We thank referee #3 for taking time to review our paper and for the helpful comments that have contributed to improving the paper. Detailed below are our answers.

The paper describes a novel aircraft-borne optical spectrometer and some atmospheric test measurements, including a validation of the inferred data with auxiliary measured data. The employed method builds on (a) the measurement of Limb and Nadir optical spectra, (b) the retrieval of the measured atmospheric parameters i.e. of UV/vis light absorbing trace gases and the aerosol extinction by means of Differential Optical Absorption Spectroscopy (DOAS), (c) forward radiative modeling of the Limb observations (whereas for the interpretation of the Nadir measurements a surrogate method is used), and (d) a mathematical inversion technique (is it linear?, see my comment I.2.b below) to infer vertical profiles of the targeted atmospheric parameters from the set of Limb observations. While the strength of the manuscript is with the description of them instrument and the validation of the inferred data by quasi-independent observations, some weakness comes with the description of the steps (c) and (d). In addition, the manuscript contains some statements, which are in contradiction to other statements made in the same manuscript and/or to the generally available wisdom in field. In a revised version of the manuscript, these deficits need to be removed.

I. Major comments:

1. Since all the essentials of the used technique and methods are not all new (even though the technique is renamed in the manuscript), it should be stated early in the manuscript (e.g. in the introduction), that the AMAX technique (or variants of it) primarily builds on the well-known Limb observation technique. For your information, the Limb technique is deeply rooted in astronomy. It started in 1907 when first limb observations were performed by the famous Catalan astronomer Josep Comas Solà. The observation led him to discover the atmosphere of Titan which is unfortunately not published in modern form. Later the technique was further developed by Milne, 1921, van de Hulst, 1950, Barth et al., 1969, Barth et al., 1971, and many others (see refs below). In the 1960's, Limb observations started in the terrestrial atmosphere from aboard the Nimbus satellites (e.g., Rodgers, 1976, Haas & Shapiro, 1985, and refs therein). Then in the early 1980's the Limb technique was employed for atmospheric observations from high flying research balloons (e.g. Water et al. 1981), and later (in the early 1990s) from research aircrafts (e.g., Traub et al., 1991). Since all these applications of the Limb technique are essentially containing the same elements as the method described in the manuscript, it appears worthwhile to trace your study back to this history (c.f. also by using a proper nomenclature). This would also provide to a wider readership a better orientation to understand the methods described manuscript.

We partially agree, and thank the reviewer for the historical background on limb observations. We have included a statement in the manuscript that mentions that the AMAX-DOAS technique *includes concepts that were first established in form of the well-known limb observations. However, viewing geometries other than limb, like NADIR, or off axis observations above and below the aircraft are other aspects of the AMAX-DOAS technique. For this reason we prefer not to extensively trace the history of limb observations at this point.*

2. Very disturbing are contradictory statements made in the manuscript, for example (a) the statement made on the DOF's and that 'the result is independent from the signal to noise' et cetera, made on page 7244, line 28 and again on page 7266 line 29. This statement is in utmost conflict with the presented equations c.f., following your equation 4, since your S(epsilon) could essentially be determined by the signal to noise, if all other measurement errors were negligible. Therefore, the solution vector (your x) is always determined by all factors going into S(epsilon), hence the measurement error as well. So at best you can state that the signal to noise of your (DOAS) measurement is not a major contributor to S(epsilon). Here, an informed reader would like to learn however, what factors of what magnitude are contributing to your S(epsilon). For example what are the uncertainties in F(x,b) (partly given in Table 5) or K, all related to your forward (RT) model? When discussing of your DOF and error budget you could also refer to Roscoe and Hill, (2002).

