Atmos. Chem. Phys. Discuss., 12, C12044–C12073, 2013
www.atmos-chem-phys-discuss.net/12/C12044/2013/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "An assessment of CALIOP polar stratospheric cloud composition classification" *by* M. C. Pitts et al.

M. C. Pitts et al.

michael.c.pitts@nasa.gov

Received and published: 28 January 2013

Final Response

Included below are our responses (in bold italics) to the reviews from the two referees and the non-referee comment on our ACPD paper. We thank the reviewers for their insightful comments which have resulted in a much improved manuscript.

Anonymous Referee 1

Major comments:

As introduced above, my main concern with this paper is the conclusions it reaches. The authors write in the abstract that their "classification scheme was found to be excellent in an overall sense", a point that is also made in the conclusion (Sect. 5). In the discussion (Sect. 4), they state that their results "provide compelling evidence that [their] composition classification is robust". Having read the paper, my own impression is that its results cannot lead to such conclusions, but instead provide a useful assessment of unavoidable limitations of lidar-based PSC classification.

We believe the performance of the classification scheme is very good and we have attempted to better substantiate our claim in the revised manuscript. However, we do acknowledge the limitations of lidar-based PSC classification and have attempted to tone down our optimism and avoided using the words "excellent, compelling, and robust."

First, as far as I know, we still don't have a good vision of PSC formation mechanisms, especially for NAT crystals. Given that, it seems to me that it's actually not possible to certify a given classification mechanism "works": we don't really know what we should classify PSC as. For instance, the P09 classification scheme considers five PSC categories (ice, mix2-enh, mix2, mix1, STS), but this number and the categories were really chosen based on the dependence of optical properties seen by lidar on the nature of PSC particles. It is not impossible that PSCs of different nature could produce similar optical signatures. In the other direction, the choice to divide "NAT mix" PSC in "mix1", "mix2", and "mix2-enh" seems rather arbitrary – are these classes really representative of distinct PSC categories, affected by different mechanisms, found in different regions or conditions, etc? My point is that while it is possible to evaluate if a classification scheme works as intended, that does not necessarily make it excellent.

Direct information on PSC composition from polarization sensitive lidar optical measurements like CALIOP's will be limited to particle shape (spherical vs non-spherical). Any attempt to infer additional information on composition requires some assumptions on what possible compositions may exist and then an examination of the potential optical signals that would be observed by a lidar probing these compositions. The most recent literature strongly suggests that PSC particles occur in one of three compositions: liquid (spherical) STS droplets, solid (non-spherical) NAT, and H2O ice (non-spherical). In addition, the non-spherical (NAT and ice) particles are thought to always occur in external mixtures with liquid droplets (either binary or ternary). This is the foundation of our composition classification that consists of three fundamental composition classes: STS, liquid/NAT mixtures, and ice. PSCs of a different composition may produce similar optical signals, but there is no clear evidence from atmospheric measurements that other compositions exist. If it is determined in the future that PSC composition is different than what we've assumed, the classification algorithm will have to be revised appropriately.

We agree that further subdividing the NAT mixture class into three sub-classes was somewhat arbitrary, but this was intended to identify differences in NAT number density/volume that might suggest different particle formation mechanisms or conditions. Indeed, two papers from the ETH Zurich group that are in preparation (Hoyle et al. and Engel et al.) indicate that low number density NAT mixtures can form without pre-existing ice, and that Mix2-enhanced clouds form primarily downstream of ice clouds. Admittedly, however, the division of the NAT mixture class into three categories is somewhat arbitrary.

Our conclusion that the classification algorithm was found to be excellent refers to the performance of the algorithm and not as a validation of the compositions themselves. As the reviewer points out, the lidar measurements cannot tell us the actual composition with certainty.

My second point is that all the evidence presented here for the evaluation of the classification 1) are circumstantial, and 2) do not necessarily support the classification. The evidence the authors rely the most on are the nitric acid uptake diagrams (Fig. 4-5) and the temperature histograms (Fig. 8-10). Regarding the former, the three figures of nitric uptake in NAT mix PSC (Fig. 5b-d) show a rather large number of points aligned with the STS equilibrium curve (Sect. 3), almost half of the points based on a visual estimation. On page 24653, the authors state that "at first glance, the mixture points... may appear to be misclassified STS". Later, they state that "However, this points represent... mixtures... [with] low number density NAT particles... [and] much more numerous liquid droplets". My impression would be that a PSC composed of a low number of NAT particles and "much more numerous" STS droplets really should be classified as STS. The classification might be working perfectly from an "optical signature" point of view, but can we actually conclude anything on the actual PSC composition, if PSC classified as NAT mix can contain either a majority of NAT or STS particles? I came from this figure with a conclusion opposed to that of the authors – that the optical properties of a PSC are insufficient to infer its composition. I expand on this in the minor comments, suggesting ways to turn this into useful results.

We disagree with the suggestion that a NAT mixture PSC should be classified as STS because the liquid particles are much more numerous. First of all, the majority of particles in all NAT mixtures are liquid (either binary aerosol or STS), with the NAT particles being by far the minority in number density, by at least 2 orders of magnitude. This is actually a very important advantage of the polarization sensitive lidar; it can detect the presence of the NAT particles even when majority of the HNO3 is condensed onto the liquid particles. The characteristics and nucleation of NAT particles are still the biggest gaps in our understanding of PSC processes. If we can detect the presence of NAT

particles, then we should definitely identify the PSC as containing NAT. It's not a matter of which particles are in the majority; otherwise all NAT mixtures would be classified as STS.

