Response to Anonymous Referee #3 on the manuscript "Observations of the spectral dependence of particle depolarization ratio of aerosols using NASA Langley airborne High Spectral Resolution Lidar".

We're pleased that the reviewer seems to very interested in our instrument and put so much effort into carefully reading our manuscript. However, some of the suggestions seem to be driving towards corrections or improvements that are quite small compared to the estimated error bars, and would not affect our conclusions. We feel that this is a very carefully designed and well calibrated instrument which it would be difficult to improve significantly at this time. However, any instrument will have potential sources of systematic error, and it may be quite difficult to assess them, the more so for a well-calibrated instrument where the remaining sources are subtle. We reiterate that this paper is primarily about the measurement case studies and the primary purpose of the discussion of systematic error sources and estimation of systematic uncertainty is to demonstrate that the measurements are accurate enough to support the conclusions we draw about them in the manuscript. It may be possible to refine some of estimates of component systematic uncertainties further with some future work, but this is out of the scope of the paper. All that being said, we certainly appreciate that many of these comments have helped us improve the exactness of our written descriptions.

Responses to specific comments can be found below. Reviewer comments are in blue and author responses are in black (manuscript text in italics)

The paper is well written, well suited for ACP. The measurement cases are well described and put into relation with other measurements and model results. Although the title of the paper emphasizes the three reported measurement cases of the linear depolarization ratio of aerosols, the part describing the instrument, its errors and the error calculation seems to be the more important of this paper, because it will serve as the reference for future papers about the depolarisation measurements with this instrument. I propose to publish the paper under consideration of following remarks concerning the description of the system set-up and the error calculations.

Chap. 2 Instrument description and measurement methodology

Because measurements of the linear depolarization ratio with the HSRL-1 and HSRL-2 are directly compared, the set-up differences between both instruments, in case they exist, should be explicitly mentioned, which could be relevant for the measurement of the linear depolarization ratio.

We will add text similar to this in the revised manuscript: "*For measurements at 532 nm and 1064 nm, HSRL-2 is identical to HSRL-1. HSRL measurements of extinction and backscatter at 355 nm are made using an interferometer rather than an iodine filter. For 355 nm measurements of depolarization discussed here, the setup is very similar to the other channels; the small differences are explained in section 2a.*" Those differences between the 355 nm and 532 nm depolarization measurements are already discussed in the first version of the manuscript.

Page 24757 Line 2

The polarization axis of the outgoing light is matched to that of the receiver with an approach similar to that outlined by Alvarez et al. (2006) using seven fixed polarization angles between ±45° , using the half-wave calibration wave plates indicated in Fig. 1.

How accurate can the offset angle between the outgoing polarization and the receiver be determined? This should be determinable from the uncertainty of the fit of the Alvarez-calibration with the seven polarization angles. It is conjecturable that the offset angle changes during a flight and between different flights due to thermal and pressure influences e.g. on the birefringence of the exit window, wherefore I would not average between different calibrations, especially not for a conservative error estimate (further discussion about the systematic error below).

The fit of the Alverez calibration with seven polarization angles is typically excellent, with chi-squared values >99%. An example can be seen below. It is a non-linear fit and we have not explicitly calculated the errors in angle or the other parameters from this fit. We acknowledge that there is a potential for the offset angle to change during a flight or between flights, which is why we have specifically examined the change in offset angle during flight when two or more calibrations could be made during a flight. We do not average between calibrations, and only look at this change in order to assess the uncertainty. We conclude that the angle is good to within 0.4 degrees over the course of a flight, as stated in the text.

Page 24757 Line 5

Following the alignment, the gain ratio between the cross-polarized and co-polarized channels is routinely determined in flight by rotating the transmitted polarization 45° relative to the receiver, ...

How accurate can the 45° angle be adjusted with respect to the receiver?

