Author comments to Referee #1 (RF1)

General comments:

The authors would like to thank RF1 for his/her extensive and very thorough review of our manuscript. Your suggestions were carefully considered and addressed.

Response to detailed comments:

1. Page 1, line 10-13: Please be more quantitative here and state approximate numbers.

   Fixed.

2. Page 1, line 19: Add ’e.g.’ before McFarquhar et al. (there are many more studies C2 who emphasize this aspect).

   Fixed.

3. Page 2, 1st paragraph: The discussion on the advantages or disadvantages of remote sensing, in-situ, ground-based or airborne should be more balanced. In-situ measurements e.g. have the advantage that they are more detailed with respect to microphysical and chemical aerosol properties while they are limited in space (point measurement). Remote-sensing can cover larger areas but their retrievals often depend on assumptions and give less microphysical detail. Airborne measurements are expensive and thus not feasible for monitoring, etc. ...

   The Introduction section was modified accordingly.

4. Page 2, line 20: Are 13 references for the retrieval techniques really needed here? The same reference chain appears on page 4 again. The authors should focus on the important publications. No discussion and references on previous validation studies are given, which should be added here.

   References were revised and repetition of references was removed. We added references of other validation studies as suggested.

5. Page 2, line 25: Most in-situ and all monitoring measurements are usually performed at dry conditions to keep them comparable. Please add this info here.

   This is mentioned a couple of paragraphs later, therefore the authors felt it would be repetitive to add this information here.

6. Page 2, line 34: DISCOVER-AQ is not defined yet.

   We moved the pertinent paragraph to the next section.

7. Page 3, line 3: The Zieger et al., 2010 reference is incorrect here. The lidarin-situ comparison (using humidified nephelometer measurements) were done in Zieger et al. (2011) and in Zieger et al. (2012).

   The references were corrected.

8. Page 3, line 13-14: Why was no hygroscopic adjustment necessary?

According to the authors of Mueller et al, 2014, it was thought that the relative humidity during the cases presented in their study was low enough and therefore hygroscopic adjustments could be neglected. In our current study, the need for hygroscopicity adjustments is emphasized.

9. Page 3, Sect. 2: I would suggest to replace the title of this section (”DISCOVERAQ”) by something more descriptive (e.g. ’Campaign description’ or something similar). The information on the number of profiles analysed (line 18) could also be moved to the campaign description section.

Section title was changed and recommended changes were applied.

10. Page 4, Sect. 3: The section heading could also be improved here. Instead of ’LaRC HSRL-2’ the authors could use ’Airborne high spectral resolution lidar’ or just ’Airborne lidar measurements’. Parts of the second paragraph of Sect. 3 are repetitive from the introduction (remove it here or there). Section 3.1 is not followed by a Sect. 3.2, so I would remove this subsection heading if nothing follows or restructure.

Section was renamed and subsections removed.

11. Page 4, line 27: Repetition and not needed here.

Repetition removed.

12. Page 5. line 15 and Table 1: This table is not really needed and could be omitted or moved to the supplement.

Table was moved to supplement.

13. Page 5, line 25: Does the 5 µm size cut relate to dry or to ambient RH? If the size cut relates to ambient conditions, then the effect of hygroscopic growth will influence the presented results since the actual size cut at elevated RH might much smaller. Please clarify.

The sentenced was changed to reflect that the cut-off size was related to dry conditions.

14. Page 5, line 28: How was the humidified nephelometer calibrated? How did the authors determine the exact RH of the wet scattering coefficient? Were salt calibrations preformed (see e.g. recommendations given in Zieger et al., 2013)?

The humidified nephelometer is calibrated in the same manner as the dry nephelometer by measuring the scattering coefficient of dry CO2 gas, whose scattering properties are known in the literature. This procedure is standard and is described in the TSI manual. The exact RH was determined from three different relative humidity probes located at the inlet, exhaust, and within the nephelometer flow cell. Ammonium sulfate was used to test the f(RH) for the nephelometers.

15. Page 5, line 16: As correctly stated, the γ-fit is one empirical fit among many. However, the real limitation is not the fit, it is rather the fact that the humidified scattering coefficients were only measured at one elevated RH due to experimental limitations (aircraft measurements). Therefore, it will remain unknown if phase transitions have occurred below 80% RH or not. In theory you can apply the γ-fit for different regimes of the humidogram (e.g. if hysteresis is present) separately (in Zieger et al., 2010, for example, the γ-fit was used to describe hysteresis effect due to deliquescent sea salt).

The original paragraph has been modified, and the experimental limitations with respect to RHwet measurements are discussed later in the manuscript, along with the discussion of chemical composition of the aerosols observed in Texas (see last paragraph of Section 6.4 of the revised manuscript).