Second are the temperature distributions of Fig. 9 and 10, which are presented by the authors as "compelling evidence" for the robustness of the classification, a point which they do not justify by any discussion. I agree that each curve seems to peak at the temperatures where we would expect to find the most particles of the expected type, but there is a *lot* of overlap between the temperature ranges covered by all curves. Almost all curves range between Tice and Tice-4K, and peak 1-2K apart (except ice). Given this strong overlap and the proximity of the peaks, for all we know a third of each PSC class could be misclassified and it would produce the same results. Since the NAT mix figures 5b-5d (discussed above) suggest such misclassification does happen often, I do not find these histograms as compelling as the authors. Moreover, this is not necessarily something that can be "fixed" by a "better" classification algorithm - temperature overlap might be unavoidable.

In noting "the *lot* of overlap between the temperature ranges covered by all curves," the reviewer is apparently referring to the histograms presented as a function of T-Tice. When transformed to T-Teq space, the histograms behave just as one would expect based on PSC particle microphysics. The STS and ice histograms each have a single mode that peaks near Teq because both STS droplets and ice crystals grow fast enough to maintain equilibrium with the gas phase. We strongly disagree that non-equilibrium NAT mixtures (where STS dominates the HNO3 uptake) are being systematically misclassified- the classification is not based on which particles are most numerous in the mixture, but rather on whether there is a clear signature of the presence of non-spherical (presumed NAT) particles in a mixture with liquid droplets. The large temperature range and bimodality of the NAT mixtures is an indication of the existence of both non-equilibrium mixtures in which STS dominates the uptake and mixtures closer to NAT equilibrium where the air parcels have been exposed to temperatures below TNAT sufficiently long for the thermodynamically favored, but kinetically limited NAT particles to dominate the uptake of HNO3. In fact we have now added an additional figure (new Fig.5) which corroborates our presumption that non-equilibrium NAT mixtures occur when exposure times are relatively short, but after long exposure times the mixtures will approach NAT equilibrium, producing the bimodal distribution in the uptake plots.

Based on these comments, I don't think the evidence presented in the paper can sustain the level of confidence displayed by the authors in the examples at the top of this section, and I think they should rewrite these sentences to be less conclusive. Unless I am mistaken in my interpretation of Fig. 5b-d, I would suggest to the authors to include a discussion of the points above as well as quantify their effects (see minor comment 24653.20-25 below). Again, I don't think the classification scheme is bad – actually quite the contrary. I think the paper would be stronger if the authors tried to acknowledge the limitations of the classification and document them. Given the classification scheme is currently state-of-the-art, it would provide a useful reference describing the limits of PSC classification based on optical properties alone.

We have modified the discussion of Fig.3 (formerly Fig. 5) to better substantiate our claim that the NAT mixture class is being accurately classified by our scheme. We do recognize that the distinction between Mix1 and Mix2 may not represent actual different particle characteristics or formation mechanisms and have modified the discussion to better reflect this conclusion.

Minor comments

24645.15: I personally agree with the authors that it is important to better understand the composition of PSCs. Out of curiosity, are the authors aware of any reason to believe that our current understanding of PSC composition somehow impedes our ability to predict the future evolution of polar stratospheric ozone?

The largest gaps in our understanding of PSC processes concern NAT particle nucleation and how these particles redistribute NOy, which limits the ability to accurately model denitrification. This is one reason we feel it is important to always identify the presence of NAT particles. In addition, a better understanding of the role of liquid aerosols (binary and ternary) in chlorine activation is needed.

24648.06: Could you provide a reference for the number densities and sizes of ice PSC particles presented here?

The number densities and sizes were taken from the Fueglistaler et al. (2003) paper (see below). We referenced this paper in the P11 ACP paper where we introduced the wave ice class and will cite it again in this paper as well for completeness.

Fueglistaler, S., Buss, S., Luo, B. P., Wernli, H., Flentje, H., Hostetler, C. A., Poole, L. R., Carslaw, K. S., Peter, T., Detailed modelling of mountain wave PSCs, Atmos. Chem. Phys., 3, 697-712, 2003.

24649: MLS retrievals of gas phase HNO3 are deemed unreliable, and are flagged as such, when tropospheric clouds are detected. In absence of evidence to the contrary, I would expect the MLS profiles of gas phase HNO3 to be also affected somehow by the presence of PSC, perhaps making them unreliable. Is there any reason/referenced evidence to believe this is not the case?

Microwave measurements such as MLS are sensitive to ice hydrometeors with diameters of 100's of microns that are present in thick cirrus anvils associated with deep convective systems and cloud-flagging is employed to identify these thick cloud events. However, PSCs have significantly smaller particle sizes and/or mass and do not impact the gas-phase retrievals. Lambert et al. (2012) stated that "microwave gas-phase measurements are insensitive to the presence of PSCs along the line of-sight, unlike solar occultation and infrared measurements, e.g., MIPAS (Wang et al., 2007), which must attempt to separate the effects of aerosol and cloud contamination on the gas-phase measurements."

24649: The horizontal resolutions of MLS and CALIOP data (400-550km vs 5km) mean that a single MLS profile can be used for up to 110 CALIOP profiles. PSC composition can change quite a bit in so many CALIOP profiles. I understand that surrounding MLS profiles were weighted according to the distance to the CALIOP profile, but the inter-profile data variability can't be created out of nothing. I find it quite surprising that this difference in resolution has no impact on the quality of the colocation. It would help if the authors could, for instance, find a way to document the variability in PSC classes across 500 km sections of CALIOP data.