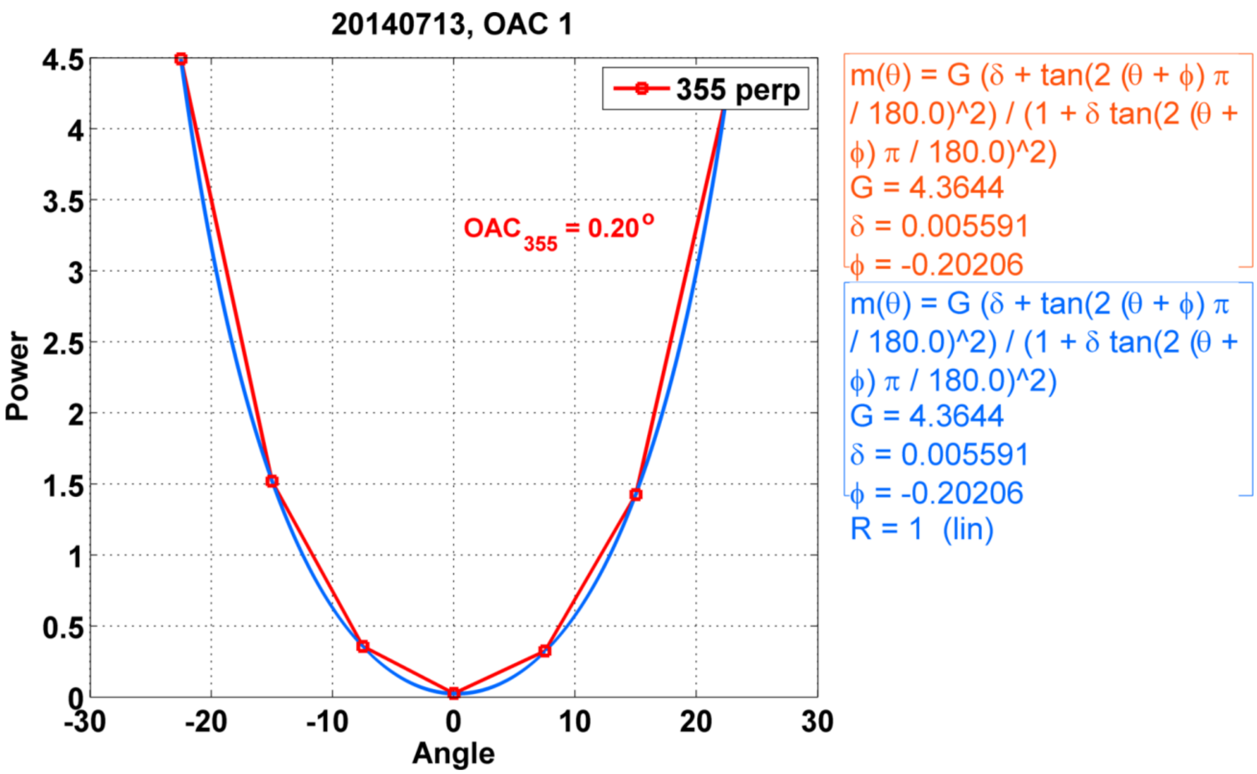
I guess that the precision is high by means of position encoders or similar. But what about the absolute accuracy? As shown by Freudenthaler et al. (2009), the high precision of the angular positioning can be used to achieve a high accuracy for the polarization calibration by means of the +-45° calibration regardless of the polarization offset angle. This could easily be done with the two of the seven calibration positions which are exactly 90° apart.

As an example, Fig.1 below shows the calibration factor (with Ic / Ip = I cross / I parallel) with assumed electronic gain ratio = 1, calculated with Eqs. (1), (A1), and (A2) of this paper for δ_{tot} = 0.1, a polarization ellipticity angle θ = 5°, and a polarization offset angle offset = -2°. The red marks show the seven measurements at nominal positions ψ as used by the authors of this paper, and the large error of the calibration factor at +45° or -45° positions. The green line shows the calibration factors calculated with the square root of the geometric mean of measurement pairs of the blue line which are 90° apart (e.g. -60° and +30°).

…

It would be very helpful for the readers if such a calibration measurement (with statistical and systematic error bars) could be shown in the paper.

Yes, the 7 angle fit does indeed give another estimate of the gain ratio which is another one of the parameters obtained in the non-linear fit. Here is an example of the 7-angle calibration from one of the flights highlighted in the paper. The red dots indicate the measured ratios and the blue curve indicates the fitted curve. The chi-squared goodness-of-fit statistic for this non-linear fit is typically >99%, computed by doing a regression on the relationship between the measured ratios and the fitted ratios.



The equation for the curve is Equation 10 from Alverez et al.

$$m_j = G\left\{\frac{(\delta + tan\text{^}2[2(\theta + \varphi_j)])}{1 + \delta tan\text{^}2[2(\theta + \varphi_j)]}\right\}$$

Where m_j are the measured ratios for each of the 7 angles, φ_j, and the parameters to be fit are θ, the offset angle; G, the gain ratio; and δ, the minimum depolarization value. This fitted gain ratio could reasonably be used as the inferred gain ratio without the need to do a subsequent gain ratio calibration, as the reviewer suggests. These two estimates are consistent (in this case agreeing to within less than 1%), supporting the use of the 45° calibration results.

Both calibration procedures (the 7 angle calibration and the single-angle 45° calibration) require the atmosphere to be stable, and the 45° calibration can of course be done more quickly. So, generally we are able to perform more 45° calibrations during the course of a flight. Since the system parameters may change during flight (due to environmental factors such as temperature changes), we use the polarization gain ratio calculated during the 45°

calibration. The reviewer has a good point that we may have sufficient time to do the calibration at both ±45° even if there is not time for a full 7 angle calibration, but historically we haven't done this.