16. Page 5, line 22: To back-up the negligence of the absorption enhancement, the authors should state the campaign mean and standard deviation of the single scattering albedo at this point.

Mean values, standard deviations, and 90^{th} percentiles of dry single scattering albedo measurements at 550 nm during DAQ CA and TX were stated.

17. Page 7, line 7: How were the multiple charged particles treated for the ammonium sulphate calibration of the LAS and UHSAS using the DMA?

Multiply charged particles transmitted through the DMA appear as distinct, smaller peaks in the optical size distributions of the LAS and UHSAS. These distinct peaks were ignored — only the singly-charged DMA peak was used.

18. Page 7, line 10-11: Please state the mean and standard deviation of the dry and wet relative humidities (preferable in the instrument section).

Mean and standard deviations of dry and wet RH were added to the in situ instruments section.

19. Page 7, Eq. 3: If g^{-} is not constant over the size range (as done later in the sensitivity study) then the change in number distribution has to be calculated as well. See Eq. 5 in Zieger et al. (2013).

All size distributions in the sensitivity study were normalized to 1 at dry conditions. Assuming the hygroscopic growth does not change the number of particles, the wet size distributions were renormalized to 1 once the growth factors were applied.

20. Page 7, line 22: Please add that an internal mixture was assumed (correct?).

Fixed.

21. Sect 5.2 and Fig. 2: Every inconsistency in the in-situ measurements (e.g. particle losses) will be balanced by the retrieved refractive index. The statement, at this point, on the consistency of the in-situ measurements can only be valid if the corresponding retrieved refractive index is shown as well. This profile should be added to Fig. 2. To save space the authors could consider to show only one wavelengths for the scattering coefficient.

Figure 2 was modified accordingly.

22. Fig. 3: The course of the hygroscopic growth factor is probably highly driven by the ambient relative humidity. For a better comparison, the RH profile should therefore be shown as well.

Figure 3 was modified accordingly.

23. Page 10, line 24-25 and Fig. 4: The authors state that the retrievals compare well to the in-situ measurements. How were these profiles chosen? Comparing Fig. 4 with Fig. 5 it seems to be that only nice examples were cherry-picked. Therefore this sentence should be rephrased.

The profiles were chosen qualitatively based on how complete the profiles were – both in terms of the in situ and the lidar retrievals. Due to the small number of points in any particular profile obtained in California (due to the shallow PBL), none of those cases were used as examples. The correspondence between Figure 4 and Figure 5 can be seen if looked carefully. Surface area and volume concentrations show better agreement while effective radii are slightly underestimated by the lidar retrievals, in comparison to the in situ measurements (adjusted to hygroscopicity). The sentence in question was not modified because the authors believe that it does not misrepresent the comparison.

24. Fig. 4: The good agreement is remarkable. While the in-situ measurements of surface, volume and effective radius show clearly the same profile shape, some exceptions can be observed in the HSRL-2 retrievals. For example, in the second profile at approx. 600 m altitude (also at 1700 m), the retrieval of effective radius, surface and volume are not in correspondence. Is this due to the fact that they are independently retrieved? A consistency check could be included in the analysis (i.e. volume and surface value should give the appropriate effective radius under the assumption of spherical particles).

The number, surface area, and volume concentrations are not independent retrievals. The effective radius is calculated from the surface area and volume concentrations, therefore, the correspondence is implied. In the examples provided by the reviewer it is not correct to state that the retrievals of effective radius, and surface area and volume concentrations are not in correspondence. At 600 m, for instance, the surface area retrieval is about 330 um^2/cm^3, the volume is about 11 um^3/cm^3, and therefore reff = 3V/S = 0.10 um (this data point in particular cannot be clearly visualized in Figure 4 but part of the error bar can be seen). At 1.7km the surface area retrieval is about 47um^2/cm^3 and the volume is about 3.9 um^3/cm^3 which results in reff = 0.25 um.

25. Page 10, line 27: Figure 5 is only described by one short sentence. Please be more detailed here. If the figure is not important then it should be removed.

Fixed.

26. Fig. 5: I find the systematic difference of the particle number concentration interesting. The HSRL-2 seems 'to see' more particles then the UHSAS. However, I would expect the UHSAS to be more sensitive to small particles which dominate the total number concentration. Is there any explanation for this? On the other hand, the surface and volume concentrations seem to agree well, while the effective radius is systematically larger for the in-situ measurements. This is surprising since the effective radius can be calculated from the surface and volume (reff = 3V /A; see e.g. Grainger et al., 1995) and therefore should agree well. Or is it differently defined/retrieved here? Please clarify.