As the reviewer correctly points out, a single MLS profile of HNO3 (and H2O) will be matched with over 100 CALIOP profiles. This issue was not addressed in our original manuscript, and we have now added a discussion in our revised manuscript. The uptake of HNO3 by PSCs observed by MLS will depend on both the spatial extent and composition of the PSC. If the roughly 4-km vertical by 400-km horizontal MLS FOV is not completely cloud-filled, the retrieved HNO3 value will be biased due to the influence of higher background HNO3

in non-cloudy regions of the FOV. Based on a number of sensitivity studies, we concluded that restricting our analyses to scenes in which the MLS FOV is at least 75% filled with PSC observations is a reasonable compromise that reduces this bias while at the same maintains a large ensemble of scenes for analysis. However, there is still uncertainty associated with variability in PSC composition over the scene. To address this issue, we analyzed the subset of scenes in which the MLS FOV was $>$75% filled with PSC observations and at least 2/3 of those observations were of a single PSC composition. The HNO3 uptake plots for this subset of scenes were very similar to those in Fig.3; leading us to the conclusion that the nature of the uptake plots is not an artifact of composition inhomogeneity. Neither the 75% cloudy FOV requirement nor the 2/3 homogeneous requirement changed the nature of the uptake plots from those in our original manuscript. This bolsters our confidence that the basic nature of the uptake plots is a true reflection of PSC growth kinetics and is not an artifact of measurement uncertainties and mismatches in resolution.

24652.29: Here the authors attribute PSC misclassifications to the presence of "negative noise excursions in the observed perpendicular backscatter signal". Why are negative values considered valid for the classification? I can imagine that these excursions happen at a resolution finer than the one used for the classification, leading to biased low average values. If that's the case, the averaged points are probably below the threshold of a reliable signal-to-noise ratio. In the presented classification scheme, is there any filtering done to remove data with low SNR?

The "negative noise excursions" we refer to here correspond to the lower half of the distribution of noise about a mean perpendicular backscatter value, not actual negative values. But indeed negative values of backscatter (both parallel and perpendicular) are valid CALIOP measurements that arise from subtracting

from each profile a DC background derived from the signal acquired at altitudes between 112 km and 97 km, where the laser backscatter signals are negligible. Subtraction of this DC background signal will result in negative-going noise excursions if the actual laser backscatter signal is small. It is critical that these negative values be included in scientific study since any analysis that involves taking some form of average will exhibit a high bias if the negative points are excluded. As an example, for the CALIOP system the theoretical molecular depolarization ratio is approximately 0.00366. If one examined a large amount of CALIOP depolarization measurements from liquid (spherical particle STS) PSC scenes only, the mean value of the measured depolarization ratio should be close to this molecular depolarization value, but due to the small magnitude of the perpendicular backscatter signal there will be a fairly broad distribution of points about this mean, of which approximately half will consist of negative values. If these negative values are disregarded, the calculated mean depolarization ratio from these liquid STS scenes would clearly be biased high and we would have thrown out approximately half of our STS measurements! The occurrence of negative values in the CALIOP Level 1 data and how to properly handle them is addressed in the CALIPSO Data Users Guide (http://www-calipso.larc.nasa.gov/resources/ calipso_users_guide/faq.phpq1_13). It should also be noted that negative values also occur in the Aura MLS gas species data and users are cautioned in the MLS Data Quality document (http://mls.jpl.nasa.gov/data/v3-3_data_quality_document.pdf) that it is critical to include these negative values in scientific study. We don't perform any SNR filtering of the data.

24653.20-25: Regarding the "NAT mix PSC actually contain a majority of STS" problem (major comment). Could you quantify the number of PSCs that follow each branch (STS and NAT equilibrium) in Fig. 5b-5d? Perhaps by counting the number of points on each side of a third curve equidistant of the STS and NAT branches? Even better, could you suggest STS and NAT concentrations (in particle number or mass) for these mixture case? Such values would be useful as they would document how the relative proportion of NAT vs. STS affect the optical properties of a given PSC and its classification in similar schemes.

The number of points that fall along each equilibrium curve in Fig. 5 (new Fig. 3) may be determined from the T-Teq histograms in Fig. 8 (new Fig.7). For Mix1, they are approximately evenly split between the branches. For Mix2, there are about a factor of 2 more in the STS branch. For Mix 2-enh, there are about a factor of 3 more in the STS branch. We can assume that the STS number density is that of the background aerosol population, or approximately 10 per cubic centimeter. The condensed mass on the STS will be dependent on temperature and cannot be accurately estimated from our analyses. The NAT number densities can vary significantly across the CALIOP optical space and within each mixture class. This non-uniqueness of the NAT mixture classes makes it difficult to say anything quantitative about the relative number density (or condensed mass) other than the liquid particles are always going to vastly outnumber the NAT particles.

24657: If I understand correctly, you say the ice arm shifts to the left because there is less HNO3 available for condensation (due to denitrification etc). I don’t understand why the availability of HNO3 would limit the growth of ice particles in ice PSC. Are you suggesting that all ice particles require HNO3 to grow in size? Probably not.

HNO3 does not affect the growth of ice, but is the limiting factor in the growth of STS droplets, hence in the maximum 532-nm scattering ratio (R532) observed in fully developed STS PSCs. Therefore, a reduction in available gas-phase HNO3 will reduce the maximum R532 observed in fully developed STS. Our

optical model assumes that ice forms on these fully developed STS. Therefore, the bottom of the ice “arm” will shift to smaller R532 (larger 1/R532) when the available HNO3 is reduced due to denitrification. We have modified the discussion in the text to make this point more clearly.