The reviewer will notice that the values of the ratios at ±45° are not equal. However, please note that after the 7 angle calibration, the waveplate is rotated to the zero-point that is inferred from the center of the 7 angle fit. The 45° adjustment for the subsequent polarization gain ratio calibration is done with respect to this new, improved angle, and so should have a smaller error. If there was an error such that the adjustment of the wave plate angle is inaccurate or that it changes during flight, it would be evident when the next polarization angle calibration (i.e. the 7-angle calibration) is performed later in the flight. That's why we use the change during flight to estimate the uncertainties.

Page 24757 Line 28

The polarization extinction ratio measured in the system is 300 : 1 ….

The extinction ratio is the ratio of the transmission of the unwanted component to the wanted component. It is defined like that, e.g., by Tompkins and Irene (2005), by Bennett (2009) in OSA's Handbook of Optics, and Goldstein (2003) writes in Chap. 26.2.1: "The extinction ratio should be a small number and the transmittance ratio a large number; if this is not the case, the term at hand is being misused." Unfortunately, searching the literature, I find "misuse" by the larger part.

We have changed it to "transmittance ratio" in the revised manuscript.

Page 24757 Line 28

The co-polarized signal and cross-polarized signal are used to determine total depolarization.

Although for an insider it is clear from the instrument description what is meant with total depolarization, the correct naming for the measured quantities are linear depolarization ratio and volume (or total) linear depolarization ratio, etc.. This is important, because some lidar systems measure the circular depolarization ratio, and it should be at least mentioned once at the begin of the paper before proceeding with the short-cuts.

Agreed. We have inserted "linear" in the revised version of the manuscript in the title, abstract, and at several places in the text.

Page 24758 Line 12

The separation of the aerosol and molecular signals is the basis of the HSRL technique for extinction and backscatter retrieval. Since it is also relevant to the systematic error in particle depolarization ratio, it will be discussed again in Sect. 2.2, below.

The discussion of the errors of the backscatter ratio and its influence on the error of the linear depolarization ratio are described not sufficiently. Page 24758 Line 15 refers to Sect. 2.2, which refers to the appendix, but there just a value for the error is given and little explained. In Hair et al. (2008) Chap. 7 a detailed error analysis was promised in a following paper: *An analysis of the systematic errors for all data products from the airborne HSRL is beyond the scope of this paper. A manuscript focused on a complete error analysis and validation of extinction measurements is currently in preparation.*

I couldn't find this paper.

It's true there's just a number in Section 2.2; both references should direct the reader to the Appendix, where the number is explained. That has been changed in the revision. In the Appendix, there are already several paragraphs of discussion about the potential sources of error in the total aerosol scattering ratio (i.e. backscatter ratio) including how the calibration factor (aerosol-to-molecular gain ratio) is assessed, how we estimated the uncertainty in the aerosol-to-molecular gain ratio, and how this propagates to uncertainty in the particle depolarization ratio. This information is sufficient for quantifying the uncertainty in the 355 nm depolarization ratio. We're not sure what else is wanted.

Likewise, we're not sure what's considered to be missing from Hair et al. 2008. That paper has very detailed information about the HSRL-1 instrument, calibration procedures, and random and systematic errors. It has served well as a useful and sufficient instrument description paper. What it didn't contain was much validation, but HSRL-1 measurements were subsequently validated in a paper by Rogers et al. [2009] in this journal. The current manuscript is not meant to be a further follow-on to Hair et al. [2008] and is not primarily meant as an instrument description paper. We consider the descriptions of the instrument in this manuscript to be sufficient to support and explain the measurements we are highlighting.

Page 24759 Line 21

Different names are used for the same thing, e.g. volume depolarization, total depolarization, volume depolarization ratio, which is confusing. Please decide for only one short-cut (see comment above Page 24757 Line 28) throughout the paper.
Similar: there are several calibrations: polarization angle calibration, backscatter gain ratio calibration, depolarization gain ratio calibration, etc.. Please use unique names and only one for the same in the whole paper, and always use the full unique name.
Similar: fractional error = relative error?
We revised the manuscript to have only one term per concept and remove most synonyms and shortcuts. We are using "relative uncertainty", "volume depolarization ratio", "particle depolarization ratio", "polarization angle calibration", "polarization gain ratio calibration", "aerosol-to-molecular gain ratio", and "total aerosol scattering ratio".

