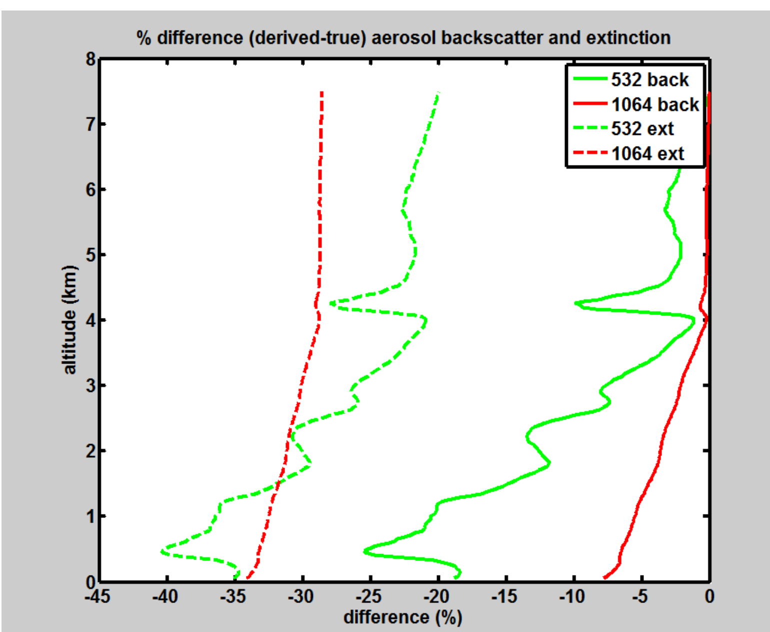
I and my coauthors thank all the reviewers for taking the time to read and comment on this manuscript. Many of the specific suggestions have been adopted in a revised manuscript. In some cases, we feel that the suggestions are out of the scope of the paper. Specific responses to the reviewers' suggestions are given below. The reviewers' comments are shown in blue (comments that don't include suggestions for improvement are not shown here). Our responses are shown in black with extended quotes from the revised manuscript set off in block form.

Response to Reviewer #1

"I assume that the interpretation of the conventional, non-HSRL measurements at 1064 nm in this context is more uncertain, since it is based on model assumptions. Therefore, a clear analysis on the suitability of the 1064 nm channels for this kind of aerosol classification is certainly mandatory."

Thanks for pointing out this omission. It should have been written explicitly that the backscatter coefficient from the 1064 nm channel depends on a conventional retrieval that uses an assumed lidar ratio of 35 sr, and this will be stated in the revision. The 1064 nm backscatter coefficient is therefore less reliable than the 532 nm backscatter coefficient which is a direct measurement using the HSRL technique. However, the retrieval at 1064 nm is less sensitive than shorter wavelengths to errors in lidar ratio (Sasano and Browell, 1989). The attached graphs shows the relative errors in the 532 nm and 1064 nm channels backscatter and extinction from standard (not HSRL) retrievals with a 10 sr error in the lidar ratio, showing that the 1064 nm backscatter has little sensitivity to the assumption of constant lidar ratio.



In addition, the 1064 nm calibration for the NASA HSRL instrument has minimal dependence on the absolute amount of aerosol in the calibration region, since the aerosol is measured well by the 532 nm HSRL channel. Instead of relying on an assumption of negligible or known aerosol in a clean region, only the less critical assumption of the wavelength dependence of the scattering in this region is needed. The backscatter ratio used here is therefore sufficient for the purpose of empirically separating aerosol classes. We have revised this section of the manuscript (in response to both Reviewer #1 and Reviewer #2). It now says, in part:

> The LaRC HSRL employs the HSRL technique at 532 nm and the standard backscatter technique (Fernald et al., 1972; Klett, 1981; Fernald, 1984) at 1064 nm, using an assumed lidar ratio of 35 sr at 1064 nm. The instrument also measures depolarization at both wavelengths. The return signal is split into components parallel and perpendicular to the polarization of the outgoing beam. The depolarization ratio here is defined as the ratio of the perpendicular to the parallel component. In contrast to standard backscatter lidars that are empirically calibrated by assuming that the aerosol contribution to backscatter is negligible or known at some altitude, the HSRL instrument is self-calibrating at 532 nm for measurements of aerosol and cloud backscatter and extinction. For the backscatter measurement which derives from a ratio of the two signals, the assessment of the relative gain ratio between the molecular and backscattering channels is performed in flight, by removing the iodine cell from the optical path and equalizing the input to the two channels. It is similarly self-calibrating at both 532 and 1064 nm for measurements of depolarization, by using the procedure given by Alvarez et al. (2006), involving rotating the half-waveplate to 22.5° to equalize the input to the "perpendicular" and "parallel" detector channels. The 532 nm extinction measurement is made from the derivative of the log of the molecular channel signal only, so the calibration constants are not required. The calibration of the 1064 nm aerosol and cloud backscatter measurement takes advantage of the internally calibrated HSRL measurement at 532 nm and therefore does not rely on an assumption of negligible or constant aerosol scattering in the calibration region. In contrast to ground-based lidar systems, the calibration region is close to the aircraft and therefore the attenuation is small. The instrument also features a unique autonomous boresighting system that insures the transmitter and receiver maintain co-alignment to very high accuracy during flight. A detailed description of this HSRL system and calibration and data retrieval techniques is provided by Hair et al. (2008).

And later:

> As stated above, the aerosol backscatter coefficient at 532 nm is a direct measurement made with the HSRL technique, while the backscatter coefficient at 1064 nm depends on a retrieval that uses an assumed lidar ratio. However, the standard backscatter retrievals at 1064 nm are less sensitive to errors in lidar ratio (Sasano and Browell, 1989), and the systematic error in 1064 nm backscatter is expected to be less than 15% at worst due to uncertainty in the choice of lidar ratio. The backscatter ratio used here is therefore sufficient for the purpose of empirically separating aerosol classes.

"Despite the authors mention the first HSRL measurement dated back to the years 1983/1992, the manuscript does not report on the state-of-the-art HSRL measurements. In the meantime new airborne HSRL instruments -other than the LaRC HSRLwith depol. capability at 532 and 1064 nm exist which have been applied in many studies for the measurement of desert dust, continental aerosol transport, biomass burning aerosol, etc. A brief discussion of these previous results should be included in this paper."

Based on the reviewer's recommendation, a reference was added to Esselborn et al. (2009) which gives an interesting analysis of lidar measurements of dust from observations by the DLR HSRL during SAMUM. The reference was added in the discussion of dust in Section 4. In addition, references to the SAMUM II papers by Groß et al. (2011) and Weinzierl et al. (2011) were added in the introduction. However, a full review of HSRL measurements is beyond the scope of the paper; rather, the focus of this manuscript is the new methodology we have applied to aerosol typing using our measurements.

"Further as a reader, I strongly miss a proof/validation of the results by in-situ measurements and/or transport model calculations to infer the origin of the measured air mass."

In the cases described in the paper, when we specifically say an airmass was advected from a certain region away from the measurement site, we have done backtrajectory analyses using the HYSPLIT tool on the NOAA READY website to arrive at or confirm this idea, but we failed to mention this in the manuscript. In the revised manuscript, we have added this information; however, we have not included figures illustrating the trajectories for space considerations. As for the more general issue of validation, we acknowledge that this is a new and largely untested methodology. The manuscript does already include some case studies where coincident aerosol composition data from the PALMS in situ instrument support the aerosol classification inferred by this method (Figures 16 and 17 and associated discussion). Our group is currently working on further comparisons with a variety of instruments and models, which will appear in another paper by our group. We believe this is a rich area for future research, but these comparisons are essentially beyond the scope of this paper. We have described the product of our new methodology as a qualitative classification; we are not attempting to provide specific aerosol microphysical characteristics that are directly comparable with in situ instruments. The methodology to categorize the very extensive dataset of HSRL intensive aerosol measurements is a tool that helps us crystallize our understanding of the HSRL data and is a springboard for future research and more advanced algorithms. Furthermore, the current manuscript is already quite long, unfortunately, longer than approximately 86% of the other papers currently under discussion in this journal, and a full comparison with other instruments would make it quite unwieldy.

"Also the methodology used to establish the aerosol classification based on the lidar measurements using LaRC HSRL is not really transparent to the reader."