24657: I don’t understand from the text how did the cloud population change going from Fig. 12a to Fig. 12b. Is Fig. 12b supposed to be documenting the presence of small NAT crystals? Why are there no more STS clouds in this Figure compared to Fig. 12a ? (i.e. there are no more non-depolarizing particles). I would have expected denitrification to lead to less NAT and more STS, but the figure suggests the opposite. Please enlighten me.

The relevant part of the 2-D optical space shown in Fig.11 (previously Fig. 12) is the ice arm. The overall PSC distribution also changed significantly between June and August, but that is not relevant to this discussion. Early in the Antarctic PSC season, liquid STS is the predominant composition. Once ice PSCs become prevalent, typically in late June, NAT mixtures become much more widespread and the CALIOP 2-D distributions look similar to Fig.11b. Since denitrification typically begins by early July in the Antarctic, we chose June as the ‘normal’ period. Indeed the character of the 2-D distributions changes significantly once ice PSCs occur because the ice PSCs serve as nuclei for subsequent NAT nucleation. We have replaced the previous Fig.12 with a “zoomed-in” version Fig. 11 that focuses on the relevant region of the 2-D space and modified the text appropriately. The main impacts of denitrification are to lower the equilibrium existence temperatures for NAT formation and to restrict the growth of both the liquid and NAT particles. The relative abundance of NAT versus STS will depend on the available NAT nucleation nuclei, although the overall growth of the particles will depend on the temperature histories of

the air parcels.

Anonymous Referee 2
<Major Comments>
1) In the previous paper P09, they introduced the four CALIOP PSC types; STS, Mix 1, Mix 2, and Ice. In paper P11, they added two more types, and classified into six CALIOP PSC types; STS, Mix 1, Mix 2, Mix 2-enhanced, Ice, and Wave ice. This categorization was made purely by optical properties of CALIOP data sets. In this study, they fundamentally followed the six category classification scheme as P11, and added two independent data sets; MLS HNO3 data and GEOS-5 temperature data. The classification of STS, Ice, and Wave ice are very clear, and easy to be understood. However, I felt that the classification between Mix 1, Mix 2, and Mix2-enhanced are not very clear, and rather arbitrary. These classes are all mixture of liquid and NAT PSCs. The only difference is the number density of particles. I felt that more detailed explanation is needed why they categorized into three groups, based on the nature of formation mechanism of NAT PSCs and/or on the nature of external mixture of NAT/STS PSCs.

We agree that further subdividing the NAT mixture class into three sub-classes was somewhat arbitrary, but our intent was to identify differences in NAT number density/volume that might suggest different particle formation mechanisms or conditions. Indeed, two papers from the ETH Zurich group that are in preparation (Hoyle et al. and Engel et al.) indicate that low number density NAT mixtures can form without pre-existing ice, and that Mix2-enhanced clouds primarily form downstream of ice clouds. However, the CALIOP data do not provide any direct information on formation mechanisms. The NAT particle number densities can vary significantly across the optical space of the three CALIOP NAT mixture classes and even within each mixture class. This non-uniqueness of the NAT mixture classes makes it difficult to say anything quantitative about the relative number density on NAT particles for a specific measurement. For the next generation algorithm, we will consider revising the number of mixture classes appropriately.

2) In this study, the authors mentioned two major weak points in the current algorithm for the classification of PSCs. One is the “speckle STS points”. The other is the misclassification of ice PSC as Mix 2-enhanced PSC in the case of severe denitrification. They mentioned that they are going to investigate methods to correct these deficiencies for their next generation algorithm. For the former effect, they mentioned about the cause of the misclassification (noise of the data), and explained a candidate to eliminate the speckle such as by applying a spatial filter. However, for the latter effect (denitrification), authors never mentioned about a possible method to eliminate the misclassification. Are they going to use the co-located MLS HNO3 data and/or GEOS- 5 temperature data for the next generation algorithm? When I looked at the ice PSC branches shown in Fig. 11a and 11b, I felt that inclined boundary would be more appropriate between Mix 2-enhanced and Ice PSCs. I expect that the authors would add more discussion on how to eliminate the denitrification effect in their next generation algorithm.

We chose the vertical ice boundary based on the assumption that ice forms on fully developed STS. However, the slope of the boundary (inclined or vertical) should not have a significant impact on the ice classification. To account for changes in the abundance of gas-phase HNO3 (denitrification) in our next generation algorithm, we will explore methods to best utilize the Aura MLS gas phase measurements which is essentially the only dataset available. One possible option would be to produce a “cloud-free” MLS HNO3 dataset using the CALIOP PSC observations to filter out Aura MLS measurements that may be

impacted by uptake by PSC particles. This "cloud-free" product would be more representative of the total abundance of HNO3 outside of PSCs and could be used to more accurately define the boundaries for the ice and Mix2 PSC regimes in our optical space during periods of denitrification. We have added a brief discussion of this in the paper.

3) Although the authors does not mention about it at all in this paper, there is another wavelength (1064 nm) in CALIOP lidar. If they can combine these two wavelength information, they may be able to derive some information on PSC particle size, which would be a very informative information for the categorization of PSC types. Is there any possibility to use 1064 nm channel data for the classification of PSCs in the next generation algorithm?

As mentioned in P11, the CALIOP 1064-nm channel measurements in the stratosphere are too noisy to use quantitatively, except for the clear robust identification of wave ice PSCs. If the calibration of the 1064-nm channel is improved in the future, we will consider incorporating it into our analyses.