Page 24760 Line 2

... we estimate a reasonable upper bound on the systematic error in the volume depolarization ratio measurement to be 4.7 % (fractional error) in the 355 nm channel, the larger of 5 % fractional error or 0.007 absolute error in the 532 nm channel, and the larger of 2.6 % fractional error or
0.007 absolute error in the 1064 nm channel.
Why is there no absolute error (offset) at 355 nm?
The absolute portion is for the uncertainty due to cross-talk that results in an offset of the observed clear air depolarization, reduced for a partial correction we applied (explained in the text and see also below). The observed offset for 355 nm is much smaller than for the other channels, so we initially didn't include it, but the reviewer is right: it is more consistent if we include it. We also realized the text was unclear about how the different channels are affected. Adding the component to 355 makes only a small change in the uncertainty, and only where the particle depolarization ratio is small. Only the lower layer (below the smoke layer) in Figure 14 is affected with now slightly larger error bars; none of the other figures or tables are affected. We will replace figure 14 in the revision. We also clarify the text like

this: "*Since 2006, we have historically measured minimum depolarization ratios in clear air that exceed the theoretical value, namely values of approximately 0.006 in the 355 nm channel, approximately 0.008 in the 1064 nm channel, and 0.0085-0.0135 in the 532 nm channel, …. An ellipticity angle of 5.8° (χ=0.980) would explain the error in the depolarization ratio at 532 nm where the error is largest…. Taking the partial correction into account, we include a component of 0.007 (absolute) due to cross-talk in the estimated volume depolarization ratio error for the 532 nm and 1064 nm channels and 0.001 (absolute) for the 355 nm channel.*"

Page 24760 Line 13

… the molecular depolarization arises only from the central Cabannes line and is very well characterized, with a value of 0.0036…

The molecular (air) linear depolarization ratio is wavelength dependent and actually 0.003946 at 355 nm, 0.003656 at 532 nm, and 0.003524 at 1064 nm. (Own calculations for air with 385 ppmv CO2 and 0% RH).

Thanks for the additional information. This is a very small difference that does not affect the results in the manuscript. We feel that we cannot quote these values in the manuscript because we don't know of an appropriate way to reference unpublished communication from an anonymous source. However, we changed the text to more precisely reflect the provenance and limitations of the calculation we use: "the molecular depolarization ratio is temperature independent and is calculated to be 0.0036 using N_2 and O_2 molecules (ignoring a negligible wavelength dependence due to non-linear molecules like CO_2) (Behrendt and Nakamura, 2002)".

Page 24760 Line 14

More critically important is ***a*** *potential systematic error in the total scattering* ***gain(?)*** *ratio. We estimate the effective upper bound of this error to be 4.1 % in the 532 nm channel from an analysis of the stability of the* ***gain (?)*** *ratio;…*

Stability (precision) is not accuracy. Furthermore, how is the error of the of the total scattering ratio determined?

The excerpt was correct as originally written, although we have clarified the revision by using the full names "total aerosol scattering ratio" and "aerosol-to-molecular gain ratio". What we are saying is that the uncertainty in the aerosol-to-molecular gain ratio gives the uncertainty in the total aerosol scattering ratio (the quantities are linearly related, so the relative uncertainties are the same). The aerosol-to-molecular gain ratio is calculated either by direct calibration (for 532 nm) or by transfer of the calibration from 532 nm to the other channels (discussed in the text). The calibration is performed one to three times during a flight, but not continuously. The difference between precision and accuracy for the 532 nm case would occur only if there was some mechanism that would cause the gain ratio to consistently be measured incorrectly during calibration. We know of no such mechanism; however, it is possible for the gains to change in flight, so the change in the measured gain ratio during a flight really is our best estimate of the uncertainty in the 532 nm aerosol-to-molecular gain ratio. The calibration transfer to the 355 nm and 1064 nm channels might conceivably result in a consistent bias, so we estimated the size of such a (potential) bias and used that to increase the calculated uncertainty for those channels, as discussed in the original manuscript.

Page 24760 Line 21

The estimates given above are intended to be a conservative upper bound on the systematic errors. The systematic errors on the three quantities, δ mol, δ tot, and R, are combined in quadrature using standard propagation of errors for independent variables, as described in the Appendix.

I do not agree, that this "standard" propagation of errors is the right one for the systematic errors mentioned here (see discussion Systematic errors below).

Response below.

Page 24762 Line 8

For that case, the particle depolarization ratios at 532 and 1064 nm are 0.33 �} 0.02 (standard deviation) ...

What does "standard deviation" mean here? Probably the propagated error due to (random) signal noise is meant (see discussion Systematic errors below).

This wasn't meant to be a systematic uncertainty. It's the standard deviation of a sample measurements immediately around the quoted measurement. It is described fully in the caption to Table 1 "samples were taken at specific times and altitudes comprising 400-4500 distinct measurement points ... the values are reported as median plus-or-minus standard deviation for the sample." Also in the text on page 24761 line 19 (of the discussion manuscript), it says "standard deviation for the sample". After that first usage, we shortened it for readability.