We have addressed the reviewer's specific comments (below) about the methodology section as much as possible, and additionally have made some revisions which we hope will make this section clearer, including adding a few more introductory sentences describing the method in general terms to help orient the reader, and an equation defining the Mahalanobis distance.

"Page 5635, line 28: the argumentation on the use of the two depol. channels for aerosol classification is not obvious and need a better justification. Is this true for all types of aerosols?"

In Section 4.2 of the manuscript, it is shown how the spectral dependence of depolarization in the HSRL measurements is found to differ between pairs of aerosol types that are otherwise difficult to separate, specifically urban vs. biomass burning and dust vs. ice crystal haze (which we see in arctic observations). The revised manuscript has added some words to the sentence on page 5635 to alert the reader that more details are in the later section.

"Page 5636, line 25: the description of the status of airborne HSRL is not in line with the state-of-the-art instruments"

The section indicated in this comment is the instrument section in which we describe the NASA Langley instrument which was used to obtain the data we analyzed. It's difficult to figure out what exactly the reviewer objects to. If the reviewer is suggesting that the NASA airborne HSRL is not "state-of-the-art," then we disagree. From what we see described in the available literature, the DLR HSRL appears to have similar capabilities to the NASA airborne HSRL. Full details of the NASA HSRL instrument are given by Hair et al. (2008). We have cut out a statement on page 5638 that uses the phrase "state-of-the-art" in a slightly different context, because it was not really necessary, in case this was the problem. On the other hand, the reviewer may have been suggesting that we should include a description of other contemporary HSRL instruments here. As indicated in response to the general comments, we have added some discussion of papers describing measurements by the DLR airborne HSRL and by ground-based HSRL instruments in the introduction and in the discussion of specific aerosol types, but not in this particular section, which is just meant to describe the instrument from which our data is obtained.

"Page 5638, lines 2-13: It is not clear how depolarisation and backscatter coefficients are derived quantitatively at 1064 nm and how the model assumption impact on the aerosol classification for the different types of aerosol"

See above in the response to the general comments. The missing explanation has been added to the revised paper. The standard backscatter retrieval at 1064 nm uses an assumed constant lidar ratio of 35 sr.

"Page 5640, lines 20-24: Why not show the FLEXPART simulations on order to better justify the source of the aerosols for this particular case."

This same example was already published in the cited WRF-Flexpart-focused article, so we chose not to reproduce the figure here. The citation of de Foy et al (2011) has been revised to include the figure number, to make it clearer that the citation refers to this exact comparison, rather than to WRF-Flexpart in general.

"Page 5641-5644: This subsection on the methodology needs considerable revision. For the reader it is not clear what algorithms already exist and how they work in the particular case."

Some wording has been added to the revision to clarify that the methodology described here is new, and how it relates to other methods that are mentioned. For example, in the introduction after discussing the aerosol classification work by Cattrall et al. (2005) and Omar et al. (2005), the revision says, "The classification methodology presented in the current work combines the strategies of contextual identification of selected cases with automated classification of the bulk of measurements, and will be described more fully in Section 4." Another sentence was added to make it clear that use of the Mahalanobis distance is not original: "The Mahalanobis distance is frequently the distance metric of choice for Expectation Maximization (EM) clustering (Dempster et al., 1977), which is a generalized form of k-means clustering." The following sentence was added to emphasize one of the main ways in which this classification differs from EM and k-means classification: "The strategy of using labeled samples to create 'seed' aerosol class models to classify all other measurements allows us to incorporate knowledge based on a relatively limited set of observations where the aerosol type is known or easy to infer."

"What is the justification of using just 8 aerosol types and what criteria are used to isolate 'known aerosol types'."

Choosing the number of aerosol types is always somewhat subjective in all of the classification methods we know about. We have added the following text to supplement the discussion already in the manuscript:

> The HSRL aerosol classification is performed in two parts. First, specific samples of known aerosol types are combined to make model distributions. Second, the full dataset of HSRL measurements are classified by comparison with these models. The number of classes depends to some extent on the cases where aerosol type is known with high confidence, and should not be considered definitive. The choice of classes also reflects a desire for the categories to be physically meaningful and suggest possible aerosol sources; to this end, the basic classes follow the existing literature (Dubovik et al., 2002; Cattrall et al., 2005; Omar et al., 2005). In addition, it is desirable for the number of classes to maximize the information content captured by the class identification. A statistic related to this idea is described in Section 4.3. The HSRL aerosol classification described herein uses eight classes, which start with labeled samples of known aerosol types. Section 4.2 describes the eight classes and how the samples were chosen.

In addition, all the subsections of section 4.2, describing the individual types, are essentially an attempt to justify the choice of classes by laying out the details of each class for the readers.

"The impact of mixing/transport, aging, and signal ambiguity need to be discussed in more detail. Despite a few obvious cases where the origin of the aerosols are more or less "visible" (smoke plumes, strong haze layers, etc.) I would expect that one need an appropriate transport model for such kind of classification."

Yes, the reviewer makes many good points here. Aerosol type identification is complicated for all the reasons listed. However, there are indeed some "obvious cases" and the point of our methodology is to leverage the information from those relatively few obvious cases to many other cases where the aerosol type is not obvious. We started with 30 samples where we felt that the type was relatively easy to infer. The classification of the rest of the HSRL data are results of the classification methodology we introduce here. In the final manuscript, we have revised the description of the 30 labeled samples to be more specific, to justify why we thought the types in these cases could be inferred. In some of those cases we did in fact use backtrajectory analysis, but it was not necessary in all 30 cases. Here is an excerpt from the revision:

Specifically, we incorporate six samples of ice haze (see Section 4.2.1) observed during the ARCTAS campaign, identifiable by the signature of fall-streaks in the lidar measurements. Pure dust is represented by two labeled samples. One of these was a plume of Saharan dust tracked by the CALIPSO lidar instrument as it was advected from Africa to the Gulf of Mexico, described by Liu et al. (2008). CALIPSO and HSRL observed it simultaneously near Houston, Texas ten days later. The other pure dust labeled sample is from a dust storm on the slope of Pico de Orizaba observed during the MILAGRO field campaign described by de Foy et al. (2011). Three labeled samples of dusty mix include cases of probably locally generated dust with intermediate values of depolarization, including two in the midwestern United States and one near Mexico City that is also discussed by de Foy et al. (2011). Clean air samples in the Caribbean provided most of five labeled samples for the maritime class. Labeling of two samples of polluted marine air from the marine boundary layer in the Gulf of Mexico and near the coast of Virginia was justified by backtrajectory analysis using the online HYSPLIT tool from the NOAA Air Resources Laboratory READY website (http://ready.arl.noaa.gov/HYSPLIT.php) which was used to track the air samples from the marine boundary layer a short time backward to urban areas. Four urban samples are used where the attribution of elevated levels of aerosol optical depth to urban sources is fairly straightforward, for example near Washington D.C. or, in one case, Mexico City in a region where the WRF-Flexpart model also indicates urban aerosols (de Foy et al., 2011). In the case of smoke (five samples) and fresh smoke (three samples), the plume was observed visually from the B200 or was measured by coincident airborne in situ measurements (Warneke et al., 2010) and/or MODIS images (Saha et al., 2010, see Figure S7a).

"Page 5647-5665: The authors do not report on adequate validation efforts which are required to prove their conclusions. They should also compare with results from recent airborne HSRL measurements during SAMUM I, II and EUCAARI campaigns."

Please see the response to the related comment in the general comments section. As to specific comparisons with recent DLR airborne HSRL measurements, we have included Esselborn et al.'s (2009) study of dust during SAMUM I in the discussion in the dust section, where we now say, "Observed values of 532-nm aerosol depolarization for this case are about 33% and 532-nm lidar ratio values are 49 ± 9 sr. These lidar ratio values are consistent with lidar ratio values of 53-55 ± 7 given by Tesche et al. (2009b) and also within the range of variability of 38-50 sr observed by Esselborn et al. (2009), both for Saharan dust nearer to the source during SAMUM (Heintzenberg, 2009). Esselborn et al. (2009) show by backtrajectory analysis that their observed variability in lidar ratio is primarily attributable to differences in source regions." References to Tesche et al. (2009) (SAMUM I), and Tesche et al. (2011), Groß et al. (2011) and Weinzierl et al. (2011) (SAMUM II) have also been added to our revised manuscript. However, comparison with DLR HSRL observations would not constitute a validation of either our method or the DLR HSRL measurements, since the NASA and DLR airborne HSRL instruments have never flown coincident flights, or even flown in the same region of the world. Esselborn et al. (2009) do an excellent job of highlighting the fact that aerosol from different sources of even the same type (in this case dust) may have differences in lidar observables and they also point out (as the reviewer does) that the lidar observables can be affected by transport. For this reason, we do not present our results as a climatology of aerosol type, but rather only as a classification of the measurements included in this study.