The reviewer is correct that the lidar is less sensitive to small particles; the relative lack of sensitivity leads to inaccuracies in the retrieved number concentrations of very small particles. Since this is a retrieval, the inaccuracy could lead to either increased or decreased number in the less-sensitive size range. The definition of effective radius used in this study is the same as the one mentioned by the reviewer. The larger apparent disagreement in effective radius is not inconsistent. The surface and volume errors are correlated and the effective radius is a ratio. Also, the surface and volume profiles have more vertical variability. An error in S or V may look relatively small compared to the relatively more constant effective radius profile. You can see that they are consistent by looking carefully at errors in particular layers. The effective radius is systematically larger or smaller for large segments but not the entire profiles. In those regions, either S is systematically smaller or V is systematically larger (but these are less obvious).

27. Fig. 6: Similar to Fig. 5, I find it remarkable that the bias of the effective radius is positive with hygroscopic correction and negative without hygroscopic correction. Should it not be similar to the surface and volume concentration?

In fact, is the other way around, as can be seen in Figure 6. What we observed is that the bias in the effective radius was negative with hygroscopic correction, and positive without the hygroscopic correction. The bias for the effective radius is not necessarily similar to the bias in surface and volume concentration because of the effective growth factor adjustment.

28. Page 10: The third paragraph is repetitive.

Fixed.

29. Page 10, line 30 and abstract: I am astonished by the fact that the bias of the total aerosol volume concentration is smaller than the on of the surface concentration. The in-situ instruments have difficulties measuring large particles (as discussed later in the manuscript) which on the other hand determine the total aerosol volume. How can this be explained?

The coarse mode volume that is not measured by the in situ measurements are not included in the comparison because we only considered the fine mode retrievals. But the reviewer's intuition agrees with what has been observed in the evaluation of the retrieval algorithm using synthetic data (Table 2). It is possible that this discrepancy observed in our study is due to the uncertainties in the retrieval of g which have not been independently assessed. Figure 6 shows that without the hygroscopic correction, the biases are smaller for surface area than for volume concentrations.

30. Page 11, line 15-16: Why is this interesting? Any explanations?

It is interesting because later on, on page 13 (lines 13-16), when summarizing the results of the simulation described in Section 6.2.1, we also observe that one of the cutoff effects seems to be that the extinction is better reproduced than the backscatter coefficient. The manuscript will be restructured in order to make this connection clearer.

31. Page 11, line 19 and Fig. 8: Looking at the graphs (especially at the first panel) I would rather talk about 'good' or 'very good' agreement (and not excellent). How were these points averaged?

"Excellent" was replaced by "very good". As described in that same paragraph, we used the HSRL-2 measurements obtained within 2.5 km and 10 minutes from the AERONET measurements for the average.

32. Page 12, line 17-23 and Fig. 9: Although the median bias slightly decreases if the LAS measurements are used, the IQR increases for these cases. This should be added and discussed as well.

Figure 9 was outdated. The authors apologize for the confusion. The figure and the discussion have been updated.

33. Page 12, line 24-27: Particle losses was only one hypotheses among many in Zieger et al. (2011). In fact, this study also compared 3β+2α lidar measurements to in-situ recordings that were re-calculated to ambient conditions. The lidar agreed much better to the in-situ measurements than the MAX-DOAS (especially during nighttime, see Fig. 12 in Zieger et al., 2011). For the MAX-DOAS, it was hypothesized that the lowest and compared layer was overestimated due to lofted layers (e.g. caused by ammonium nitrate partitioning). In addition, the MAXDOAS retrieval could have been influenced by horizontal gradients in aerosol concentration. Nevertheless, in this work, the influence of coarse mode particles is definitely a hot candidate for the underestimation of the in-situ data. To further investigate this, the authors could, similarly to Zieger et al. (2015), compare their in-situ optical properties to the columnar measurements of AOD (Fig. 8). Zieger et al. (2015) also found a clear underestimation of the in-situ derived AOD and hypothesized as one possible reason that coarse particles were not sufficiently sampled (e.g. being lost in the canopy or within the inlet system) due to the pronounced wavelength-dependency. The calculated fine-mode fraction could be added to Fig. 10 which would be more convincing.

This paragraph was modified. We worked on the reviewer's suggestion of comparing integrated in situ measurements of extinction to AERONET AOT measurements and new figures were added to the manuscript accordingly.

34. No subsection is followed after 6.2.1. Therefore I would re-order and add an extra subsection or remove this heading.

Sub-subsection 6.2.1 was changed to subsection 6.3.