Figure 14

The x-scales could be adjusted for each wavelength to make the data better visible.

We feel that the full range is needed to capture the large values in the smoke plume at 355 nm.

Page 24776 Line 20

The calibration procedure has been carefully designed and used successfully on both the HSRL-1 and HSRL-2 systems since 2006, and the stability of the offset angle is high. Changes indicated during calibrations are at most 0.4◦ of polarization (0.2◦ rotation of the half-wave plate) for all channels (assessed, as before, using the mean plus two standard deviations for all flights having multiple calibrations during the latest field mission).

This tells us only something about the stability (or precision) of the 45° angle adjustment, but nothing about the accuracy, which is the basic important value.

See above for why we consider the stability of the calibration to be a good measure of its uncertainty.

What does "*Changes indicated* " mean?

If the Alverez calibration indicates that the center of the polarization curve shown in the Reviewer's Figure 1 is not at 0 degrees, the waveplate angle is adjusted so that it is at zero degrees. This has been clarified in the text.

Page 24777 Line 12

Change:

*This effect on the measured gain will be reflected in the ~~stability~~ **error** of the gain ratio, ...*

Changed as suggested.

Page 24777 Line 18

The stability of this gain ratio was assessed in a similar manner to the offset angle and polarization gain ratios given above.

Again: precision (stability) is not accuracy. Please explain.

The same explanation applies. The aerosol-to-molecular gain ratio calibration measurements occur at least once during flight and twice when possible (occasionally 3 or more times), but not continuously. There is some possibility of the gain ratio changing during a flight due to environmental factors (e.g. thermal changes). Therefore the amount of change during a flight is the best estimate of the size of the uncertainty. The uncertainty was estimated to be the amount of change in flights where two or more calibrations took place. For this study, we calculated statistics of this change over multiple flights to give a confidence limit for the uncertainty.

Systematic errors

A well-founded error calculation for lidar products is a really laborious task. The effort done in this paper is ambitious. Nevertheless, I would like to make a general remark and some comments in the following:

Error bars are essential in several respects, e.g. for the retrieval of micro-physical aerosol parameters with model calculations, for the comparison of results from different instruments, or for aerosol classification. The two scenarios A and B in Fig. 2 show their importance: in scenario A the two values 1 and 2 cannot be measurements of the same object, because the error bars don't overlap.

At least if we take the error bars seriously. In scenario B we cannot exclude that value 1 and 2 are measurements of the same object. They are not distinguishable considering the accuracy of the measurements.

…

Furthermore, if we know from other measurements that object 1 and 2 are actually the same, as it is sometimes the case from simultaneous measurements in multi-sensor field campaigns, we must conclude from scenario A that there are unaccounted instrumental errors, and from scenario B that the true value of the object is in the small overlap region of the two measurements.

This shows first, that error bars are very valuable and powerful information, and second, that we must be careful because other scientist will take our error bars and interpret them in their context if we don't specify them sufficiently.

We agree on the general discussion and philosophy of error bars. I don't think we said anything in the manuscript that goes counter to these general statements.

Models need the experimental error bars as constraints. They often produce results with statistical probabilities from many trials. Often Gaussian like distributions arise due to the central limit theorem, even if the original parameters are evenly distributed. But for that more than about ten different input values for each parameter are required.

But an instrument like a lidar system has only one set of system parameters at the time of a certain measurement, which usually shouldn’t change during the measurement. Therefore the application of the statistical error propagation for independent parameters (sum of squares), which assumes a Gaussian distribution of the "erroneous" parameters, is not appropriate for the error propagation of fixed systematic errors. Should a system parameter nevertheless change during a measurement, its behaviour should be determined and an appropriate error propagation developed. This would be the preferred method, but it is often too complex to accomplish. Also in this case an error calculation using the extreme bounds is the conservative way.

Here is where we disagree. First, it is not true that statistical error propagation for independent parameters assumes a Gaussian distribution. Statistical error propagation and summation in quadrature are only dependent on the definition of variance, and variance is a concept that is applicable to any distribution. The variance is defined as $E[X^2]-E[X]^2$ where E[] denotes the expectation value, and so this is not dependent on Gaussian distributions. The proof that variances add depends only on this definition (see the proof given here, for example: http://apcentral.collegeboard.com/apc/members/courses/teachers_corner/50250.html, accessed 11/3/2015).

