal trace gas distribution. iDOAS measurements of anthropogenic point source emissions of NO₂ were performed by Heue et al. (2008) and by Schönhardt et al. (2015) above the South African Highveld plateau and above a German power plant, respectively. General et al. (2014) performed measurements of volcanic emissions of BrO, OCIO, and SO₂ at Mt. Etna, Italy. Urban NO₂ distributions were measured by Popp et al. (2012) above the city of Zürich, Switzerland, and by Lawrence et al. (2015) above the city of Leicester, England.

• P. 2, L. 35: Last sentence. Maybe add a sentence why that is better.

*Mapping of NO₂ distributions above cities with airborne iDOAS provides a holistic picture on pollution levels across the city and may be used identify the contribution of different NOₓ sources, such as industry and traffic, to pollution levels in a city.*

• P. 3, L. 21: Formatting problems for reference 

*Fixed.*

• P. 3, L. 25: ‘. . .overlap of the swaths between adjacent flight tracks. . .’ That’s difficult to understand without the instrument description first.

*In this part of the text we describe the research flights and the flight patterns. In this section, the reader shall be informed that the flights aimed at producing gapless maps. Further explanations about AirMAP’s viewing geometry follow in Section 3.1. A reference to Sect 3.1 is now added to the text.* 

*This pattern provides sufficient overlap of the swaths between adjacent flight tracks to produce a gap-less map (see also Sect. 3.1).*

• P. 3, L. 26-28: Please add the local time of flight since this is important in relation to the SZA and rush hour.

*The times of the flight are listed in Table 1. The footer of Table 1 the states that: “Local time (EEST) is UTC+3”.*

*The sentence describing the condition was now changed to include the information, that the flights were performed ‘around local noon’.* 

*Both flights were performed around local noon under cloud free and sunny conditions with low wind speeds*

• P. 4, L. 3: DOAS was introduced before.

*Fixed.*
The Airborne imaging Differential Optical Absorption Spectroscopy instrument for Measurements of Atmospheric Pollution (AirMAP) has been developed for the purpose of trace gas measurements and pollution mapping.

- **P. 5, L. 10:** ‘For data safety reasons.’ Is that due to memory effects of the CCD chip?  
  *No, this procedure is not related to a memory effect. The data on the CCD is stored in a circular buffer. In order to prevent possible data loss, for example from buffer overflow, the circular buffer is flushed every minute (60.7s). In this way, no more than one minute of data is lost by such a problem. This is what was meant by “data safety reasons”.*

- **P. 5, Table 2:** The FOV_along needs an aircraft speed to be meaningful.  
  *Table 2 now states the instantaneous field of view of the instrument. The instantaneous FOV is independent of the aircraft speed.*

- **P. 6, L. 1-7:** I feel this information should be provided earlier on in the description. Also, I would sort the 4 sentences in the order of 2,1,3,4. Also, do you know what happens at the edges of the pixels?  
  *In this section we first describe the instrumental setup and later the viewing geometry. It is a matter of personal taste which topic should be discussed first. The subsequent section describes the post-processing of the CCD images to individual viewing directions. Therefore we think that it is better to keep the two parts (viewing geometry & post-processing) close together.*

  *With increasing VZA the pixels become larger. This increase of the pixel width in across-track dimension can be up to 20% for the outermost viewing direction during a level flight. See also answers to Referee #1.*

- **P. 6, L. 7:** Remove ‘even’ and ‘individual’.
  *With the AirMAP setup it is thus possible to examine the sub-pixel variability within one OMI pixel (13×24km² at nadir), or a S5p-satellite pixel (3.5×7km² at nadir).*

- **P. 6, L. 11:** how many adjacent rows?  
  *The number of averaged adjacent rows is on average \( \frac{\text{SizeCCD}}{\text{#ViewingDirection}} = \frac{512}{35} = 14.6 \). Due to small variations in the cross-section of the individual light fibres and the cladding in the fibre bundle, the exact number of averaged rows depends on the extent of an individual light fibre imaged on the CCD and ranges from 13 to 15 rows.*

- **P. 6, L. 18:** ‘as is the case for ‘some’ satellite measurements’  
  *Using an extraterrestrial solar spectrum as background spectrum in the DOAS analysis, as is done in some satellite retrievals, yields slant column densities (SCDs), which are the number densities of an absorber, integrated along the light path.*

- **P. 6, L. 21:** This back-scattered radiance spectrum.  
  *This back-scattered radiance spectrum may contain small amounts of the absorber.*
• P. 6, L. 23: Replace ‘close’ with ‘slightly smaller or equal’. What are the expected background values in percent in comparison to the enhancements when polluted? → P.6 L.10
In our retrieval we assume a background NO$_2$ VCD of $1 \times 10^{15}$ molec/cm$^2$. Considering a moderate pollution level (of the presented flights) of $2 \times 10^{16}$ molec/cm$^2$, the NO$_2$ contained in the background spectrum has a share of 5% to the signal in polluted areas. Stating that the dSCD is equal to the SCD, thus does not seem appropriate. ‘Slightly smaller’ however may be better than ‘close’. The sentence now reads:
Because the background spectrum is taken at a clean location, the dSCD is only slightly smaller than the SCD.