"List of references need to be updated"

We have added several references as mentioned above, and also as recommended by the other reviewers.

"Many of the figures are too busy and axis labelling not readable"

We have increased the font size on the axis labeling on many of the figures and changed the gray axis labeling in Figures 3 and 10 to black.

Response to Reviewer #2

Many thanks to the reviewer for the positive comments and helpful detailed suggestions.

In partial response to the general comments, we have increased the font size on the axis labeling on several of the figures, to help with readability.

Specific comments:

p. 5636: "Since the intensive variables do not depend on the amount of aerosol loading, there is a much smaller effective limitation on the loading that can be used for classification than what was required for Dubovik et al. (2002) and Cattrall et al. (2005)." This is probably true, but the observation would be more useful if the authors provided an estimate of how large layer AOT or height-resolved backscatter needs to be to obtain adequate signal for a successful classification.

We estimate the minimum requirement for the aerosol backscatter signal to be approximately 0.0003 km^{-1} sr^{-1}. This signal level generally produces measurements of the four intensive parameters that are reliable and pass the criteria given in Table 2, and yield an inference of aerosol type that seems reliable from the standpoint of consistency among nearest neighbors. This is equivalent to about 0.015 km^{-1} in aerosol extinction, or 0.015 in aerosol optical thickness, assuming a lidar ratio of about 50 sr and assuming the aerosol is contained in a 1 km thick boundary layer.

p. 5637: "The HSRL instrument is self-calibrating at 532nm for measurements of aerosol and cloud backscatter and extinction". A few words as to why the technique is self-calibrating would help.

The HSRL calibration techniques are described much more thoroughly in Hair et al. (2008) than can be done here, but this section has been revised to help round out the description somewhat. Specifically, the text has been revised to say,

> In contrast to standard backscatter lidars that are empirically calibrated by assuming that the aerosol contribution to backscatter is negligible or known at some altitude, the HSRL instrument is self-calibrating at 532 nm for measurements of aerosol and cloud backscatter and extinction. For the backscatter measurement which derives from a ratio of the two signals, the assessment of the relative gain ratio between the molecular and backscattering channels is performed in flight, by removing the iodine cell from the optical path and equalizing the input to the two channels. It is similarly self-calibrating at both 532 and 1064 nm for measurements of depolarization, by using the procedure given by Alvarez et al. (2006), involving rotating the half-waveplate to 22.5° to equalize the input to the "perpendicular" and "parallel" detector channels. The 532 nm extinction measurement is made from the derivative of the log of the molecular channel signal only, so the calibration constants are not required. The calibration of the 1064 nm aerosol and cloud backscatter measurement takes advantage of the internally calibrated HSRL measurement at 532 nm and therefore does not rely on an assumption of negligible or constant aerosol scattering in the calibration region. In contrast to ground-based lidar systems, the calibration region is close to the aircraft and therefore the attenuation is small. The instrument also features a unique autonomous boresighting system that insures the transmitter and receiver maintain co-alignment to very high accuracy during flight. A detailed description of this

HSRL system and calibration and data retrieval techniques is provided by Hair et al. (2008).

p. 5638: "aerosol wavelength dependence, which is the Ångström exponent for aerosol backscatter". It would be better to give the equation for Angstrom exponent for backscatter.

Added.

p. 5638: "Ames Airborne Sun Photometer (AATS-14)". "Tracking" is missing between Airborne and Sun. Sun Photometer should be Sunphotometer: Ames Airborne Tracking Sunphotometer (AATS-14).

Fixed.

p. 5638: "3% (0.01 km^{-1}) . . . 50% (0.015 km^{-1})." Something seems wrong in the comparison as stated. 50% is 17 times 3%. But 0.015 km^{-1} is only half again larger than 0.01 km^{-1}. Is something amiss?

This appears to be a typo in the abstract of the Rogers et al. paper which I quoted uncritically. It should be 0.001 (3%) for the bias difference, consistent with the values shown in Rogers et al. Table 2 for individual comparisons.

p. 5639: "The campaigns include many process-oriented field projects for NASA, the Department of Energy (DOE), and the Environmental Protection Agency (EPA)". NOAA is missing from this list but included in Fig. 1 (GoMACCS, CalNex).

Added.

p. 5639: "locations of these missions has enabled". "has" should be "have".

Fixed.

p. 5639: "". Another reference to ARCTAS smoke measurements is: Shinozuka, Y., Redemann, J., Livingston, J. M., Russell, P. B., Clarke, A. D., Howell, S. G., Freitag, S., O'Neill, N. T., Reid, E. A., Johnson, R., Ramachandran, S., McNaughton, C. S., Kapustin, V. N., Brekhovskikh, V., Holben, B. N., and McArthur, L. J. B.: Airborne observation of aerosol optical depth during ARCTAS: vertical profiles, inter-comparison and fine-mode fraction, Atmos. Chem. Phys., 11, 3673-3688, doi:10.5194/acp-11-3673-2011, 2011.

Added.

p. 5642: "The classification algorithm itself is not sensitive to outliers and noise and including or eliminating them has no effect on the classification of the remainder of observations." This is an unexpected statement and seems to contradict the following statement from p. 5651: "For the HSRL data classified into eight classes as described above, Wilks' lambda is 0.083. If outliers are also included, the value is 0.137". The apparent contradiction should be resolved.

It's good that the reviewer mentioned this, because I didn't anticipate this way of reading the statement. What was meant by this was that including or excluding outliers or noisy points does not have any effect on the classification of other points. In contrast, unsupervised clustering methods like k-means and expectation maximization rely on global minimization, so changing the number of points classified, particularly if they are outliers, can potentially have a huge effect on the results, potentially making many changes to the identification of individual points. The current method does not have that disadvantage, since the classification of individual observations is independent of the classification of other points, the number of points classified, the overall variance in each dimension, etc., none of which is true for k-means or Expectation Maximization. However, it wasn't meant to imply that the classification works equally well for bad points and good points. The classification of less reliable data (outliers or noisy points) will be less reliable (i.e. more likely to be wrong) than the classification of high-quality measurements. So a global statistic taking all points into account will be degraded by the inclusion of less reliable measurements. Filtering criteria are therefore still useful, for helping to ensure that the bulk of classifications are reliable, but filtering is not critical for success in the way that it is for the global clustering methods discussed above, and so the filtering can be more conservative and is not really necessary. In the revised manuscript, this sentence has been reworked: "Typing of outliers or noisy points with the aerosol classification algorithm will be less reliable than for well-behaved points, but since each measurement is classified independently, the inclusion or exclusion of outliers has no effect on the classification of the remainder of observations, in contrast to global minimization algorithms like k-means (MacQueen, 1967)."

p.5649: "Russell et al. (2010) indicate that absorption Ångström exponent [AAE] derived from AERONET shows promise for separating them [urban and smoke aerosols]". Actually, Russell et al. (2010) speak of "Ambiguities in aerosol composition or mixtures thereof, resulting from intermediate AAE values" and note the need to reduce these ambiguities "via cluster analyses that supplement AAE with other variables, for example Extinction Angstrom Exponent (EAE), which is an indicator of particle size." They also note: "the value of combining several different types of remotely sensed information in multidimensional cluster analyses to derive the most information on aerosol type (e.g., to reduce the ambiguity resulting from the partial overlap of the biomass burning and urban-industrial clusters in Fig. 5 [a plot of AAE vs EAE])". Some rewording of the p. 5649 statement is in order to note these ambiguities and the need for multidimensional analyses.