Second, while we agree that it would be inappropriate to use statistical error propagation for repeated (fixed) systematic errors, that is not what we are doing. That is, to determine the error from repeated trials with a system having fixed systematic errors, you would not assess the error for each trial and then add them in quadrature or indeed add them at all. However, what we are doing in this section is different: assessing the overall effect of multiple independent sources of uncertainty (which happen to be systematic sources). Since they are independent and the value of the actual error for each source is unknown, it is appropriate to add them in quadrature. It would only be appropriate to add them absolutely if they were exact known values. That is, if we knew that the error due to the polarization gain ratio was exactly x and the error due to the aerosol-to-molecular gain ratio was exactly y, etc., then it would be appropriate to say the total error is x plus y. Although in that case, if so much was known exactly, we would probably choose to correct the errors rather than report them as error bars. In fact, we do not know the values: they are unknown values from an unknown distribution.

So, it's calculating the variance that's problematic, not adding them in quadrature. For random, normally distributed errors, the variance is well defined, and assessing multiple trials is usually the best and most straightforward way to calculate it. Granted, systematic errors are not best described as "random". For a constant systematic error from a single source, if the distribution is a delta function, it would not make much sense to talk about variance, as the reviewer points out. But in our case, even though it may not be right to describe the errors as "random", the errors are unknown so they each belong to some probability distribution which is not a delta function. Unfortunately the probability distribution is also unknown. So, we don't know a formula for how to calculate the variance. Instead we do the best we can to make an estimate of a confidence interval for these errors, such that we expect that the error from a given source is most likely less than the limit we specify (and which we therefore called a bound or limit in the discussion paper). We do not say that the error is exactly equal to that confidence limit. In fact, we do not know these values but they are unlikely to all be equal to the upper limit. So adding the errors absolutely would result in unnecessarily large error bars, which, as the reviewer points out below, are not particularly helpful to data users. On further reflection, we see that calling these confidence limits an "upper bound" may have added to confusion if the reviewer took that to mean all our uncertainty estimates denote absolute maximum. We really intend this to be analogous to 95% confidence interval (although we don't wish to take "95%" too literally, since the distributions are not Gaussian), a threshold that the systematic error is "most likely" to be below with high confidence. In the revision, we use "uncertainty", making a more correct distinction between the systematic error (the amount by which a measurement is incorrect due to system parameters) and "systematic uncertainty" (our best estimate of the value below which the error should fall) and we think this greatly clarifies the discussion. It's true that if we knew more about the probability density of the true errors within our uncertainty, it would be possible (though probably still challenging) to make the estimates of the variance and confidence interval more exact and the results might differ from what we reported. However, we still would need to add the errors from different sources, and since they are unknown values and independent, we believe that adding them in quadrature is the most reasonable approach.

Furthermore, if the lidar error bars are too large, a too large variety of model results fall within the error bars, and if the lidar error bars are too small, the model solutions which would come close to the reality might be excluded. If lidar error bars are getting smaller and reliable, the lidar measurements can be really helpful to improve the model developments.
We have no disagreement with this statement. We have made fair estimates of the systematic uncertainty of our measurement system and demonstrated that the error bars are small enough to support the conclusions we draw about the measurement cases.

Detailed comments:

1. The details of the error calculation should not fall back behind the one presented by Freudenthaler et al. (2009).
We are not sure what the reviewer means by this. While we do compare our measurements to those of Freudenthaler et al. (2009) for a dust case, to show that our results are consistent with earlier published results, there is nowhere in the manuscript that suggests that the Freudenthaler et al. (2009) error calculation should be used to understand our own measurements.

The equations for the F_x-values should be presented as well as the ones for the calculation of the error of the backscatter ratio due to the HSRL technique.
Equation A6 of the Appendix gives the F_x factors as partial derivatives calculated from Eq. (2). In the revision we will add a sentence clarifying this: "*The partial derivatives are calculated easily from Eq.* ***Error! Reference source not found.*** *which relates the particle depolarization ratio to the factors R,* δ_{tot}, *and* δ_m." We don't write out the partial derivatives, since they are straightforward for any reader to calculate.

As already mentioned, the "manuscript focused on a complete error analysis and validation of extinction measurements" promised by Hair et al. (2008) is missing.
This is a criticism of an earlier published paper, not the manuscript under review. However, we will point out again that Hair et al (2008) does indeed include a "complete error analysis" of the extinction— and depolarization— measurements of HSRL-1. The validation was demonstrated in a later paper, Rogers et al. (2009). The current manuscript does not present any extinction measurements, so the error analysis of HSRL-2 extinction measurements is outside the scope of this paper. To the extent that the extinction and backscatter retrieval is

relevant to the depolarization ratio uncertainties, we have discussed it already in the manuscript.

2. The absolute error (offset) of the volume linear depolarization ratio can only be positive. The only way to decrease the depolarization is a polarization filter, which is the case if the receiver optics has diattenuation. But this effect is in principle fully corrected with the polarization calibration. (I propose to use "polarization calibration" instead of "*de*polarization calibration".) Therefore, this error would have a one-sided distribution if many different instrument adjustments were done, which is clearly not a Gaussian distribution.