We agree with the reviewer and we have removed the statement regarding the signal to noise from the manuscript. We have added a section 'error analysis' to the manuscript, which gives the relevant details about the noise error, forward model error and smoothing error. In our inversion, the S(epsilon) matrix is built using the square of the slant column density fit error as the diagonal elements. The off diagonal elements are set to 0. We do not include the forward model parameter errors in the S(epsilon) matrix, because of the limited gain this would provide for a significantly increased computational effort to calculate the Jacobians with respect to the model variables. However, we did conduct sensitivity studies, that varied e.g., g over a range from 0.6 to 0.75, and take the error of the forward model parameters into account in our error bars, by calculating the S(total error) as the sum of noise error, forward model error and smoothing error. All error bars in the graphs of the manuscript reflect this total error.

The manuscript by Roscoe and Hill (2002) does not include any mentioning about DOF. However, we find it a useful reference in context of attainable vertical resolution, and have added it in our Section 3.5 where we discuss the vertical resolution of our limb measurements.

(b) The statements made in the manuscript regarding an 'iterative approach' are very disturbing (c.f., Page 7258, lines 8 and 24, and on page 7264 line 20) since the mathematics of your 'a posteriori' inversion scheme does not require iterations (see your equation 4). So what inversion technique did you really use? Did you apply a non-linear scheme in order to solve your equation 3, or did you just play around with parameters until you found the result 'convincing'? Explain.

The aerosol profile retrieval is a non-linear problem, and requires an iterative approach. On the other hand, the trace gas retrieval is a linear problem. In fact, there was a subscript missing in *our original manuscript equation (3), but as described in the following text, we do apply an 'onion peeling' type iterative approach to retrieve the aerosol profile. We have modified the equation (3), and the text in manuscript now connects with equation (3) to make this clearer.*

3. Radiative modeling: Even though treated to be independent, the SSA and the asymmetry parameter g are not total independent parameters. This can be seen from the definition of the SSA (k(scatter)/[k(scatter)+ k(abs)]), the definition of the (effective) free mean path length (i.e. the length until photon are directionally randomized) which is frequently implemented in RT codes i.e. k(scatter)-1 = lambda(eff) = lambda(Mie)/(1- g), where lambda(Mie) is the free path for Mie scattering, and g is the asymmetry parameter. How these different definition of k(scatter) =1/lambda(eff) (in SSA) or k(scatter) =1/lambda(Mie) are dealt with in the used RT needs be clarified before the firm conclusion regarding the uncertainty or (cross) sensitivity of the asymmetry parameter (e..g, on page 7259 line 12, and page 7268 line 28) and its independence on the SSA and inferred extinction can be made. In fact I recommend to parameterize the model with respect to the SSA for g = 0 and for the final discussion/inter- comparison rescale (see above) the inferred extinction with an actual or assumed typical g value for urban aerosols. Here for the aerosol retrieval it could also very helpful to increase the information content by not only using information from measured O4 slant column but also from relative radiances.

We agree that single scattering albedo (SSA) and asymmetry parameter, g are not independent parameters. We disagree that g=0 would be a reasonable assumption given the ample knowledge about inferences of g from sensors like AERONET in urban environments. Our value of g = 0.68±0.07. is based on the climatology of g-values in urban environments (Dubovik et al., 2001). While our sensitivity studies only assess the error for a particular parameter assuming all other parameters are known, the asymmetry parameter retrieved from AERONET measurement at Caltech, Pasadena CA has a 3 day average and standard deviation of 0.67 and 0.01 respectively and a 7 day average and standard deviation of 0.69 and 0.03 respectively. This is in excellent agreement with the climatology for other urban locations. With such a small range we don't think retrieving aerosol profile with g=0 adds much to the discussion.

The combined use of O_4 slant column and relative radiances for aerosol retrieval has been proposed before as a way to increase the information content but so far has not been implemented as it is computationally very intensive. We believe this topic deserves its own publication, and is beyond the scope of this methods paper.