Indeed, the thrust of Russell et al. (2010) should have been described more fully and accurately. The statement has been revised to "Russell et al. (2010) indicate that absorption Ångström exponents (AAE) derived from AERONET are strongly correlated with aerosol type, but display some ambiguity between urban-industrial aerosol and biomass burning aerosol. They demonstrate that multi-dimensional analysis consisting of the combination of AAE and extinction Ångström exponent (EAE), for example, shows potential to more fully resolve these types."

p. 5649: "The elevated smoke layer also has slightly higher particulate depolarization (8–10%)". At 532 nm? This is one of several examples where "depolarization" means "depolarization at 532 nm". Being more explicit would help clarity.

Agreed. I added "532 nm" here and in many other places to try to improve clarity.

p. 5651: "The dust category and the "fresh smoke" category". I think "dust category" here means "pure dust category". Clarify.

Yes, pure dust. Fixed.

p. 5652: "The ranges of the intensive parameters applicable to each of the aerosol classes are displayed in Fig. 10". Showing the Wilks partial lambda value in each frame of Fig. 10, and referring to it here, would help show the correspondence of this parameter to the spread of measured lidar parameters (relative to their standard deviations) among aerosol types.

Thanks, this is a nice idea. In the revised manuscript, the figure has been annotated with the Wilks' partial lambda values, and a brief description is repeated at this point in the text.

p. 5654: "the sum of the AOT for the two classes exceeds the dust fraction as computed using the Sugimoto and Lee (2006) algorithm." I don't understand how one can compare dust fraction and absolute AOT. Is the wording amiss here?

Yes, the wording was poor (but the figures are fine). The partial optical depth attributed to dust should not have been described as "dust fraction". The wording has been reworked in this paragraph and the figure caption to say "dust partial optical depth" or "optical depth due to dust" wherever it previously said dust fraction.

p. 5672 (Fig. 2) and other 6-frame figures like it: I think it would improve clarity if the Frame (f) right axis label were "Backscatter Wavelength Dependence" or "Backscatter Angstrom Exponent" rather than "Aerosol Wavelength Dependence".

All of these figures were remade with the backscatter color ratio instead of the backscatter angstrom exponent (aka wavelength dependence) for improved consistency.

p. 5673 (Fig. 3): Here the terminology (and variable) are "Backscatter Color Ratio" rather than "Aerosol Wavelength Dependence", "Backscatter Wavelength Dependence or "Backscatter Angstrom Exponent". It would help to choose one variable and terminology and stick with it throughout.

All the figures have been converted to backscatter color ratio for consistency.

p. 5674 (Fig. 4): A more explicit description of the "two-sigma covariance" used to draw the ellipses would help (maybe in an appendix?).

More explanation and a reference have been added in the paragraph on page 5674 that describes figure 4. The revision states "Two-sigma covariance ellipses for these models are also shown in Figure 4. The ellipse used to visualize a covariance matrix is determined by its eigenvalues and eigenvectors (see, e.g., Rodgers, 2000). The square roots of the eigenvalues are the major and minor axes (doubled in this figure to represent two-sigma variability). These are not necessarily aligned with the variable axes or equal to the standard deviations for each variable, since the covariance terms (off-diagonals) of the variance/covariance matrix are not necessarily zero. The directions of the major and minor axes are given by the eigenvectors." I hope an appendix isn't necessary. Since the technique of illustrating covariance matrices with ellipses isn't original, an appendix may not even be appropriate.

Also: "Aerosol Depolarization" on the horizontal axis of the right frame. It's probably at 532 nm, but it should be stated.

Revised the figure to specify the wavelength of the aerosol depolarization.

p. 5680 (Fig. 10): See p. 5652 comment above. And again: "Aerosol Depolarization" on the horizontal axis of the top left frame. It's probably at 532 nm, but it should be stated.

The clarification of the wavelength, and also the Wilks' partial lambda annotations, have been added in the revised manuscript.

Response to Reviewer #3

"Page 5633, line 21 : Add reference to McCormick et al., BAMS, 1993."

Added.

"Page 5637, line 6 : Add a reference on extinction profile retrieval, may be Burton et al., 2010."

Actually, at this line, the text in the manuscript is still summarizing the HSRL method from the HSRL references. A few lines later, I added these references on extinction profile retrievals using the standard backscatter technique: (Fernald et al. 1972; Klett et al. 1981; Fernald 1984.)

"Page 5638, lines 19 and 20 : numbers in parentheses are not coherent with percentages given."

This appears to be a typo in the abstract of the Rogers et al. paper which I quoted uncritically. It should be 0.001 (3%) for the bias difference, consistent with the values shown in Rogers et al. Table 2 for individual comparisons.

"Page 5640, line 5 : Add an older reference to Sassen on depolarization and cirrus clouds"

Added.

"Page 5641, line 26 : 'aerosol' badly written"

Fixed.

"Page 5642,line 16 : the identification of 8 classes appears as a magic numbering. It is discussed in the text further on from observations, as why this number has been identified, but the authors may discuss more precisely the objectives at the beginning, not only ona posteriori basis(should they have a large SNR on all signals a large number may be possible which may not be needed)…."

Ideally the choice of the number of classes would maximize the fraction of the variability that is captured by the class identification. The Wilks' lambda statistic is one way to probe this. However, this is not the only consideration. We also wish the classes to reflect categories that make physical sense and to some extent follow the existing literature of aerosol classification represented by Dubovik et al. (2002), Cattrall et al.(2005) and Omar et al.(2005). The desire for physically important distinction played a role in choosing some of the classes. For example, we would not want to leave out the "ice" class and lump those cases with dust, because these are clearly very different physical categories, even though these are somewhat difficult to separate, as discussed in the manuscript. On the other hand, "polluted maritime" is believed to be a mixture of pollution and marine air and therefore its usefulness as a distinct type is more debatable. However, the cases falling in this category have observable differences in the aerosol intensive properties compared to the maritime and pollution classes, and the inclusion of this class does indeed increase the amount of variability captured by the class identification (that is, it decreases Wilks' lambda – from 0.095 without the class to 0.083 with it). The distinction in the measurements between the two smoke classes (what we have called smoke and fresh smoke) was found empirically. Including smoke as two distinct classes increases the overall success of the classification both in terms of increasing the fraction of variability captured by the classification and in terms of identifying cases known a priori—but not specified—to be smoke plumes. The classes presented here are not necessarily definitive. They also depend on the number of cases where we know with high confidence the type of aerosol present. As we continue to make measurements in more field missions, and as we continue to make comparisons with other instruments, we may find that additional classes are called for. The following text has been added near the line number indicated by the reviewer:

> The HSRL aerosol classification is performed in two parts. First, specific samples of known aerosol types are combined to make model distributions. Second, the full dataset of HSRL measurements are classified by comparison with these models. The number of classes depends to some extent on the cases where aerosol type is known with high confidence, and should not be considered definitive. The choice of classes also reflects a desire for the categories to be physically meaningful and suggest possible aerosol sources; to this end, the basic classes follow the existing literature (Dubovik et al., 2002; Cattrall et al., 2005; Omar et al., 2005). In addition, it is desirable for the number of classes to maximize the information content captured by the class identification. A statistic related to this idea is described in Section 4.3. The HSRL aerosol classification described herein uses eight classes, which start with labeled samples of known aerosol types. Section 4.2 describes the eight classes and how the samples were chosen.

"…A short discussion on the needs with respect to in situ observations and occurrence, improving and extending what has been done already (as mentioned in the introduction) as well as need for modelling/understanding interactions and radiative forcings may be appropriate to include in this section, as an introduction."

In the revision, we made slight additions to the discussion of future needs. The last paragraph of section 4 has been revised in part to say:

> Further confidence in the results can be gained by comparisons with other data sets, both comparisons with the properties of aerosol types already presented in the literature, and more specific comparisons between coincident measurements with aerosol in situ composition instruments. These comparisons are begun in the next section and carried forward in a future paper (Ferrare et al., paper in preparation). Additional research on comparisons with aerosol in situ instruments and models is ongoing.

"Page 5643, line 1 : some reference on k-means clustering may be useful here (MacQueen)

In the revision, the MacQueen reference has been repeated here and in other places where k-means clustering is mentioned.

"Page 5649 and 5650 and conclusion : properties of aged carbonated particles are different as discussed for smoke in p. 5650. It is not clear if this difference can be really taken into account. Discuss a little with respect to SNR and potential improvements."