35. Sect. 6.2.1 and Fig. 10: The argumentation is very speculative. The CA dataset contains much less datapoints than the TX dataset so the statistics is different. Looking at Fig. 7 again, it is hard to see a clear and significant difference between the two datasets. I would suggest to move this figure to the supplement. Alternatively, the authors could further test their hypothesis in a more convincing way, e.g. by colour-coding the points in Fig. 7 by the fine mode fraction or by plotting absolute or relative differences of the retrievals vs. the fine mode fraction of the AOD.

The statistics presented in Figure 10 were calculated using AERONET data only, and therefore do not correspond to the same statistics presented in Figure 7. The AOD fine fraction box plot presented in Figure 10 corresponds to the order of 19,000 data points for California and 21,000 for Texas.

36. Page 13, third paragraph and Table 4: The choice and definition of the cut-off diameters is not clear to me. Sedimentation or diffusion losses should be low between 100 and 1000 nm, so I don't understand the choice of 0.7 and 0.4 µm. In addition, the hygroscopic growth factors were much larger during the campaign (up to 1.6 at elevated RH, see Fig. A1). Therefore, the particles (if the UHSAS sampled at dry conditions) where actually much larger at ambient conditions due to hygroscopic growth and the cut-off diameters should be set to values above 1 µm. The interpretation (see 4th paragraph) should then be adapted. I would interpret the sensitivity study that coarse mode particles above ∼ 1.2 µm are only really relevant for the β1064-measurement. Please clarify and adapt accordingly.

The simulations presented in Section 6.2.1 had the objective to show how much of each component of a 3+2 dataset would be reproduced at varying cut-off diameters. In the simulations we probe cutoff diameters from 11 nm to 21 um. The choice of reporting the reproduced fractions at cutoff diameters of 0.4, 0.7 and 1.2 um were chosen arbitrarily. As correctly stated, the hygroscopic factor values ranged from 1 to about 1.6. The value 1.2 for g was based on the median g values for both campaigns, which were both about 1.2.

37. Sect. 7 (Discussion) and Sect. 8 (Summary and conclusions): These sections are again very repetitive and often dissipate. Please focus and discuss the main findings. Both sections can be combined. The references to previous findings are missing and should be added to the discussion. The limitations of the lidar retrieval technique are not discussed or even mentioned at all in the conclusion part, which has to be added. It would be beneficial to the paper if the authors would add a short and precise outlook and recommendation part to their work.

The sections were modified.

38. Sect. 7.1: The phase transitions are important mainly for pure compounds. In the ambient atmosphere, clear and distinct phase transitions or hysteresis effects have been observed (using humidified nephelometers that look at the overall/integrated effect) mainly when sea salt was present (see e.g. Zieger et al., 2013, for an overview). Organic compounds and mixtures with other inorganic substances will most likely lead to a smooth hygroscopic behaviour without pronounced deliquescence. Figure 12 also shows a dominance of WSOC and NO3 and thus makes deliquescence quite unlikely. In addition, the particles in the ambient atmosphere will most likely be on the upper branch of the hysteresis curve if they have experienced an elevated RH before the time of measurement. For the ambient optical properties, which are studied here, the authors should look and discuss the related scattering enhancement factors, which is an integrated value while HTDMA's only look at distinct (and fine mode limited) dry sizes. There is no Sect. 7.2 following 7.1, so please restructure.

This discussion has been updated.

39. Page 26, Table 3: The linear regression and correlation coefficients should also be given for the comparison of the microphysical parameters from Fig. 5.

A table was added with the correlation coefficients and fit parameters for the microphysical properties in the Supplemental Material.

40. Page 26, Table 2: The choice of biases smaller than 50% seems quite arbitrary. How is this justified?

50% is a number that has been used as a worst case benchmark when testing the lidar retrieval algorithms. This threshold will most likely change in the near future to take into account the measurement requirements set by ACE (Aerosol-Cloud-Ecosystems).

41. Figure A1 and Sect. A1: This part is to reviewer's opinion quite important since it demonstrated the validity of the presented retrieval method for $\bar{g}$ and the good quality of the recorded in-situ data. It could be moved to the main part of the manuscript. However, it is unreasonable to show $\bar{g}$ vs. the ambient RH because the wet scattering coefficient was always measured at a constant RH (80-85 %). A comparison to κ is therefore not appropriate since the entire curvature of $\bar{g}$ is predetermined by the here used γ-parametrization. The authors should show a distribution plot of $\bar{g}$ at RH=80-85 % (or preferable at one fixed RH by recalculating the wet scattering coefficients to one fixed RH). These values can then be compared to literature values.