Our calculation does not depend on its being Gaussian. Please see above. We acknowledge that an offset error is one-sided; however, like the reviewer, we think it is reasonable to try to correct for it (as described in the text). This correction may overshoot in particular cases, so the error after correction is not necessarily one-sided. Furthermore, the other potential systematic error sources are not one-sided (errors in the gain ratios, for example), so adding them all absolutely would result in an unnecessary overestimate. We feel that what we have done is the most reasonable way to combine all the various sources of uncertainty, given that it is impossible to assess systematic error by repeated trials. (We agree about using "polarization calibration" and have revised the manuscript accordingly.)

3. Eq. (1) of this paper corrects only for different electronic gain and optical transmission after the polarizing beam splitter, but not for the cross talk of the polarizing beam splitter as shown by Freudenthaler et al. (2009) Eqs. (15) and (16). Although the extinction ratios of the polarizing beam splitter assemblies used in the HSRL-2 receiver are quite good, the error from neglecting their cross talk is maximal for low depolarization and amounts for the molecular linear depolarization ratios to +0.0023 at 355 nm and +0.0010 at 532 and 1064 nm using the transmission ratios in page 24757 line 29. The linear depolarization ratio values presented in the paper could be easily corrected for that effect.

However, this calculation also shows, that the effect is not sufficient to explain the assumed molecular linear depolarization ratios of 0.0085 to 0.0135 measured since 2006 (Page 24775 Line 14).

The molecular linear depolarization ratio is the only calibration standard we have for depolarization measurements. Deviations from that can be due to an offset, due to a calibration factor, and due to a combination of both. Assuming that the error of the calibration factor can be reduced to a few percent, the offset can be determined and all measurements can be corrected for that error with the appropriate equations. The remaining error is then the unexplained spread of the assumed-molecular linear depolarization ratios of 0.0085 to 0.0135.

As the reviewer points out, the correction for cross-talk in the polarization beam splitter is not sufficient to correct the total systematic error due to presumed cross-talk. So we have used the molecular linear depolarization ratio as the means of estimating this source of systematic uncertainty, as the reviewer suggests. Since, as the reviewer points out, the offset error will be in one direction and fairly consistent, we feel that we can correct for it, as already described in the Appendix. However, even this correction is not perfect, and so we include an offset portion in our reported uncertainty as well.

4. Elliptically polarized output light can be separated in the Stokes vector in a pure linearly polarized and a pure circularly polarized part. The circularly polarized part is detected by the

linear polarization analyser, i.e. the polarizing beam splitter in the lidar receiver, as depolarization and gives a more or less constant offset contribution to the linear depolarization ratio (decreasing slightly with increasing atmospheric depolarization). It doesn't influence the polarization calibration factor.

If by "polarization calibration factor" the reviewer means "polarization gain ratio" we agree that ellipticity does not affect that quantity. It does affect the measured depolarization ratio; the amount is given by Equation A3.

In contrast, if there is a rotation of the plane of polarization of the emitted light with respect to the receiver, it is probably also there for the polarization calibration, which results in a relative error of the gain and therefore in a relative error of the linear depolarization ratio (see above comment to Page 24757 Line 5). Therefore, the two systematic errors, i.e. elliptical polarization and angle of the plane of polarization, cannot be treated identically as cross-talk (Page 24776 Line 8).

We agree that there is also a relative uncertainty due to the effect of cross-talk on the gain ratio, which we discussed separately in the paper in the following paragraph. The cross-talk also has an offset effect, described by Equation A3. The offset portion does not depend on whether the problem is a rotation of the plane of polarization or an ellipticity. Although this was already discussed in the manuscript, we have attempted to clarify this by changing the wording. We now say, "*Eqs* ***Error! Reference source not found****. and* ***Error! Reference source not found****. make no distinction between the ellipticity and polarization offset angles* θ *and* ψ*. Therefore, we can model cross talk due to either source using the same correction, although noting that an offset angle would additionally affect the polarization gain ratio*."

Page 24776 Line 17

Taking this into account, we include a factor of 0.007 (absolute) due to cross-talk in the estimated volume depolarization error.

The value 0.007 is not a factor, but an absolute offset. The cross-talk error should be a relative error. See discussions above.

We agree it is not a factor; we changed the word to "component". As discussed above, the cross-talk has both a (fairly constant) offset and a relative portion, both of which are discussed in the text. This sentence refers to the offset portion.

Table 2

Instead of somehow arbitrary value combinations the real values for Table 1 should be used, and maybe some extreme values to show certain aspects. Furthermore, the equations used to calculate the factors and errors should be shown, which would be valuable for the readers to improve their own error calculation.

