Some comments on

An assessment of CALIOP polar stratospheric cloud composition classification by Pitts et al., 2012

The manuscript gives a nice example of the combination of different A-Train sensors. It generalizes the work presented by Lambert et al. (2012) for the entire CALIPSO data set and uses independent measurements with MLS for an assessment of the CALIPSO PSC classification scheme presented by Pitts et al. (2009). The latter paper will be referred to as P09 in the following text.

After reading the manuscript, I have quite a number of questions that concern (1) issues of the classification scheme that were not addresses in P09 and (2) the present work of combining CALIPSO PSC data with observations of MLS.

I will start with the issued regarding the **CALIOP PSC composition classification** scheme described in P09 and applied in the present paper. To me, the presented classification scheme is hard to comprehend and evaluate. This is mainly due to a lack of information provided by the authors. It would be nice to see an improved description of the classification scheme in the revised version of the manuscript. My concerns are listed below:

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.
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!
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.
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://www-calipso.larc.nasa.gov/resources/calipso_users_guide/essential_reading/index.php#altitude_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?
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?

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?

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.

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?

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?

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.

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?

3. 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.

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.

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.

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

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 non-spherical. This might be okay for PSCs but it's not the general case in the atmosphere.

Page 24653, line 8: Regarding the discussion of speckle, try vertical smoothing as discussed above. This way, homogeneous regions will appear quite easily.

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).

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?

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).

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.

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

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?