1) Figure 5a: They claimed "The main STS data cluster does not lie at slightly lower temperatures than the reference equilibrium curve, a finding consistent with Lambert et al. (2012)". An explanation of possible cause of this offset would be helpful.

We believe the reviewer meant that "The main STS data cluster does lie at slightly lower temperatures. . ." We know of no physical explanation for why the STS data would be clustered at colder temperatures than the reference equilibrium curve. This result is similar to that found by Lambert et al. (2012) based on analyses of a few weeks of data from the 2008 Antarctic season and may be an indication of a low bias in the GEOS-5 temperature data. Comparisons with COSMIC GPS data for the same time period (Antarctic 2008) show a similar GEOS-5 temperature bias. We have added text attributing the offset to a possible bias in GEOS-5 temperature data.

2) Figure 5e: Why the authors do not show "Wave-ice PSC" case in Figure 5f? If it is because there is no difference between ice and wave-ice PSCs, please mention about it.

Due to the relatively small number of wave ice observations, we opted to combine them with the synoptic ice class in Fig. 3 (previously Fig. 5). Theoretically, all ice PSCs should exhibit similar behavior in terms of HNO3 uptake: as the temperature drops to the frost point, all available HNO3 will have condensed onto the liquid STS and/or NAT particles. So both synoptic and wave-ice observations should correspond to near-zero gas-phase HNO3 values near or below the frost point and the uptake plots for both should look the same. But, as we discuss in the paper, wave-ice PSCs occur on relatively small spatial scales and, as a result, these small-scale features are not well resolved in the lower-resolution MLS measurements and GEOS-5 analyses. Therefore, a wave-ice only panel in Fig.3 would appear very similar to Fig.6a and the discussion of that figure would apply.

3) Is there any possibility to further separate PSC category of "NAT PSC" from Mix 1, Mix 2, and Mix 2 enhanced categories? If you separate PSC points by two groups which are closer to STS and NAT equilibrium curves, within HNO3-(T-Tic) plots in Figure 5, you may be able pull out pure NAT PSC category?

Pure NAT clouds are not likely because the NAT number densities are always

small compared to the number of liquid particles (binary or STS) in the external mixtures. Although the NAT PSCs that occur near the NAT equilibrium curve in Figure 3 (previously Fig. 5) are behaving like "pure NAT" clouds, the NAT number density is still very small and there are many more liquid (mostly binary at this point) aerosols.

Non-Referee Comment

1. A proper description of how particle depolarization ratios (PDR) are determined from CALIOP level 1 data is completely missing in both the present paper and P09. PDR profiling is not trivial (even for ground-based lidars) and the parameter is crucial for the presented PSC composition classification scheme. It is not sufficient to just define the PDR as the ratio of perpendicular to parallel particle backscatter coefficient. It should rather be explained in detail (equations!) how PDR is determined from level 1 attenuated backscatter coefficients. Please elaborate the respective part in the revised version of the manuscript.

P09 describes the procedure for calculating particulate depolarization ratio. Specifically, P09 states that we first apply a first-order correction for molecular and ozone attenuation to CALIOP Level 1 attenuated backscatter coefficients. Then we calculate perpendicular and parallel components of particulate backscatter by subtracting estimates of the perpendicular and parallel components of molecular backscatter from these (molecular and ozone) attenuation-corrected CALIOP backscatter coefficients. The particulate depolarization ratio is then computed as the ratio of the perpendicular and parallel components of particulate backscatter following Cairo et al. (1999). We feel that this process was described adequately in P09 and that there is no need for further description and equations herein. We do note that technically the particulate depolarization ratios still include a component of attenuation due to

the particles themselves.

2. The thresholds applied in the retrieval algorithm are not clear. How can you obtain classified pixels at R=1.25 in your PDR-vs-1/R-plots (e.g., Figs. 1 and 3 of the present paper), when it is stated in P09 that the lowest threshold (for the highest horizontal average of 135 km) is R=1.32? These values are even higher for shorter averaging lengths!

As stated in P09, the threshold value of R=1.32 is a typical mid-season value for the 450-550 K potential temperature layer for the 135-km smoothing scale. Actual thresholds are computed daily and vary slightly from day to day and with altitude (potential temperature layer). The reason that PSC pixels show up at values of R below the 135-km R threshold in Fig. 1 is because they are being identified solely through enhancements in the perpendicular backscatter and thus are identified as containing non-spherical (NAT) particles: Mix1. These mixtures likely consist of relatively low numbers of NAT particles at temperatures too warm for significant development of STS - thus the overall scattering ratio is below our threshold, but there are NAT particles present in sufficient numbers/volume that they produce a detectable enhancement in perpendicular backscatter. There is no requirement that PSCs exhibit an enhancement in both R and perpendicular backscatter, and, in fact, the majority of the Mix1 class is identified solely through enhancements in perpendicular backscatter.

3. The total backscatter ratio is used in the classification together with the perpendicular backscatter coefficient. Why isn't a perpendicular backscatter ratio used? This would seem more intuitive and maybe even less noisy.

The goal of the paper is to assess the P09/P11 detection and composition classification algorithm. P09/11 uses R and perpendicular backscatter coefficient for PSC detection. P09 chose the variables 1/R and particulate depolarization ratio as a basis for composition classification based on historical precedence. Utilizing perpendicular backscatter ratio instead of perpendicular backscatter coefficient would not be less noisy. Although the majority of the noise is likely coming from perpendicular backscatter coefficient, changing to perpendicular backscatter ratio would just change the scale but not the relative noise.