4. The statements made on the T dependence and 'pressure effects?' of the O2-O2 absorption bands (page 7258, line 8 to 23 page 7263, line 10 to 25) are scientifically incorrect, hence not useful. First from O2-O2 collision experiments in the laboratory, the nature, structure, orientation dependent as well as the thermally averaged welldepth (De(O4) (= -(1130+80) J/Mol)) is well known (Aquilanti et al., 1999). Second, the nature of the O2-O2 absorption bands is well understood from theoretical studies (e.g. Biennier et al., 2000). Third, the weak T-dependence of the O2-O2 absorption bands found in the laboratory (e.g. Long and Ewing, 1973) as well as in

the atmosphere (e.g. Pfeilsticker et al., 2001) can concisely be reconciled to the nature of the O2-O2 interaction. So there is no need to speculate on (a) either pressure effects, (b) an alpha factor for which you fortunately infer alpha = 1, or (c) the T-dependence of the O2-O2 absorption. In fact I suppose that studies claiming an alpha =6 1 are subject to deficits in correctly dealing with the RT, for example neglecting the dependence of the SSA on g, and k(abs) et cetera.

We fully agree with the reviewer that alpha = 1 is the expected result. However, it is a reality for those who follow the MAX-DOAS literature that for reasons outside our control other studies apply alpha < 1 in order to scale O_4 measurements, and bring inferences of aerosol extinction into agreement with other sensors, like AERONET (e.g. Clemer et al., 2010). An increasing number of other studies have applied alpha < 1 (e.g. 0.75-0.83 Zieger et al. (2010), 0.89 Merlaud et al. (2011). In fact, we are only the second study that found alpha compatible with 1; the other study found alpha = 0.95-1.00 (see http://www.knmi.nl/omi/documents/presentations/2010/ostm15/OSTM15 AIS Spinei O2-O2 Cross Sections.pdf). Hence we consider it worth mentioning that our data shows that such a factor is not needed. We have modified the language to place our results into the current literature context.

We further agree for the most part with the summary about O_4 absorption bands. Indeed, we have conducted a very detailed investigation of O_4 absorption spectra of our own, using high signal-to-noise optical cavities that we develop in the Volkamer group that enable us to study the temperature dependence of O_4 absorption bands under atmospheric pressure. Our results reproduce the measurements by Long and Ewing 1972, though with lower error bars, and over a broader spectral range. We further reproduce the temperature dependence of the peak absorption cross sections by Pfeilsticker et al. 2001. However, we find that the integral absorption cross sections of O_4 is constant at all wavelengths (340-630nm), in particular it does not vary with temperature between 200K and 298K. The increase in the peak absorption cross section is found to be due to a change in the band shape with temperature. This is relevant, because it assigns O_4 as collision induced absorption (CIA) in the atmosphere. The study by Aquilanti et al., is relevant for atmospheric O_4 only to the degree that the low bond energy shows directly that bound O_4 molecules can only exist in molecular beams (such as used by Aquilanti), and at temperatures much lower than those found in the atmosphere. At atmospheric temperatures the kinetic energy of colliding O_2 molecules prevents the formation of bound complexes. As such, no information can be inferred about the bond strength of O_4 from atmospheric observations. The noted pressure effects refer to differences in the band shapes that we observe in comparing our atmospheric pressure spectra to those of Greenblatt et al. 1990, where pressure broadening contributes to the apparent band shape. As such there is nothing 'scientifically wrong' with our statements. We have clarified language, and have added a reference to the 'manuscript in preparation' to support the statement about 'pressure effect'. We do not consider this manuscript to be the place to go into further detail regarding the CIA nature of O_4 under atmospheric conditions.

Reference:

Thalman and Volkamer, Temperature Dependent Absorption Cross-Sections of O_2-O_2 collision pairs between 340 and 630nm at atmospheric pressure, 2013, manuscript in preparation.

II. Other comments:

(a) I strongly recommend to inter-change section 3.3. with 3.4 and 3.5, because the way you deal with your Nadir observations is a 'poor man' (or approximate) version of the inversion methods introduced in section 3.4 and 3.5. Also, I recommend to strictly separate the Nadir and Limb observations (in the nomenclature as well as by sections) in order to make clear what result is obtained from what observation geometry.