• P. 6, L. 32: ‘. . ., whose pixel center.’ Please rephrase → P7.L3
The measurements are spatially binned by the pixel centers. All measurements with pixel-center coordinates falling into a grid cell are assigned to that grid cell. Multiple measurements in one grid cell are averaged using the unweighted arithmetic mean.

• P. 7, Table 3: ‘Fitted absorption cross-sections and ‘other’ important settings. . .’
Fitted absorption cross-sections and important other settings used in the retrieval of NO$_2$ dSCDs.

• P. 7, L. 2: ‘in aircraft turns’ → P7. L6
This effect can be observed in aircraft turns when the projected footprint becomes larger.

• P. 7, L. 5: add comma after second date.
Fixed.

• P. 7, L. 7: How many spectra do you add up for the background spectrum?
Table 2 lists that the exposure time is 0.5s. In table Table 3 we write that the background spectrum is a 60s average (also P 6 L. 7). Consequently 120 individual spectra are averaged for the background spectrum. This information was now included in the text in Sect. 3.2.
The background spectrum $I_0$ used in the DOAS analysis of the AirMAP spectra is an average over 60 seconds (120 individual spectra) taken from a scene of the same flight having low absorber abundances.

• P. 7, L. 8: ‘Some slightly negative values. . .’
Some slightly negative values occur in the background region as a result of instrumental noise, because the dSCD-values are scattered around zero.

• P. 7, L. 9: That sentence implies that those high values are caused by the pollution which is not the case. Maybe rephrase?
The maximal dSCD value is caused by high pollution levels in conjunction with bright surfaces, which enhance the NO$_2$ signal. The dSCDs are impacted by many factors, which are addressed in the derivation of the VCDs (Sect. 4).
A new sentence was added, which emphasizes the influence of bright surface on the measured dSCDs. See answer to next comment.

- P. 7, L. 11: with high values enhanced by up to 50% in comparison to the rest of the plume?
  A detailed examination of the derived surface reflectance in the region, highlighted in Figure 2, reveals surface reflectances of around 0.03 for non-built-up areas and surface reflectances in the range of 0.08-0.14 for the ‘bright’ pixels. Taking the data from Figure 5, an increase in surface reflectance from 0.03 (AMF=1.4) to 0.1(AMF=2.1) results in an AMF increased by 50%. This indicates that enhanced values of 50% in comparison to the rest of the plume can be explained by the surface reflectance.
  We have now added more explanations to the text.

  [...] Some lightly negative values occur in the background region as a result of instrumental noise, because the dSCD-values are scattered around zero. In the most polluted areas, dSCD-values of up to 6.1×10^{16} molecm^{-2} are observed. The dSCDs show a plume of NO\(_2\) spreading south-westwards from the city center. The NO\(_2\) field inside the plume shows small-scale structures with high values. Some of these structures, e.g. the pronounced values at the ring road in the South-West, are not associated with NO\(_2\) emissions from traffic, but are related to bright surfaces. These bright surfaces can enhance the NO\(_2\) dSCDs by about 50% compared to neighboring pixels, having a darker surface, see also Fig. 5 in Sect. 4.2.

  [...] The calculation of the AMF for the background spectrum is performed in the same way as for all other measurements. This means that all parameters listed in Table 4 of the paper are accounted for. Besides the flight altitudes and the angles, this includes the aerosol and NO\(_2\)-profile as well as the scene-specific surface reflectance. The text now includes more information.

  [...] The background spectrum is an averaged spectrum from 120 individual spectra, which may have different AMFs caused by changing conditions (geometry, surface reflectance) between these measurements. For each individual measurement during the integration time of the background spectrum, the AMF is computed for the respective condition during a single exposure, as will be
shown in Sect. 4.2. The AMF of each single spectrum is multiplied with the VCD\textsubscript{trop}, see Eq. 2. The average of the product is used as the tropospheric part of the reference background (SCD\textsubscript{trop}).