We acknowledge that the differences between the categories we have labeled "smoke" and "fresh smoke" are not yet well understood, and that the four intensive variables we have give only an incomplete picture of the distinction. Coincident HSRL and in situ measurements of smoke plumes of a variety of ages would of course help. Also, a future airborne $3\beta + 2\alpha + 2\delta$ lidar instrument (that is, an instrument with HSRL capability at 355 nm in addition to the capabilities of the airborne HSRL described here) would give additional discriminatory power, as shown by the microphysical retrievals described by Müller et al. (e.g. 1999, 2007).

Response to Reviewer #4

"The paper contains very interesting and original material, obtained with an excellent airborne HSRL. However, many points of the manuscript must be improved, are not just carefully written. For example, a better literature review concerning airborne aerosol HSRL is necessary. At the moment the reader gets the impression, only North American groups are active in this field. That is simply not true. Furthermore, many observational findings are again set into the context of North American aerosol observations, widely ignoring the long list of papers and efforts regarding multiwavelength lidar and aerosol typing since the beginning of EARLINET or even earlier."

We were not so ambitious as to present this paper as a review paper. We have no objection to including more references and in the revised manuscript have adopted many of the specific suggestions for additional references that the reviewer provides, but we still do not consider a review of all lidar measurements of aerosol to be a goal of this paper. We spent time in the introduction discussing the Cattrall and Omar papers because these papers are most related to the aerosol classification method presented in this manuscript. That is, we are specifically discussing classification methodologies, not measurements of aerosols generally. We certainly agree with the reviewer that many additional case studies of aerosol of different types have been published since the Müller et al. 2007 paper. That paper is excellent and seems to represent a truly exhaustive summary of lidar measurements of aerosol up to that time, but the current manuscript is not meant to be a sequel to that paper. Rather than a catalog of case studies of aerosol types, we present the paper as a methodology paper addressing a new technique for classifying the measurements that our group has made with our instrument. Other attempts to classify large datasets of many aerosol types in diverse locations are relatively few, we still believe, and the Dubovik, Cattrall, and Omar papers are indeed the key papers providing the groundwork for the study described in this manuscript. Similarly, we are not presenting this manuscript as an instrument paper. However, in the revision we have added some of the suggested references which describe other lidar measurements of aerosols of particular type. Some examples of revised text are included below in the responses to specific comments.

"A statistical analysis of aerosol observations performed in the framework of 18 field campaigns is presented. That is unique! But, there is surprisingly no attempt to use any air mass transport analysis to distinguish between different aerosol types and characteristics, and, at least, to demonstrate, for some cases, shown in the figures, that the statistical approach makes really sense and is in agreement with the backward trajectories. This would then be convincing.."

For the cases described in the paper, when we have said that an airmass was advected from a certain region away from the measurement site, we have done backtrajectory analyses using the HYSPLIT tool on the NOAA READY website to arrive at or confirm this idea, but we failed to mention this in the manuscript. In the revised manuscript, we have added this statement. We have not included figures illustrating the trajectories in the manuscript for space considerations, but some of them are included below in response to the specific comments of the reviewer.

"Some quantities in the paper are not explained. Sometimes AOT is used, sometimes AOD. Instead to use backscatter-related Angstrom exponents throughout the paper, the authors present: e.g., aerosol wavelength dependence, backscatter color ratio, backscatter spectral ratio, aerosol backscatter-related Angstrom exponent, backscattering Angstrom exponent. In the case of the wavelength dependence of the depolarization ratio: ratio of aerosol depolarization ratios, spectral depolarization ratio, depolarization spectral ratio, particle depolarization spectral ratio, aerosol depolarization spectral ratio."

Revisions have been made to use AOD rather than AOT everywhere, and to consistently use the backscatter color ratio rather than the related angstrom exponent (which was called by various names in the original manuscript). In the case of the spectral depolarization ratio, we also made some changes, for instance to consistently use the phrase "spectral ratio" rather than "wavelength ratio". The depolarization and spectral depolarization ratios used throughout have the molecular component removed. We usually refer to the result as "aerosol depolarization ratio"; however, when talking about ice crystal haze, we use "particulate" instead of aerosol.

"p. 5633, lines 16-27: This is a good example of ignoring the non American papers on aerosol typing. Furthermore: Why do you concentrate here on simple backscatter lidars?"

This seems like a strange objection, which I suppose results from misunderstanding how the introductory text is organized. We are attempting to summarize previous work on classification methodologies that present a complete set of aerosol types, starting with the oldest. This paragraph concentrates on backscatter lidars, but the very next paragraph points out the weakness of backscatter lidar and sets up the discussion of HSRL and Raman lidar later in the introduction. Both the introduction and later parts of the paper include references to non-American papers, even if they don't appear in this particular paragraph. The paper by Müller et al. (2007) in particular is referenced repeatedly and many of the studies referenced in that paper are also mentioned specifically in this manuscript.

"p. 5634, lines 7-23: Now the focus of the literature review is on AERONET. This is also not satisfactory. There are at least 30 lidar papers since 2000 (or even earlier, starting with the Ferrare papers, JGR 1998) dealing with high quality lidar observations with focus on aerosol typing. You may not know that the Cattrall paper is based on modelled lidar ratios. No measured ones are presented. There was later on one lidar paper (Muller 2007). In this paper the Cattrall results were compared with real world lidar ratios (from Raman lidar). And it was shown (if I remember well) that the Cattrall data are partly questionable (because they are modelled, rather then measured). So, I suggest to perform a literature review on HSRL, and on aerosol typing with appropriate lidars (Raman lidar, HSRL)."

Again, we are summarizing previous work that is directly relevant to the current study's focus on aerosol classification methodology. Of course we know that Cattrall is based on modeled lidar ratios; that point was already acknowledged in the submitted manuscript, and that is why we proceeded from the discussion of Cattrall et al. (2005) to discussion of Müller et al. (2007) which, as the reviewer says, focuses on aerosol typing using lidar measurements. The submitted manuscript already makes these points at lines 8-13 on page 5635:

> The advantage of lidar is the ability to provide vertically resolved measurements, and Raman lidar (Ansmann et al., 1990) and High Spectral Resolution Lidar (Shipley et al., 1983; Grund and Eloranta, 1991; She et al., 1992) have the additional key advantage over backscatter lidar , in that they measure aerosol extinction and backscatter coefficients independently, without using models or assumptions about aerosol type.

In the revision, we also added a description of two aerosol typing papers that appeared in the SAMUM II special issue:

> Measurements of the lidar ratio from ground-based Raman lidars along with aerosol depolarization values were shown to be useful for the separation of aerosol types by Groß et al. (2011), but from a more limited set of observations

> and aerosol types including pure dust and biomass burning mixed with dust from the recent Saharan mineral dust experiment-2 (SAMUM-2) field mission. Weinzierl et al. (2011) classify aerosol from the same mission using these two lidar observables and absorption Ångström exponent calculated from in situ data.

We take the reviewer's point that there are additional appropriate lidar measurements of aerosols of specific types such as those described in the Müller et al. (2007) paper but later than that paper. Therefore, we have also added the following sentence to the revision:

> More recently [i.e. than the referenced studies that use sun-photometers], many additional high quality case studies characterizing vertically resolved aerosol optical properties of specific aerosol types world-wide have been made with ground-based Raman and HSRL lidars (e.g. Müller et al., 2007a and references therein; Amiridis et al., 2009; Noh et al., 2009; Tesche et al., 2009; Giannakaki et al., 2010; Alados-Arboledas et al., 2011; Tesche et al., 2011 and many others) and with airborne HSRL (Esselborn et al., 2009).

Some of these references have been added in response to the reviewers' comments. Several others as well as many that have been referenced in Müller et al. (2007) were already discussed in the submitted version of the manuscript in the later sections of the paper dealing with specific aerosol types, but I see that including only the Müller et al. (2007) reference in the introduction was off-putting, so the sentence above was added to the revision. As indicated in the quote above by "e.g." and "and many others", we realize that we are probably still missing other lidar measurements of specific aerosol type, but we do not expect this manuscript to be viewed as a review paper.

"p. 5636, line 24: : :..: Please add Chinese HSRL observations (Liu, 2002), Esselborn (Tellus, JGR, Appl. Opt) on HSRL, Wandinger (JGR, 2002) showing airborne HSRL data. The list of papers dealing with aerosol HSRL is not that long, thus an exhausting review is justified here."