The discussion on growth factors have been modified, and the suggested distribution plot was added to the manuscript in the methodology section.

42. Sect. A2: This section is quite difficult to read and understand (again many paragraphs consisting of only one sentence). A flow chart could help here. It is not clear on why Fig. A2 has to be shown. It is probably sufficient to state that the simulations were done for similar conditions as the HSRL-2 measurements. Table A2 is hard to interpret as well. Why was the noise not added to the RH of the ambient and humidified nephelometer measurement? The same is true for the influence of the coarse mode, which might not have been sufficiently sampled. Both aspects will have a clear effect on the retrieval uncertainty (see Appendix A and Fig. A1 in Zieger et al., 2013). The sensitivity to the ambient RH is not discussed at all and should be added here. Were the ambient RH measurements of the two aircrafts compared?

Figure A2 was removed. In the "noise-added scenario" the 20% error in the ambient scattering coefficient is assumed to account for the errors of all variables used to calculate the scattering, i.e. gamma, dry scattering coefficient, and ambient RH. This simulation is a best case scenario assessment of how the in situ retrieval algorithm would perform if all in situ instruments were consistent among themselves in terms of cutoff diameter. A note has been added in the summary section of the revised version. It is unclear whether there were RH measurements onboard the B200. But even if there were, such comparison would not be relevant in our study since the B-200 was flown at a constant altitude (approx. 8.5 km, i.e. higher than the P3B) and the measurements at the B200 aircraft level are not really used in this study.

43. Fig. 10: How exactly was the scaling of the volume distribution to the aerosol layer height performed? And why were two different heights (1 km and 3 km) for the two campaigns chosen. Maybe it would have been easier to just normalize all volume size distributions to 1 and then calculate the average values.

AERONET's volume distribution retrieval is a column-integrated retrieval which is reported per unit area. The scaling was not strictly necessary in this case since we are not comparing AERONET retrievals to lidar retrievals. The purpose of the mean size distributions figure was to show qualitatively the proportion between the coarse and fine modes. In this case the scale factor was chosen as an average mixing layer height estimated from the lidar and in situ measurements. In California the PBL was very shallow and most aerosols were confined within the first kilometer. In Texas the PBL was higher and 3 km was used as the scaling factor.

Response to technical comments:

1. Page 5, line 5: Replace 'mum' by µm.

Fixed.

2. Throughout the manuscript: Please don't put the unit meter in italics.

Fixed.

3. Page 8, line 22: One web-link is probably sufficient here.

Fixed.

4. Fig. 2: Please use the introduced variables for the scattering and absorption coefficients (y-labels) and avoid abbreviations like Scat450 or Abs532. Instead of ’recalc’ it would be better to use an abbreviation like ’retr’ (retrieved).

Fixed.

5. Fig. 3: Please use the correct variables for the axis-labelling (see comment above). Units should be next to the numbers and not in the next row.

Fixed.

6. Fig. 2 and 3: To be consistent in the equations, the scattering efficiency should also depend on the dry or ambient particle diameter.

Fixed.

7. The section headings for 5.2 to 5.4, 6.2 are quite long and complicated. They can be shortened to be more concise (e.g. ’Retrieval of dry complex refractive index’ or ’Retrieval of the effective growth factor’ or ’Optical closure study’).

Fixed.

8. Page 13, line 2: Please replace here and throughout text (where possible) 3β+2α by ’extinction and backscatter coefficients’. This will improve the reading flow since the acronym is very specific for the lidar community and not known to the majority of the readers.

The authors explicitly describe the meaning of 3\beta + 2 \alpha in the introduction.

9. Please replace ’optical particle counters’ by ’optical particle size spectrometers’. UHSAS and LAS are not just counting particles like a CPC, they also size them.

Fixed.

10. Fig. 4: Please add that these profiles are for the fine mode fraction only.

Fixed.

11. Fig. 8: Units missing in the statistics text blocks.

Fixed.

12. Fig. 11: Define SZD at the beginning of the caption.

Fixed.

13. Page 17, line 9: κ can range according to Petters and Kreidenweis (2007) up to 1.3 (sea spray).

Fixed.

14. Page 17, line 29: The correct formula for ammonium sulphate is (NH4)2SO4

Fixed.

15. Sect. A1: Please harmonize the variable names (i.e. the refractive index is given in different ways).

Fixed.

16. Fig. 1: Replace ’King Air’ by ’B-200’ as shown in the figure (or vice versa).

Fixed.

17. AOT and AOD are not used in the same way throughout the manuscript. For example, in Fig. 8 it is AERONET AOT and in Fig. 10 it is AOD for the fine fraction. Please harmonize.

Fixed.