Polar Stratospheric Cloud (PSC) classification using LIDAR measurements from the recent SAGE III Ozone Loss and Validation Experiment (SOLVE).

Melvin Felton and Ali H. Omar

Center for Atmospheric Sciences

Hampton University, Hampton, VA 23668

Phone: 757-7275108 Fax: 757-7275090

Email: Melvin.Felton@hamptonu.edu

Abstract

Lidar measurements from the recent SAGE III Ozone Loss and Validation Experiment (SOLVE) have been used to identify the number of classes of Polar Stratospheric Clouds (PSCs) and their corresponding characteristics. The backscatter lidar, flown aboard a DC8 aircraft, measures profiles of backscatter at 532 nm and 1064 nm, and depolarization at 532 nm. These data along with the color ratio were used to categorize PSCs using a clustering algorithm. For each group or cluster, the central values or medoids describe the optical characteristics of the clouds. For this preliminary study, five such groups were determined with medoid scattering ratios of 1.13, 11.90, 47.68, 71.07, and, 88.96 at 532 nm. The mean quality index of 0.82 for the five groups shows that the clusters are sufficiently distinct from each other.

I. INTRODUCTION

The largest uncertainties in the most recent estimates of climate forcing are due to the forcing effects of aerosols and clouds. There is a need to improve the characterization and classification of aerosols and clouds using their measured optical properties so that the uncertainties of their effects on radiation can be minimized. Understanding these effects can lead to an improved ability to quantify such climate effects as global warming. The physical properties of aerosols and clouds also have an effect on stratospheric chemistry. In particular, chemical reactions that cause ozone depletion take place on the surfaces of clouds and aerosols.

This study examines the properties of polar stratospheric clouds (PSCs). PSCs occur in both polar regions whenever the ambient temperature falls below about 198 K. PSCs lead to increased ozone loss caused by the heterogeneous chemical reactions that occur on their surface [1]. Using a set of polar stratospheric data, this study identifies the number of PSC types and their corresponding characteristics.

The data is from the lidar measurements of the recent SAGE III Ozone Loss and Validation Experiment (SOLVE). SOLVE was a high latitude (Arctic) mission which was conducted over the course of the 1999-2000 Northern hemisphere winter. It was designed to examine the processes which control polar to mid-latitude stratospheric ozone levels and to acquire correlative measurements needed to validate the SAGE III satellite mission.

An aerosol lidar aboard the DC8 aircraft measured profiles of aerosol and cloud backscatter at 532 nm and 1064 nm and aerosol and cloud depolarization at 532 nm. From these measurements, the aerosol and cloud optical properties can be obtained. For instance, information on the relative concentration and spatial distribution of aerosol and cloud particles can be obtained form the backscatter profiles at the two wavelengths. The two wavelength measurements also provide information about particle size [2]. In addition, depolarization measurements from the particles can be used to infer particle shape and therefore phase [3].

Fig. 1 is a SOLVE data plot of aerosol depolarization, color ratio, and backscatter coefficient.

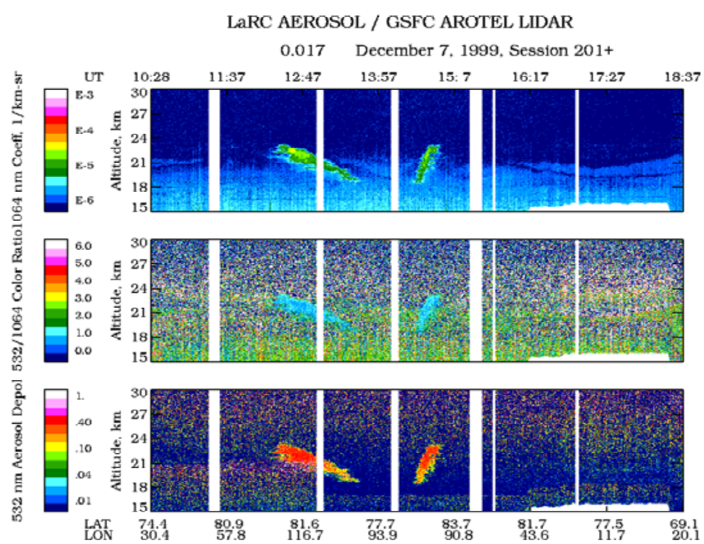

Fig. 1. 532 nm aerosol depolarization, color ratio (532/1064), and 1064 nm backscatter coefficient for the SOLVE measurements of Dec7, 1999.