Some of the suggested papers have been added to later parts of the manuscript where specific aerosol types are discussed, and in the introduction as quoted above. However, the reviewer is suggesting adding a review of all HSRL measurements at the point in the manuscript where the NASA Langley airborne HSRL instrument is described. We included a description of the NASA airborne HSRL because that is the instrument that was used to obtain all of our data. No offense is intended in the fact that the introduction doesn't reference the DLR airborne HSRL instrument paper or earlier ground based HSRL instruments, such as the one described in Liu et al. (2002), but again, we do not feel that it is our responsibility to provide a thorough review of the HSRL measurement technique in this manuscript nor provide a review of all HSRL aerosol-related measurements. We focus instead on describing previous lidar measurements that have been use to identify and classify aerosols. This study is not meant to introduce our instrument or our measurements. That task was done already in the paper by Hair et al. (2008) which indeed includes many more references to other HSRL instruments and a much more extensive description of the HSRL technique and calibrations. There is ample precedent for describing only the instrument which was used to collect the data being analyzed. For example, a quick check confirms that the NASA Langley airborne HSRL is not referenced in papers focused on the measurements from the DLR airborne HSRL (e.g. Freudenthaler et al. 2009, Wandinger et al. 2009, Esselborn et al. 2009, Weinzierl et al. 2011, etc.) or ground-based HSRL measurements. I'm not surprised by this, since they are also data-driven papers and correctly describe the instrument(s) from which the analyzed data is obtained. The reviewer is probably correct that the list of all HSRL measurements is not that long, but a review of them should not be necessary here in a data analysis paper.

"p. 5639, l. 25, please review not only the work of Japanese and American groups. It is almost impossible to measure desert dust depolarization ratios east of China. The dust is mixed and contaminated. Most depolarization ratios are then clearly below 30%. The best available depolarization ratio data set for desert dust is presented by Freudenthaler et al. (2009). They measured desert dust directly close to the source."

Freudenthaler et al 2009 has been added to the list of references here in the revised manuscript. Of the other references given at this line, Sugimoto and Lee do a calculation of dust fraction that includes a value for pure dust, so is a valid reference here. Liu et al. tracked a dust plume across the Atlantic from a dust storm in the Sahara and discuss how the lidar properties are fairly constant as the dust layer was transported.

"p. 5640, lines15-16: The lidar ratio of desert dust is around 50sr because of the irregular shape. Absorption does not play any role. Marine particles seem to be always spherical (because RH is always very high). The lidar ratio of marine particles is then determined by their large sizes (and thus close to 20sr). Concerning absorbing particles, please check the latest SAMUM special issue (Tellus 2011, Gross papers, Tesche Papers, Weinzierl paper, the latter includes HSRL observation!!!)."

In the revision "weakly absorbing" was deleted so that the phrase now simply says "low values of approximately 20 to 50 sr at 532 nm for coarse mode particles". The observation that absorbing particles have much larger lidar ratios was well established before the SAMUM 2 special issue and the references given in the submitted manuscript are sufficient to establish the point.

“p. 5640, line 22: Lower lidar ratio and higher depolarization ratio : : : suggest higher concentration of dust? Please be more specific: : :! Is the reference level around 90sr? Then it would make sense.”

This paragraph is describing the measurements shown in Figure 2 and contrasting the eastern part of the city with the western part of the city. “Lower” and “higher” should be interpreted in that context.

“p. 5641, first paragraph: discussion is confusing!”

The writing has been revised to try to make it clearer.

“p. 5641: aerosol classification: Why do you not use Angstrom exponents as most of the groups do, why do you still use color ratios.”

Perhaps it is not really necessary to justify this choice. Both are valid intensive parameters that can be presumed to vary with the type but not the amount of aerosol and both variables are correlated with particle size. In our research, a trial of the classification was made using backscatter-related Ångström exponent instead of backscatter color ratio and we concluded that the variable space using the color ratio was slightly better suited to separating types than the Ångström exponent (i.e. using Wilks’ lambda). The difference is no more than a mathematical transform of one axis of the four-dimensional space used. The effect on the classification was not large, but we decided to use backscatter color ratio for this study partly for this reason.

“p. 5642, line 4: scattering ratio: : :. Please define!”

Aerosol scattering ratio is commonly defined as the ratio of the aerosol backscattering divided by the molecular backscattering. Likewise, total scattering ratio is aerosol plus molecular divided by molecular. However, since it was only mentioned in order to say that it isn’t used in the study, it’s simpler to delete the phrase from the revision, and that is what we have done.

“p. 5644: Another example of unsatisfactory literature review: The text suggest that Cattrall, 2005, and Omar 2005 seem to be basic, fundamental papers regarding aerosol typing. Why do you permanently relate your findings to AERONET observations (now including Dubovik, and again Cattrall)? Photometers measure extinction effects, and height integrated as column values. So there is almost no chance to separate different layers, to obtain information of ‘pure’ aerosol types. On the other hand, you present lidar results! You mainly deal with backscatter effects, profiles, why do you not use the huge number of Asian studies (Murayama, Noh, Liu, Franke, : : :.) and European findings (Amiridis, Papayannis, Mona, Mattis, Muller, Gross, and many others in JGR, ACP, Tellus: : :.) for comparison? There is already so much knowledge on aerosol types based on lidar work, that one does no longer need this Cattrall paper.”

Cattrall et al 2005 and Omar et al 2005 are indeed basic, fundamental papers regarding classification *methodology* for a complete set of simple aerosol types, which is what the statement on page 5644 (and the manuscript generally) pertains to. We agree with the reviewer's point that there is a large literature of lidar measurements of aerosol of particular types which improve the understanding of the optical and microphysical characteristics of specific aerosol types beyond what is described by Cattrall and Omar, but they are generally case studies. Also, AERONET measurements are being used as a global aerosol reference database for satellites and other applications. We suppose that if we did not discuss Cattrall et al. and Omar et al., we may have gotten at least as much negative feedback from a reviewer in the satellite field! This is not to say that comparisons to case studies should not be included, but only that they don't supersede the discussion of the methodology papers we have referenced. We have revised the manuscript to include several of these suggested references, as described above. Besides the sentences added to the introduction (already quoted above), some comparisons were added to the sections describing specific types to supplement the comparisons to other lidar measurements that were discussed in the submitted manuscript. For example, section 4.2.2 on the dust and dusty mix types now includes this sentence:

> Observed values of 532-nm aerosol depolarization for this case are about 33% and 532-nm lidar ratio values are 49 ± 9 sr. These lidar ratio values are consistent with lidar ratio values of 53-55 ± 7 given by Tesche et al. (2009b) and also within the range of variability of 38-50 sr observed by Esselborn et al. (2009), both for Saharan dust nearer to the source during SAMUM (Heintzenberg, 2009). Esselborn et al. (2009) show by backtrajectory analysis that their observed variability in lidar ratio is primarily attributable to differences in source regions.

Section 4.2.4 on the urban and biomass burning types now includes this comparison:

> Noh et al. (2009) observed only a slight difference in lidar ratio at these two wavelengths [355 and 532 nm] for smoke and pollution events in Korea for which carbon particle analyzer data were also available, but found somewhat more variation in retrieved single scattering albedo (SSA).

References to Tesche et al. (2009, 2011) have also been added in appropriate spots. Other references to lidar measurements of aerosols of specific types were already included in the submitted manuscript (e.g. Alados-Arboledas, Amiridis, Ansmann, Fiebig, Freudenthaler, Giannakaki, Hoff, Müller, Murayama, Sakai, Shimizu, Sugimoto, Wandinger). Several of the other papers mentioned by the reviewer describe interesting and valuable lidar measurements that are nevertheless not directly relevant to our discussion of aerosol classification and so are not included.

“p. 5645: I would leave out such a chapter on ice particles. The review is again frustrating in view of the huge amount of cirrus lidar papers.”

Section 4.2.1 is not about cirrus. Ice crystal haze is included in the classification due to the high frequency of occurrence in Arctic observations by HSRL. Both AERONET and models represent these ice particles as aerosols and not as cirrus, so identifying these cases is important.

“p. 5646, l12: It is interesting to notice that the Freudenthaler 2009 paper is referenced in the paper, but not here. Exactly here, a citation would be appropriate.”