4. How do the authors account for the change in CALIOP measurement resolution at 20.1 km in their retrieval? This is a serious issue that is not addressed in any of the papers that describe the PSC composition classification algorithm. Note that CALIPSO initially retrieves profiles with a vertical/horizontal resolution of 60 m/1000 m and 180 m/1667 m below and above 20.1 km height, respectively (see also the CALIPSO User's Guide at: http://wwwcalipso.larc.nasa.gov/resources/calipso_users_guide/essential_reading/index.phpaltitude_array. In other words, the high-horizontal-resolution 5-km and 15-km classification of PSCs above 20.1 km height is actually based on 3 and 9 independent 180 m x 1667 m measurements, respectively. This introduces a significant amount of noise to the classification (see also next point and the noise-induces misclassification described in Fig. 6). Can you please comment on that?

We are certainly aware of the change in spatial resolution with altitude of the CALIOP Level 1 data products and how ignoring this change in resolution could introduce artifacts into our analyses. In fact, we explicitly discussed this in detail in Pitts et al. (2007), the first in our series of CALIOP PSC papers. The fundamental sampling resolution of CALIOP is 30-m vertical and 333-m horizontal. Due to data downlink limitations, an altitude-dependent on-board

averaging scheme is employed that provides the highest resolution in the troposphere and lowest resolution in the stratosphere. In Pitts et al. (2007) we list the spatial (horizontal and vertical) resolutions of the downlinked CALIOP data as a function of altitude in Table 2. We then discuss at length in Pitts et al. (2007) section 3.1 "Data preparation" how mixing data with different noise characteristics across these resolution boundaries would be inappropriate and therefore our first step is to smooth the Level 1 data onto a common 180-m vertical x 5-km horizontal grid to produce a dataset with consistent resolution and noise characteristics over the entire altitude range of interest. We mention this again in P09 in section 2.1 "Data preparation" where we state "Since the spatial resolution of the Level 1B CALIOP data products is altitude dependent (Winker et al., 2007), we first smooth the data to a uniform grid with 5-km horizontal and 180-m vertical resolution."

5. Depending on the exceedance of threshold values, the authors apply quite significant horizontal averaging of up to 135 km in the classification scheme. This refers to 135 1-km horizontal resolution profiles and 81 1.666 km horizontal resolution profiles below and above 20.1 km, respectively. However, they stick to the original vertical resolution of 180 m without applying vertical smoothing. Why is that? Keeping the original (above 20.1 km height) and 3- bin averaged (below 20.1 km height) resolution is likely to introduce a lot of noise and artifacts to the classification. In fact, the example presented in P09 shows that different PSC compositions (P09, Fig. 10) relate quite well to the different horizontal averaging lengths used in the classification (P09, Fig. 2). Actually, it would be very interesting to see PDR-vs- 1/R-plots according to different horizontal averaging lengths. Introducing vertical smoothing of 5 to 10 height bins (i.e., 900 to 1800 m) should decrease the noise in the lidar data without influencing the contained information. However, the threshold values for PSC classification might have to be adapted to such an approach. Can you please comment on why you don't perform vertical smoothing of the CALIOP data and discuss the implications of your

choice?

The first step in our processing is to put the data onto a uniform grid of 180-m vertical x 5-km horizontal resolution. Each data point on this grid will have similar SNR since it is produced from data with the same fundamental sampling resolution (30-m vertical x 333-m horizontal). We were concerned that averaging the data vertically (beyond that required to establish the common 180-m vertical grid) before detection and classification could actually introduce aliasing in our composition classification. So we decided to perform all the analyses at 180-m vertical resolution. The averaging scales and composition classes shown in Figures 2 and 10 in P09 look related mostly because the vast core of that PSC is ice which is always detected at 5-km resolution. In general, STS and the NAT mixtures are detected at varying resolutions. We agree that we could reduce noise in the classification of NAT mixtures and possibly STS by doing additional vertical smoothing. We will explore the possibility of additional vertical smoothing in our next generation algorithm.

6. It seems odd that the authors don't reject physically meaningless data points in their classification scheme. If points are classified according to their amount of backscatter ratio, perpendicular particle backscatter coefficient, and PDR, than ALL these three quantities should be in a meaningful range. For instance, points with noise-induced negative PDR should not be considered. Wouldn't the occurrence of 'unusual' data imply that there is something wrong with (or at least room for improvement of) the used approach?

We identify PSCs as positive statistical outliers in either R or perpendicular backscatter coefficient (β-perp), so all PSC data points will at least have a meaningful R or β-perp value, depending on which component is an outlier. The

seemingly physically meaningless numbers come from the "other" component: negative measured values of β-perp or negative calculated values of particulate depolarization ratio for PSCs detected as R outliers; and – to a much less extent - calculated values of R$<$1 for PSCs detected as β-perp outliers. These negative values are simply a result of the low SNR of spaceborne lidar measurements. However, it is important to note that negative values of backscatter (both parallel and perpendicular) are indeed valid CALIOP measurements that arise from subtracting from each profile a DC background derived from the signal acquired at altitudes between 112 km and 97 km, where the laser backscatter signals are negligible. Subtraction of this DC background signal will result in negative-going noise excursions if the actual laser backscatter signal is small. It is critical that these negative values be included in scientific study since any analysis that involves taking some form of average will exhibit a high bias if the negative points are excluded. For example, if one examined a large amount of CALIOP depolarization measurements from liquid (spherical particle) STS PSCs only, the mean value of the depolarization would be near the molecular depolarization value of 0.00366, but there will be a fairly broad distribution of points about this mean, of which approximately half will be negative values. If these negative values are disregarded, the calculated mean depolarization ratio from these liquid STS scenes would clearly be biased high and we would throw out half of our STS measurements! The occurrence of negative values in the CALIOP Level 1 data and how to properly handle them is addressed in the CALIPSO Data Users Guide (http://www-calipso.larc.nasa.gov/resources/calipso_ users_guide/faq.phpq1_13).