We have explicitly separated the sections into nadir and limb observations but kept the section numbers as it is. We also think that having geometric approximation before introducing inversion methods minimizes any ambiguity if present.

(b) The statements made anywhere in the manuscript on advantage of an EA angle control need to be fine-tuned. First, an active control of EA is certainly an advantage in maximizing the DOF of the measurements, if EAs are carefully chosen in order to maximize the DOF. Here a sensitivity study could help a lot to demonstrate what set of EAs is most relevant to increase the DOF in aircraft-borne Limb observations. Further, as far as I know, predecessor instruments had always means to learn the actual EAs, although most of them were not actively controlling the EAs. So the statements regarding the active control of the EAs made throughout the manuscript need to be accordingly fined-tuned. Moreover, the averaging kernels shown in Figure 11 do not really indicate that for these measurements the instrument was actually scanning through a series of EAs. Rather the AK pointed that the observations were made for a more or less stable EA (= 0? degree) during aircraft ascents and descents which is nice, but also not new. Also I found it rather disturbing how uncertainties in the EAs are discussed, in particular the coarse delta(EA) resolution (x-axis) shown in Figure 3. Here a higher resolution version would be helpful, and additionally a more concise explanation needs to be given in section 2.5 why the actual pointing error is smaller than expected based on the (Gaussian) sum of the individual errors.

We don't think there are any arguments regarding the advantages of having the ability to control EAs. Regarding the 'need to fine tune' language, we do think that our statements about previous AMAX-DOAS instruments on page 7246, lines 11-16 'pitch and roll is used during post-processing [of data] ... 'accurately reflects what is being done. As such we are not sure what the reviewer's point is here.

As the reviewer mentioned, active control of EA along with careful selection of EA allows for maximization of the DOF. This is only possible with the ability to actively control EA during the flight and not with ability to correct after the flight. One could even totally miss an elevated pollution layer during unexpected attitude change of balloon or aircraft. The extent of aircraft pitch and roll effect on lines of sight can be seen in Fig 8 (Merlaud et al., 2011), the LOS deviates by as much as 20-30°, leading to a significant loss of sensitivity and vertical resolution.

So we believe our statements regarding advantages of motion compensation system are justified. To our knowledge there is no other AMAX-DOAS instrument with motion stabilization.

We agree with the reviewer that the averaging kernels appear as if there were no other EAs but EA 0 used. This highlights the fact that EA 0 is the most sensitive EA during ascent and descent of the aircraft and provides the most independent information. This has also been stated by Merlaud et al. (2011). Bruns et al. (2004) performed theoretical sensitivity study regarding choice of EA for maximizing DOF while flying at a constant altitude. Based on the study by Bruns et al. (2004) and our experiment, we recommend maintaining EA 0 during descent/ascent and scanning only at the lowest point during the ascent and descent. We have included this in the manuscript as recommendation for future airborne missions and have also implemented it in our recent mission.

The resolution of delta EA was chosen based on the resolution of the motor internal encoder, which is ~0.2˚. The actual microstep drive of the motor is more precise (~0.01˚) than the encoder position read back. The width of the histograms is determined by the encoder readback. Choosing a finer grid would be rather arbitrary, and not add any information, as has been included in the manuscript.

III. Minor comments:

Here I summarize typos, improvements to the English, missing information, references and other stuff.

- Page 7246, line 15: âAŽthe first true et cetera‘ see my comment above ˘

It is not clear what the reviewer is referring to here.

- Section 2.3: Please provide information somewhere on the dispersion pixel/FWHM, since then the magnitude of over, or under-sampling for the detection of each gas can be assessed.

FWHM values in nm are mentioned in the text in Section 2.3. The FWHM are 7.7 pixels and 6.7 pixels for O_4 and TG spectrometers respectively. We have added this information to the manuscript.

- Page 7250,line 28: clinometer –> inclinometer

Clinometer has been changed to inclinometer in the manuscript.

- Page 7251, line 2: It not clear whether systematic errors should and can be Gaussian added. Explain.