Agreed. Freudenthaler et al (2009) has been added here and elsewhere where depolarization for “pure dust” is discussed.

“p.5646, l.20: Ice lidar ratios are measured since 1990 (Ansmann, Reichardt, Eloranta, Whiteman, Sakai : : :.) and are of the order of 20-30sr (in the case of off zenith pointing) and can be as low as 2 sr in the case of nadir pointing, and cirrus lidar ratios are thus lower by roughly a factor of 2 and more with respect to dust lidar ratios. That should be clearly stated.”

Revised the text to be more specific about the lidar ratios expected for ice particles.

“p. 5647, l.5, again no reference to Freudenthaler et al.. Asian dust observations are always a problem. I believe that it is almost impossible to observed pure dust over Japan. And, if they measure almost 40%, I would start to think about there efforts to assure calibration quality.”

The Freudenthaler et al. (2009) reference was added here and elsewhere and 40% was revised to 35% in this part of the general discussion.

“p. 5647, l.12, Heese et al., 2009? There are so many better papers on this topic in that SAMUM special issue. Obviously the references are arbitrarily selected. According to Freudenthaler, more than 33% for pure dust at 532nm is not possible.

No, the references are not arbitrarily selected. I made a technical mistake here with my bibliography software. The paper I meant to reference is Heese and Wiegner (2008), not Heese et al. (2009). Measurements showing that dust mixed with something else gives a smaller depolarization than pure dust are common enough that it's hard to imagine exhaustively referencing every measurement that has shown this. However, the Heese and Wiegner paper gives a good example of a measurement of smoke mixed with dust close to the source where they put particular emphasis on the depolarization measurements. It seems like a reasonable reference to use at this point in the paper. I have corrected the mistake and added "e.g." to indicate that we recognize this is not a comprehensive list of references reflecting the observation made in the statement. We have also added a list of references of some papers that derive mixing ratios or microphysical properties of dust mixtures (Léon et al., 2003; Sugimoto and Lee, 2006; Tesche et al., 2009a; Groß et al., 2011; Weinzierl et al., 2011). See below for the exact wording of this note.

Section 4.2.3. See measurements of Gross et al. (SAMUM, Tellus 2011) Esselborn (SAMUM1, Tellus 2009). Separation of dust and smoke was also published in JGR (Tesche et al., 2009, based on SAMUM results).

We don't quite understand this comment. Section 4.2.3 is about maritime and polluted maritime aerosols, not dust or smoke. Separation of specific pairs of aerosol types is not discussed much in this manuscript except for the particularly challenging pairs dust vs. ice crystal haze and smoke vs. urban.

Section 4.2.4. See Petzold et al. (SAMUM 1, Tellus 2009, SAMUM 2, Tellus 2011) on this topic."

It does not appear that Petzold et al. (2009) or Petzold et al. (2011) have anything to say on the topic of the distinguishing biomass burning and urban aerosols using lidar, which is the focus of section 4.2.4. The papers are about urban and dust (2011) and dust alone (2009). We find the separation of urban and biomass burning aerosol to be a particular challenge which is why we have put emphasis on this topic in the manuscript. The broad separation of dust from urban or biomass burning (i.e. separation of cases that do or don't have a significant amount of dust) is relatively straightforward when depolarization measurements are available, so that topic has received much less focus in our manuscript.

Naturally I was not thoroughly familiar with the papers in the September SAMUM II special issue when this manuscript was submitted in August, but we do recognize that the SAMUM I and SAMUM II field campaigns have resulted in a large set of high quality papers pertaining to case studies of dust and dust mixtures, including calculations to estimate optical and microphysical properties of these mixtures. On the other hand, the methodology presented in this manuscript separates cases of many different aerosol types more broadly without attempting to characterize mixtures for specific cases. Specifically we have not presented work in this manuscript dealing with characterizing dust mixture cases by quantifying dust fraction such as, for example, Tesche (2009), Petzold (2011) or Weinzierl (2011) have. The SAMUM papers did not get a lot of emphasis because they did not seem to overlap much with the current manuscript, since these papers refer to more detailed calculations of the characterization of mixtures but are limited to case studies that are dominated by dust, without observations of other pure types. However, to acknowledge these papers and to attempt to clarify the point that the current study is not focused on dust mixtures in the same way, the following sentence was added to section 4.2.2.

Various other studies (e.g. Léon et al., 2003; Sugimoto and Lee, 2006; Tesche et al., 2009a; Groß et al., 2011; Weinzierl et al., 2011) have attempted to characterize the optical and microphysical properties of case studies of mixtures of dust with other species. In the broad classification methodology presented here, "dusty mix" labels a general category which may include cases of dust mixed with a variety of other species.

"Section 4.2.5. Isn't it possible that depolarizing mineral particles are injected too during hot fires? The discussion concerning fresh smoke is not convincing. Relatively low lidar ratios in the case of fresh smoke? That means they do not absorb much, or are fresh smoke particles large, and this causes the drop in the lidar ratio? Please provide an improved argumentation. How is fresh defined? Did you use trajectories to estimate transport time."

The label "fresh smoke" was used because the category was derived from measurement samples in which the B200 with HSRL aboard flew directly over smoke plumes quite close to the source and in which the pilots could see the plume leaving the surface, and also because the measurements are found to differ significantly from other cases where smoke was observed many days and thousands of kilometers away from fires. The types differ primarily in lidar ratio and not much in depolarization, so mineral particles are unlikely to be the primary driver of the observed differences. We would rather not speculate at this time about the mechanisms to explain the observed measurements but instead simply present the observations as an empirical finding of our classification methodology and a topic of future research. Without supporting data we could do no more than speculate.

"Section 4.2: General remark: How can one show an aerosol classification without any airmass transport study to identify the most likely aerosol sources, to study the transport ways (chances to add further aerosols from a variety of sources) and temporal periods (impact of aging on the aerosol properties)?"

The point of our methodology is to leverage the information from relatively few cases where aerosol type can be inferred easily to other cases where the aerosol type is not obvious. In the final manuscript, we have revised the description of the 30 labeled samples to be more specific, to justify why we thought the types in these cases could be inferred (see response to Reviewer #1 for the precise wording of the revision). In some of those cases we did look at backtrajectories. Some other backtrajectories for the example results are provided below in response to more specific comments.

"Figure 2 should include a clear indication of the Mexico City area (city boundaries, down town area)"

Figure 2 has been annotated with a gray bar to indicate the approximate boundaries of the Mexico City metropolitan area along the flight track segment.

“Figure 2: Please provide backward trajectories for the interesting layers, around 4 km height, and the different aerosols below 3.5 km height before and after the black column.”

Please see de Foy et al. (2011) where this example is discussed in more detail including source characterization.

“Figure 3: Two times Caribbean?”