- P. 11, L. 5-6: Only for the respective observation geometry? What about aerosols and albedo here?  
  *See previous answer.*

- P. 11, L. 6: Refer back to equation (2) here.  
  *Done.*

- P. 11, L. 16: Please add reference to support this statement.  
  *The bias between SCIAMACHY measurements the stratospheric model, OSLO CTM2, was investigated in (Hilboll et al., 2013). The method presented there was applied to investigate the performance of the B3dCTM model. Results can be found in (Hilboll, 2014). The text now includes these references.*

- P. 11, L. 24: Formatting issue for reference  
  *Fixed.*

- P. 12, Table 4: 500m layer for NO\textsubscript{2}: Is that a valid guess? Where does this number come from?  
  *We agree that the justification of the assumed NO\textsubscript{2} profile was too short. We have now provided more explanations to justify the assumed NO\textsubscript{2} profile.*

  No information about the NO\textsubscript{2} vertical distribution is available for the conditions of the flights. Thus assumptions have to be made. In order to keep the definition of the unknown profile simple, we have chosen to use a box profile with a homogenous mixing ratio. The altitude in which the NO\textsubscript{2} resides depends on many parameters, such as emission altitude, boundary layer height, orography, temperature etc. Vlemmix et al., 2015 derived NO\textsubscript{2} profiles from MAX-DOAS measurements in China, showing that the NO\textsubscript{2} profile height is between 500 m and 1000 m. Mendolia et al., 2013 studied the urban NO\textsubscript{2} profile of Toronto and found the average characteristic profile height to be around 500 m during summer. These studies suggest that the assumption of a 500m NO\textsubscript{2} layer is a reasonable guess.

- P. 12, L. 8: I would suggest moving this sentence 4 sentences down. Figure 4 only supports the statement of the first half of the previous sentence. Hence it is a bit misleading here.  
  *We prefer to keep that sentence where it is.*

- P. 12, L. 8: BAMF: either explain or give reference here.  
  *Added citation.*  
  For a detailed explanation of the BAMF concept see Rozanov and Rozanov (2010) and references therein.
• P. 14, Figure 6: I suggest updating this figure if the authors feel artsy: The way it is sketched now implies that the instrument on the plane is scanning across track.

We have updated the figure.

• P. 14, L. 2: RAA is already described on the page before.

Thank you for this comment. The RAA was not the only angle explained a second time... The changed text segment now reads:

The observation geometry relevant for the RTM calculations is described by three angles: Solar Zenith Angle (SZA), Viewing Zenith Angle (VZA) and Relative Azimuth Angle (RAA). Figure 6 illustrates the meaning of these angles. The VZA is the deviation from the direct nadir observation geometry. As the VZA increases, the light paths get longer. The VZA changes with the viewing direction, but is also altered with the aircraft’s attitude. The RAA is the difference between the Solar Azimuth Angle (SAA) and VAA of the measurement. Following the SCIAMAT convention, the RAA is defined as 0° if the instrument is pointed towards the sun (forward scattering) and 180° for the direction away from the sun (backward scattering). The SZA is the angle between the zenith and the center of the sun’s disc and impacts on the length of the light path through the Earth’s atmosphere.

• P. 14, L. 9: 500m: see comment above.

See answer above.

• P. 15, L. 4-5: Why are there no measurements available close to the ground?

This is a general problem of the 2 types of measurements.

The Raman-Lidar measurements do not provide measurements in the lowest few hundred meters. This is mainly caused by geometrical reasons, because the telescope cannot capture the entire backscattered signal from close distances.

The FUBISS-ASA2 airborne sun-photometer has to point directly towards the sun. Vertical aerosol information can only be derived during vertical motion of the aircraft. If the instrument cannot be pointed towards the sun during ascent / descent, no information can be retrieved. The instrument is pointing always to the right with respect to the flight direction. The flight track during ascend and descend follows the regulations of the airspace and cannot be chosen by the instrument operator. Thus it is often not possible to point the instrument towards the sun, especially during take-off and landing.

• P. 15, L. 6: Why was the profile with the lowest AOD chosen?

Climatological data of the AOD in Bucharest is available on the AERONET website (https://aeronet.gsfc.nasa.gov/new_web/V2/climo_new/Bucharest_Inoe_500.html#2014). For September 2014, the monthly average of the AOD at 450nm is 0.26 +0.1 (mean+-std). The used profile 31a (AOD=0.22) is the closest match to the climatological value and is within the standard deviation of the monthly average. We thus assume that the used profile can be regarded as a typical profile for this location and season. We have now added this information in the text.
For the analysis of both flights in Sect. 7, we have used the profile FUBISS 31a. The AOD of this profile is the closest match to the monthly average AOD, available from AERONET measurements, having a value of AOD$_{450} = 0.26\pm0.1$, (mean±stdev) (https://aeronet.gsfc.nasa.gov/new_web/V2/climo_new/Bucharest_Inoe_500.html#2014).

- **P. 16, L. 15:** Remove ‘unfortunately’.
  
  Done.

- **P. 16, L. 16:** Remove 2nd comma.
  
  Done.

- **P. 16, L. 29-30:** But the surface elevation surely was not constant?
  That is true. The surface elevation was not constant. However, the area in and around Bucharest has a flat topography. At a constant flight altitude, modulations in the altitude a.g.l are only caused by buildings. In general the buildings have a height of about 10m or 20m. These modulations are rather small in relation to the flight altitude of 3360m a.g.l. Thus changes in the measured intensities, caused by different altitudes can be neglected here.