We do not think the motor resolution and MMQ angle uncertainty are systematic errors.

- Section 2.6. Please provide information on the change in pressure within the cabin and how the change in p and hence change the refractive index within the spectrometers affect your wavelength recording.

The Twin Otter aircraft is an unpressurized aircraft and the pressure in the cabin changed from 760 -500 torr for the flight shown in Fig7. The change in the refractive index, as well as any pressure differentials that could give rise to a wavelength shift equally affect the Fraunhofer lines, and Earth atmospheric absorbers, and are being accounted for by linear shifting of the measurement spectrum during analysis. The observed shifts are up to 0.18 nm for the TG spectrometer between the spectra taken at the highest and the lowest altitude.

- Page 7253, line2: It is great to have a separate O3 monitor aboard, but why then the O3 data are not used to be compared with retrieved d O3 profiles?

We did not retrieve O_3 profiles using AMAX-DOAS. O_3 was measured by a differential absorption lidar. It is very difficult to retrieve O_3 as the signal is dominated by the stratospheric O_3 and hence did not retrieve O_3.

- Page 7255, line 21: McArtim may simulate much more than stated here. In fact in the math following your equation (4), you use c.f., computed Jacobians from McArtim.

We have modified the sentence in the manuscript and now reads "McArtim has the capability to simulate Jacobians of trace gases and aerosols needed for the interpretation of AMAX-DOAS data".

- Page 7259, line 19: Delete this sentence because it is not well based on facts (see e.g. Kritten et al, 2010, Merlaud et al., …)

We have modified to sentence in the manuscript to reflect the fact that angle accuracy is often ignored. It now reads; "Angle accuracy uncertainties is often not considered for error estimates for vertical profiles from airborne DOAS measurements, but it could be the most important and largest source of error in the retrieved profiles, especially for transition layers".

- Section 4.2: Validation of NO2 vertical column –> Validation of the Nadir measured NO2 vertical column… since actually you do not infer the column by integration of an inferred NO2 profile (see my comment above II.a)

We have divided the results into nadir and limb measurements. Validation of NO_2 vertical column is now as subsection of nadir measurements and hence we have left subsection heading as it is. The new section number reads as follows:

4. Results

4.1 Nadir Observations

4.1.1 Horizontal distribution

4.1.2 Validation of NO_2 vertical column

4.2 Limb Observations

4.2.1 Determination of O_4 SCD in the reference spectrum

4.2.2 Aerosol extinction profiles

4.2.3 Trace gas vertical profiles

- Section 4.4.1, fourth line. … between the measured and modeled O4 SCD. A better wording were … between the measured and best-guessed (see above) or inferred ..

We do not agree that best guessed is the better wording. We leave it as it is.

- Page 7264, line 26: Error contribution in retrieved extinction due EA uncertainty –>The error contribution in retrieved extinction due to EA uncertainty… and then … It illustrates

We have made changes in the manuscript as suggested by the reviewer.

- Page 7265, line 223. 1ppb _= 2.46×10^{10} molecules cm2 reconsider the correctness of the dimension cm2?

Changed to cm^{-3}.

- Page 766, line 1: …error is slightly smaller in the FT, where aerosol extinction presents less of a limitation,…. a limitation of what, and less with respect to what?

With both the reference spectrum and the measured spectrum taken under very similar conditions, i.e. FT with lower aerosol load compared to BL, the DOAS fitting error is smaller in the FT compared to BL. This has been added to the manuscript to make it clearer.

- Page 7266, line 29: Reconsider the statement with respect to the comment I.2.a.

We have removed the sentence from the manuscript.

- Page 7288, line 1: though the presence of elevated layer was observed as well –> though the presence of an elevated pollution layer was observed as well…or –> though the presence of elevated pollution layers was observed as well

We have changed it in the manuscript to "though the presence of an elevated pollution layer was observed as well".

- Figure caption 7: mention here … Nadir observations

We changed the figure caption and now it reads "NO_2 vertical columns derived from Nadir measurements".