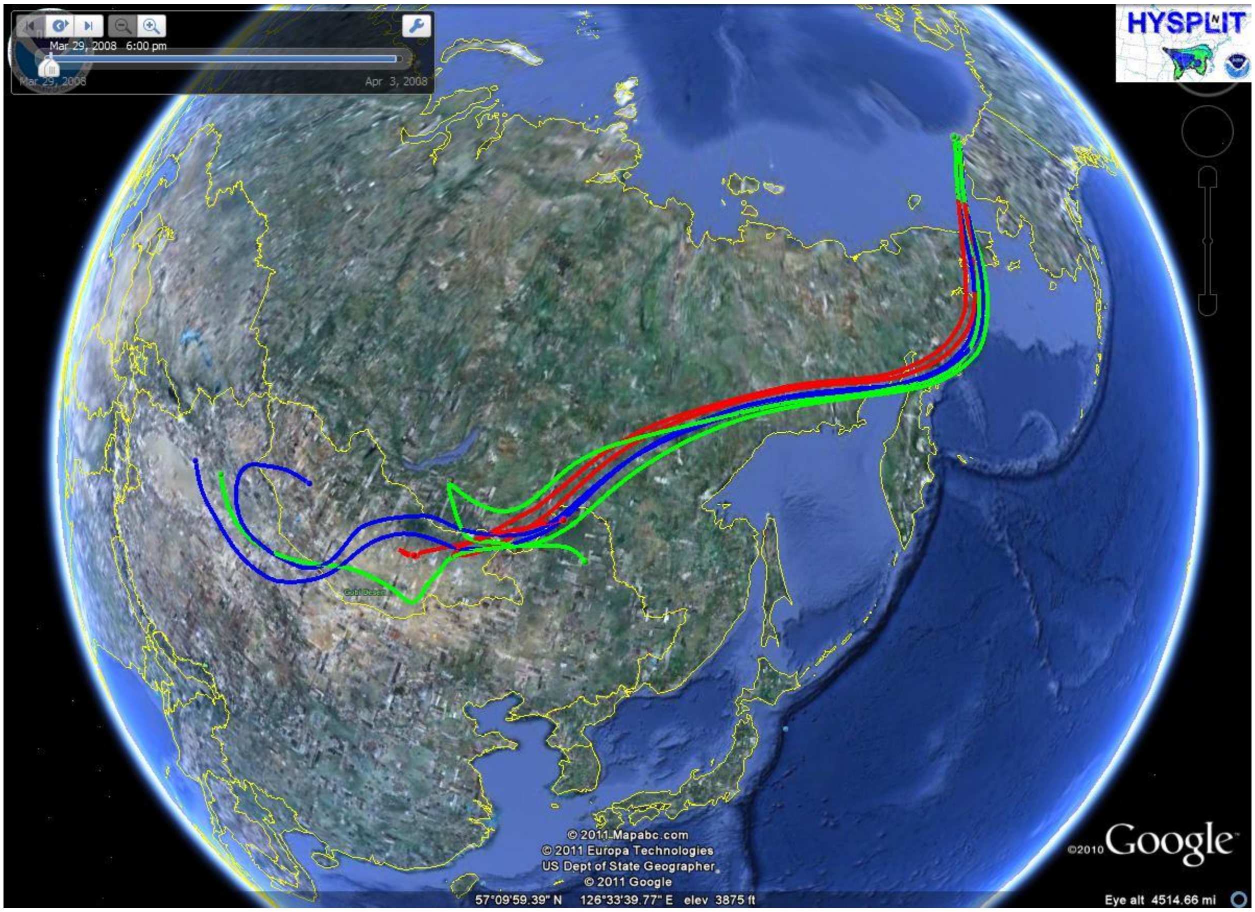
There were two different Caribbean campaigns, as shown in Figure 1 and as discussed in the text. Another sentence was added to the Figure 2 caption to decrease the potential for confusion.

“Figure 4: In our polluted world measurements of pure dust features are almost impossible. Only the SAMUM team may have observed pure dust. Is that included in your analysis?”

Please see Liu et al. (2008). A coherent dust layer was described in this paper that was observed continuously from the source in the Sahara to the Gulf of Mexico 10 days later. The depolarization and lidar ratios for this layer were relatively constant as the dust was advected across most of the Atlantic. As we have seen depolarization values at 532 nm around 33% (which the reviewer proposes is the maximum possible value for dust) in lofted layers in the Caribbean in summertime using the NASA HSRL, it is consistent to call these examples “pure dust.”

“Figure 5: Without showing air mass backward trajectories, all the discussion around Figure 5 is just speculation, and thus not acceptable. Please show backward trajectories to convince me that you really measured dust over Alaska. How can you be sure. May be it is fresh volcanic ash from all the volcanoes in that area. By the way, volcanic aerosol is not considered in your classification scheme.”

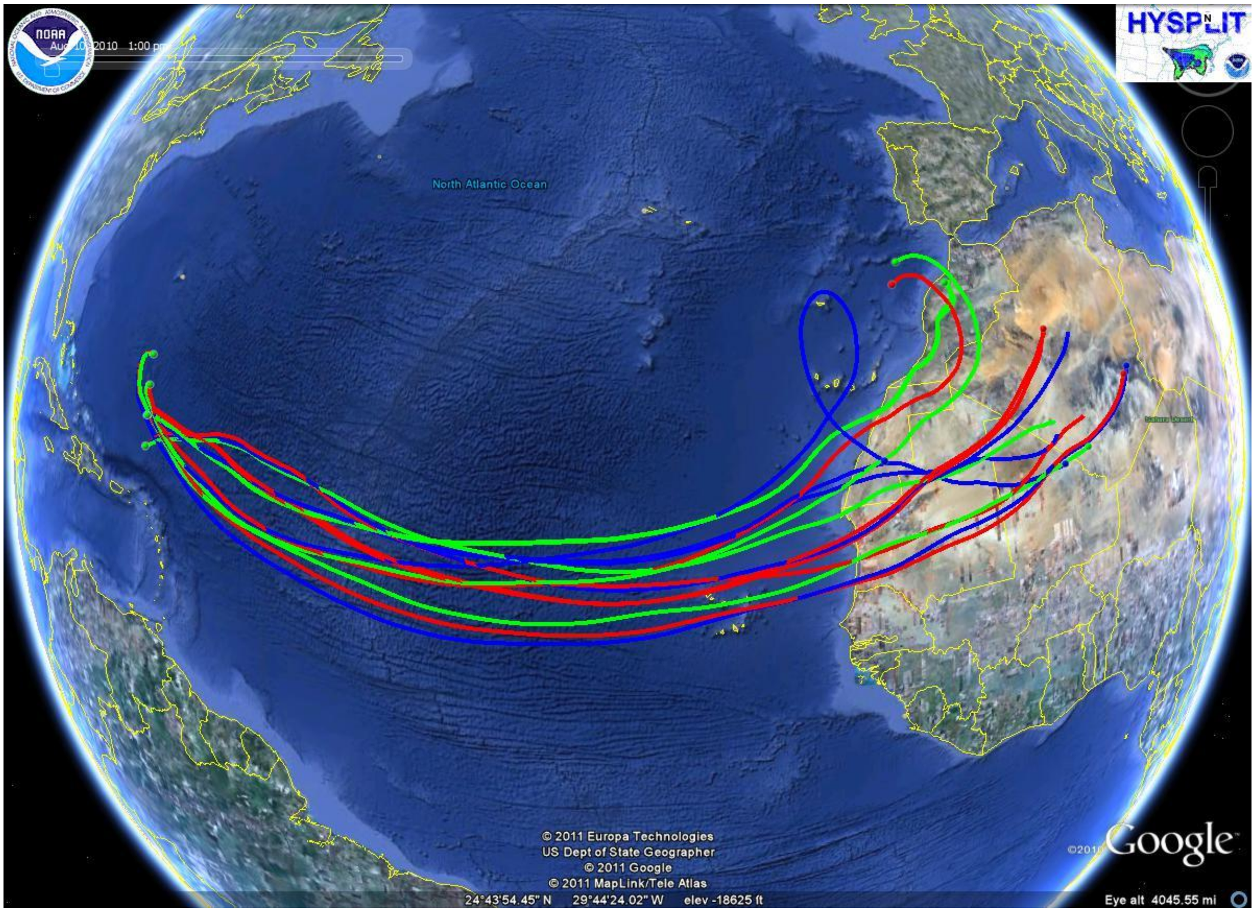
Backtrajectories from HYSPLIT using the NOAA Ready tool (http://ready.arl.noaa.gov/HYSPLIT.php) (Draxler and Rolph, 2011) are consistent with Asian dust in this location, as shown below.



The reviewer is correct that we do not include volcanic aerosol in our classification scheme, as we have never knowingly measured volcanic aerosol with this instrument. In response to Reviewer #3, we stated in the revised manuscript, "The number of classes depends to some extent on the cases where aerosol type is known with high confidence, and should not be considered definitive." So, if our group gets the opportunity to measure volcanic aerosols in the future, we would work on adding that type to our classification.

"Figure 6: The same: A discussion that does not include an airmass transport analysis is useless."

As expected, backtrajectories for this airmass in the Caribbean in August are from Africa, as seen using the online HYSPLIT tool from the NOAA READY website (http://ready.arl.noaa.gov/HYSPLIT.php) (Draxler and Rolph, 2011). The trajectories are given below. We revised the discussion to explicitly state that we have made this check; however, we did not add the figure to the revision for space considerations.



"Figure 8: I would leave out this case."

We disagree. This is one of the types used in our classification and the paper would be incomplete without having an example, especially since the observation of very distinct lidar observables (especially lidar ratio) for different smoke plumes is relatively uncommon in the literature. We acknowledge that there is much we don't know about the explanation for this observation, but we see that as an opportunity for interesting future research.