7. Having addressed the previous concern, vertical smoothing (see previous point) is likely to reduce the amount of points with negative PDR. To test this, I took a look at the example cloud discussed in P09. I did not come up with negative PDR in any of the vertically smoothed profiles within the cloud and I found that the range of PDR is

getting narrower with a maximum at realistic values. Taking a look at the data also showed that vertically unsmoothed data are still quite noisy even after considerable horizontal averaging.

We agree that vertical smoothing may reduce the number of negative points and will consider this option for our next generation algorithm.

8. The particle size distributions that built the foundation for the classification presented in P09 seem to come out of nowhere. While I agree with the results for STS and ICE, I found the results for NAT somewhat confusing. As a person working with aerosol lidar it just seems odd to have something that shows almost no total backscatter (i.e., total backscatter ratio close to unity) but has a high PDR. I believe that, if the comments above are considered in the retrieval and the analysis is screened off noisy signals, the amount of pixels in upper left part of the PDR-vs-1/R-plots and below the PDR < 0 line will decrease quite significantly. What do you think?

The size distributions are described in detail in P09 (i.e., see section 3.1 Optical calculations for liquid-solid PSC mixtures and references therein) and are based on in situ observations of particle size and number density reported in the literature. It is completely plausible for an external mixture of NAT particles and liquid droplets to exist that exhibits essentially no enhancement in R, but a clear enhancement in particulate depolarization ratio. For instance, one possibility is an external mixture of background stratospheric aerosols inside the winter polar vortex (R around 1.05, but no particulate depolarization) with large NAT crystals in low number density (that do not enhance R but had considerable depolarization). The spread in pixels along the PDR axis at high values of 1/R (small R) will decrease with additional vertical averaging, but the points will not disappear. They are real, valid measurements.

9. How is the horizontal averaging done in the classification scheme? Do you use a sliding window (move one profile further and average anew) or neighboring windows (average over profiles without considering the ones contained in the previous average)? If you use a sliding window, does this produce an oversampling of data with long averaging lengths, and hence, to many points classified with this resolution?

Neighboring averaging windows are used.

The second part of this review deals with the combination of CALIPSO PSC data and observations of MLS presented in the present paper.
1. It is stated that MLS provides HNO3-profiles with a vertical and horizontal resolution of 3.5 – 5.5 km and 400 – 550 km, respectively. This is rather coarse and I miss a discussion about how this is addressed in the combination with CALIOP profiles. Given the horizontal (5, 15, 45, or 135 km) and vertical (180 m) resolution used in the CALIOP PSC classification, very different kinds of PSC composition coincide with identical MLS observations. This issue should be addressed before conclusions can be drawn. Please include a respective discussion in the revised version of the manuscript.

As the reviewer correctly points out, the resolution of MLS HNO3 data is much coarser than that of the CALIOP PSC data. This issue was not addressed in our original manuscript, and we have now added a discussion in our revised manuscript. The uptake of HNO3 by PSCs observed by MLS will depend on both the spatial extent and composition of the PSC. If the roughly 4-km vertical by 400-km horizontal MLS FOV is not completely cloud-filled, the retrieved HNO3 value will be biased due to the influence of higher background HNO3 in

non-cloudy regions of the FOV. Based on a number of sensitivity studies, we concluded that restricting our analyses to scenes in which the MLS FOV is at least 75 percent filled with PSC observations is a reasonable compromise that minimizes this bias while at the same maintains a large ensemble of scenes for analysis. However, there is still uncertainty associated with variability in PSC composition over the scene. To address this issue, we analyzed the subset of scenes in which the MLS FOV was $>$75% filled with PSC observations and at least 2/3 of those observations were of a single PSC composition. The HNO3 uptake plots for this subset of scenes were very similar to those in Fig.3; leading us to the conclusion that the nature of the uptake plots is not an artifact of composition inhomogeneity.

2. The same holds for GEOS-5 data. It doesn't seem right to just reject points that are considered as anomalous without further discussion of why the non-anomalous points should be trustworthy! Please comment on that. How do point (1) and (2) affect the outcome of the investigation summarized by Figs. 4, 5, and 7?

With regard to (new) Fig. 3, our sensitivity studies show: (1) Uncertainty and bias due to $<$ 100% cloudiness over the MLS FOV accounts for no more than 1-2 ppbv of the spread in data points along the HNO3 axis; and (2) uncertainty in MLS H2O and GEOS-5 temperatures produces $\sim \pm$ 1K spread in data points along the T-Tice axis. Therefore, neither of these can explain the separation of data points into distinct families along the STS and NAT equilibrium curves, which we interpret as a manifestation of a real difference in PSC growth kinetics.

It would be good to provide some background of how you combine data from a limb sounder with nadir-looking CALIOP. Maybe a sketch of the different viewing geometries would help? Also, no information is given on the accuracy of MLS measurements.