- **P. 17, figure 8:** What is cts?
  ‘Cts’ was an abbreviation for ‘counts’. We have now changed the labels of Figs. 8, 11, 12 and write the full word without abbreviations.

- **P. 25, L. 13:** Remove ‘very’.
  Done.

- **P. 26, L. 22:** And what is the error on the ADAM reference surface reflectance?
  The surface reflectance database also provides uncertainties associated with each grid cell, which is based on the variability of the MODIS measurements. For both of the surface reflectance reference grid cells a variance of $1.83 \times 10^{-5}$ is given. Thus, the uncertainty (1sigma) of the surface reflectance value is 0.0043. With a surface reflectance of 0.0394, this results in a relative uncertainty of 11%. The uncertainty is now stated in the text.

  Two grid cells of the ADAM database were used as the reference region, both having a surface reflectance value $A_{\text{Ref}}$ of 0.0394±0.0043.

- **P. 28, L. 3:** remove comma after larger.
  Done.

- **P. 28, L. 6, 8:** ‘antagonistic’ is poor choice. Maybe ‘opposing’.
  Changed antagonistic to opposing
• P. 28, L. 5: Remove 2nd ‘in a’.
  
  Done.

• P. 28, L. 10: Remove ‘a’.
  
  Done.

• P. 29, L. 10: do you mention somewhere what the absolute stratospheric column is? Maybe repeat it here.
  
  We have now added information about the absolute stratospheric column, as well as information on the impact of the stratospheric correction in Sect. 4.1.2 (Stratospheric correction)

  As the measurements shown in this study were performed around noon, the diurnal variations in SCD\text{strat} are very small and stratospheric correction is of minor importance.

  [...]

  For the two flights on 2014-09-08 and 2014-09-09 the stratospheric VCD was estimated to be around $3 \times 10^{15}$ molec cm$^{-2}$. The maximum change in the stratospheric SCD with respect to the reference spectrum, $\Delta \text{SCD}_{\text{strat}}$, was $1.5 \times 10^{14}$ molec cm$^{-2}$ and $3 \times 10^{14}$ molec cm$^{-2}$, respectively.

• P. 31, L. 1: ‘scatter is smaller’.
  
  When comparing to the UGAL instrument the scatter is smaller, presumably because the spatial inhomogeneities in the NO$_2$ field affect the comparison to a much smaller extent because the instrument is pointed to the zenith.

• P. 31, L. 9: Does that time lag correspond to the speed of the car?
  
  The speed of the car surely impacts on the ‘time lag’, but is not the only reason. The ‘time lag’ discussed is caused by errors in the geolocation of the compared NO$_2$ measurements. We have used the car’s location as co-location criterion which is a point. However, the measurements were not only performed at that location but along the route of the car (line) during the integration time. If the NO$_2$ field is not homogeneous, this results in spatial misregistrations. Furthermore, the MPIC measurements were performed under an elevation angle of 22°. In the manuscript we have assessed the potential spatial mismatch between the car’s position and the location of the measured NO$_2$ signal. The comparison between airborne measurements and off-axis car-DOAS measurements may be improved in future analyses, taking into account the azimuth as well and the elevation angle and the NO$_2$ profile. For such a correction of the geolocation, car-DOAS measurements with a higher temporal resolution would be desirable to avoid changes of the azimuth during the integration time.

• P. 32, figure 20: It’s a bit misleading in this case here that the top panel in Figure (b) doesn’t have markers on it since there is this large measurements gap. Maybe break the lines to show this. Why not make it panels A-C?
  
  We excluded the markers, because the plot became messy and overloaded. Breaking the lines is a good suggestion and we have changed the plot accordingly.
P. 33, L. 21: Polar coordinate system?
We have used a Cartesian coordinate system. First, the grid of NO₂ VCDs was converted to Cartesian coordinates. Then the sampling angles $\phi$, were defined from 0° to 360° in steps of 0.1°. In the next step, the sampling locations were obtained by conversion from polar to Cartesian coordinates, using the respective radius $R$.

$$x = \sin\left(\phi \cdot \frac{\Pi}{180}\right) \cdot R$$

$$y = \cos\left(\phi \cdot \frac{\Pi}{180}\right) \cdot R$$

P. 36, L. 14: What is APEX?
APEX is the “Advanced Prism EXperiment” and was already introduced on P.15 L.14 in the sentence:

To our knowledge, no data product provides information on the ground spectral surface reflectance in sufficient spatial resolution to be used for our measurements. Such data is only acquired on campaign basis from instruments such as APEX (Itten et al., 2008) or HySpex (Baumgartner et al., 2012).