We included Table 2 to help build a sense of how the functions behave. If we used measurements for Table 2, all quantities would vary simultaneously, and it would be more difficult to discern the effects of varying each column. The actual measurements are used in the calculation of the error bars quoted in Table 1 (the table that shows the measurements) and in the figures, so we are not losing any information by additionally illustrating arbitrary combinations in Table 2. The equations are already given: A5 and A6 of the appendix. The partial derivatives aren't written out, but they are easily derived from Equation (2), which we have now clarified, as stated above.

The uncertainty for R is only +-5%, but for 1064 nm +-20% are mentioned in the paper.
Yes, table 2 is arbitrary values as already noted. However, Table 1 and the figures use the larger error bars for 1064 nm as discussed in the text, so there is no inconsistency that needs to be corrected.

Summary

The offset errors and the errors of the calibration factors should be separated as much as possible.
They are already separated in the discussions in the paper. We have added some clarification to the text as described above.

The polarization calibration error can be decreased and separated from the measurements error of the polarization angle by using the +-45° calibration.
See above. The waveplate is physically rotated to the inferred zero point (i.e. the angle that minimizes the depolarization), and so the offset angle is removed before science measurements are made; therefore we do not agree that the polarization calibration error can be decreased. If we had obtained +-45° calibration measurements after rotating the waveplate, it might be useful to confirm that there is no lingering error in the offset angle after the adjustment but we have not seen any indication of a consistent bias in setting the angle, which would be evident when looking at consecutive polarization calibrations. In any case, after adjusting the waveplate as indicated by the 7 angle calibration, we only did calibrations at +45°, so we cannot make the suggested change to the measurements in this paper.

The error of the polarization angle should be determined for each calibration separately and propagated to the corresponding measurements.
See response above. The waveplate is physically rotated to the inferred zero point as part of the calibration procedure, so there is no error in the polarization angle for subsequent measurements, as far as we can tell. We do not do another assessment of the angle immediately after adjusting the waveplate, so we have no information to use for correcting for an error in the angle. However, a small error is possible to the extent that the waveplate position may change between measurements, which is why we use the in-flight change of the measured calibration to estimate the systematic uncertainty for measurements obtained between calibrations.

The cross talk error from the polarizing beam splitters should be corrected.
As noted in the text, the cross-talk error from the polarizing beam splitter is made negligible compared to other error sources by the addition of "clean-up" PBS cubes. This is already acknowledged in the reviewer's comments "the extinction ratios of the polarizing beam splitter assemblies used in the HSRL-2 receiver are quite good" and "the effect is not sufficient to explain the assumed molecular linear depolarization ratios". The reported systematic uncertainties cover all sources of cross-talk. For the purpose of this paper, which is to highlight the spectral dependence of the aerosol linear depolarization ratio for the selected case studies, the systematic uncertainty estimates are sufficient small to show that the highlighted differences are real differences.

The determination of the backscatter ratio error should be described more detailed and its influence on the error of the linear depolarization ratio should be made more clear.

We have included a fair amount of detail about the total aerosol scattering ratio uncertainty and its influence on the linear depolarization ratio. We point out that the systematic uncertainty for the total aerosol scattering ratio is the systematic uncertainty in the gain, and that for 532 nm, the aerosol-to-molecular gain ratio calibration is the only source. For the other channels, there is additional uncertainty related to transferring the gain ratio calibration from 532 nm, and we describe how we estimated that additional uncertainty. The influence of the uncertainty in the total aerosol scattering ratio on the uncertainty in the linear depolarization ratio is given by equation A5. The uncertainty in the total aerosol scattering ratio is ΔR and the uncertainty in the linear particle depolarization ratio is $\Delta\delta_a$ on the left-hand side. The propagation factor F_R is a partial derivative given by equation A6 operating on Equation 2.

With a small error of the calibration factor, the more or less constant offset error can be accurately determined, and the values of the linear depolarization ratio can be corrected for that.

We already discussed correcting the more or less constant offset error in the original manuscript. We included a correction, and because the correction is not perfect, we also included a component in the uncertainty.

References:

Alvarez, J. M., Vaughan, M. A., Hostetler, C. A., Hunt, W. H., and Winker, D. M.: Calibration Technique for Polarization-Sensitive Lidars, J Atmos Ocean Tech, 23, 683-699, 10.1175/jtech1872.1, 2006.

Rogers, R. R., et al. (2009), NASA LaRC airborne high spectral resolution lidar aerosol measurements during MILAGRO: observations and validation, *Atmos Chem Phys, 9*(14), 4811-4826, doi: 10.5194/acp-9-4811-2009.