CALIOP nadir and MLS limb tangent tracks are maintained to horizontal separations of less than $\pm$ 10 km and time differences of under a couple of minutes by coordinated A-train satellite operations. The main issue when combining these datasets is the different horizontal and vertical resolutions. We have attempted to better quantify the impact of this mismatch in spatial resolutions in our revised manuscript. The precision of the Aura MLS HNO3 and H2O data were listed in Section 2.2. The accuracy of the v3.3 HNO3 and H2O data is discussed in the MLS data quality document (http://mls.jpl.nasa.gov/data/v3-3_data_quality_document.pdf). For the pressure range appropriate for PSCs (100 – 32 hPa), the H2O accuracy (2 sigma) is estimated to be 0.2-0.4ppmv and HNO3 accuracy (2 sigma) is estimated to be 0.5-1 ppbv.

4. The authors should be careful when they connect the theoretical conditions for formation or existence of a certain PSC composition to the actual measurements. Note that CALIOP observations only give a snapshot of the situation and it's not always clear what happened outside of the field of view and/or before/after the observation. Regarding this fact, it's not necessarily a misclassification of a certain substance is detected at temperatures (data originate from the coarse GEOS-5 data set?) above its existence or formation threshold. I think this deserves a more detailed discussion in the revised version of the paper.

We have not attempted to connect theoretical conditions for PSC formation to our measurements, only temperature existence regimes. The uncertainty and coarser spatial resolution of the GEOS-5 temperature data induces $< \pm$1K spread in the histograms, but we think that "seeing" PSCs systematically – not randomly - at temperatures above their existence threshold is a clear indication of something wrong. The two cases where we observe temperatures that are

systematically too warm for the PSC class and an indication of misclassification are STS "speckle" and wave ice.

Finally, there are some specific comments:
Page 24648, line 14: A discussion about how measurement noise actually affects the classification is actually missing in the manuscript. It doesn't seem sufficient to just refer to things being a little fuzzy. In fact, the comments on the composition-scheme algorithm above should help to shed some light on the effect of data noise.

We have added a more detailed discussion of how noise impacts our composition classification.

Page 24651, line 14: It would be nice to briefly explain at this point why denitrification should impact the composition classification at all.

We have added a brief discussion of why denitrification can impact composition classification here.

Page 24652, line 24: Actually, enhancements in perpendicular backscatter coefficients mark a presence of non-spherical particles. The way you state it implies that all solid particles are nonspherical. This might be okay for PSCs but it's not the general case in the atmosphere.

We have changed the wording to non-spherical.

Page 24653, line 8: Regarding the discussion of speckle, try vertical smoothing as

discussed above. This way, homogeneous regions will appear quite easily.

We will explore additional vertical smoothing in our next generation algorithm.

Page 24654, line 13 and 21: A discussion of the implications of using data on very different scales is totally missing. It would be very interesting if these points would be elaborated (see detailed comments above).

We have added a more detailed discussion of the implications of using data on very different spatial scales to Section 3.

Page 24656, line 18: This is just a statement! Please elaborate on this compelling evidence in a real discussion. How do you account for uncertainties in the modeled temperature profiles?

We have expanded the discussion of the multi-year histograms to better indicate their implications on the robustness of our composition classification scheme. We have also added a discussion of how temperature uncertainty impacts our analyses.

Page 24657, line 14: The discussion in this paragraph seems quit arbitrary. How is this motivated? How did you decide on the chosen latitude bands? Where does the information on HNO3 abundance come from? Also, note that CALIPSO only covers latitudes to 82 °S (not 90 °S as stated in this paragraph or the caption to Fig. 13).

We have revised Figure 12 (new Fig. 11) and the relevant discussion to more

clearly describe the effects of denitrification. As described in the text, we are not referring to actual latitude bands here, but are using equivalent latitude (EqLat). EqLat is a potential vorticity based coordinate system that is commonly used in studies of polar processes, especially in the stratosphere (e.g. Nash et al., 1996). It can be thought of as a vortex-centered coordinate system where 90 degrees equivalent latitude represents the center of the polar vortex. So although CALIPSO only covers geographic latitudes up to 82 degrees, it frequently makes observations at higher EqLats up to 90 degrees.

Nash, E. R., P. A. Newman, J. E. Rosenfield, and M. R. Schoeberl, An objective determination of the polar vortex using Ertel's potential vorticity, J. Geophys. Res., 101 (D5), 9471-9478, 1996.

Page 24659, line 15: How is misclassification accounted for in the classification scheme? I would assume there is a group called unclassified according to the three criteria used for classification. Yet such a group is never mentioned.

Our six composition classes cover the entire PDR and 1/R optical space, so each observation will fall into one of the six classes. There is not a class called "misclassified" in our present algorithm, the performance of which we are assessing in this paper. Technically, the points classified as STS that are clustered around the NAT equilibrium curve are "misclassified."

Fig. 11: I miss a critical discussion of the theory behind these plots/calculations. Where do they come from? Please elaborate.

Figure 10a (previously 11a) is based on the theoretical calculations that were used to produce Figures 5 (NAT mixtures) and 6 (ice) in P09 for nominal conditions of 50 hPa, 10 ppbv HNO3, and 5 ppmv H2O. Figure 10b (previously 11b) shows the same theoretical calculations, but for 5 ppbv HNO3 (as indicated on the figure). We have restricted the range of the 1/R532 axis to [0.6, 0] to match the scale of the new Fig. 11.

Fig. 12: It would be very interesting to see such plots resolved according to the averaging length. Are there any objective criteria for placing the white line or is this a by-eye decision?

The axis of the ice arm is drawn by eye only.

Interactive comment on Atmos. Chem. Phys. Discuss., 12, 24643, 2012.