There has been a need to classify PSCs based on their optical properties and physical characteristics since they were first observed [4]. An analysis of lidar backscatter and depolarization ratios for PSCs revealed five different types [7]. Water ice PSCs were found to have relatively large backscattering and depolarization ratios. A class of clouds that are not depolarizing at neither the visible or near infrared wavelength and with low total backscatter were also observed. These are identified as Type 1b PSCs and are composed of ternary solutions of $H_2SO_4/HNO_3/H_2O$. A third class of clouds that have high depolarization at both lidar wavelengths and relatively low backscattering ratios are named Type 1a PSCs. These clouds are composed of nitric acid tri- or dihydrate (commonly known as NAT). The fourth class of PSCs is depolarizing at visible wavelengths but not near infrared wavelengths. Termed Type 1c PSCs, these clouds are composed of small solid particles. Lastly, clouds that have no depolarization at the lidar's visible wavelength but significant depolarization at the infrared wavelength are likely to be mixtures of Type 1a and 1b PSCs.

II. EXPERIMENT

To minimize the noise effects on the analysis, the SOLVE data was smoothed using a vertical averaging window of 150 m. To ensure that the averaging did not substantially alter the data, percent residuals ($R(z)$) were calculated using

$$R(z) = \left(N(z) - \overline{N(z)}\right) / \left[\left(N(z) + \overline{N(z)}\right)/2\right] \times 100\% ,$$

where N(z) is the original data point and $\overline{N(z)}$ is the averaged data point. The residuals are generally within the range of 10% of the measurements. Fig. 2 and 3 show the averaged and raw data, and a profile of the residuals respectively.

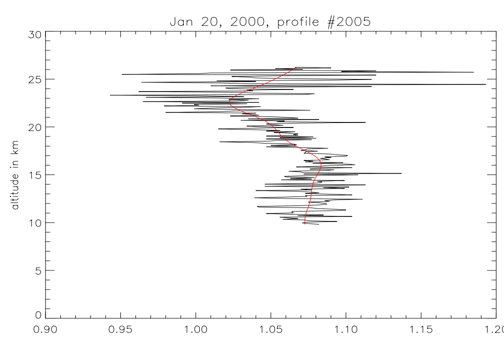

Fig. 2. 532 nm backscatter ratio. Black line represents the raw data and the red line represents the data averaged 150 m in altitude.

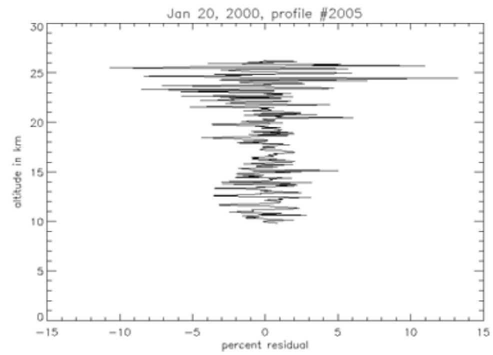

Fig. 3. Percent residuals after applying a vertical averaging window of 150 m.

From the averaged data, the cloud containing measurements were identified for use in the categorization. The criterion used to identify the presence of clouds is a backscatter ratio at 532 nm greater than 1.1, as [5].

The clustering algorithm was used to classify the clouds into mutually unknown groups based on combinations of their principal components. Cluster analysis starts with a data matrix, where objects (measurements of clouds) are rows and observations (optical properties) are columns. Proximity matrices are then constructed where objects are both rows and columns and the numbers in the table are measures of similarity or differences between the observations. This measure of similarity is chosen to be the squared Euclidean distance. To prevent one measure from overwhelming the others, the measures can be transformed to standard scores. The proximity matrices resulting from the squared Euclidean distance that result are then summed to produce a combined distance matrix. Based on these distances, the measurements can be divided into groups.

In particular, a partitioning method has been chosen to perform the cluster analysis. The technique is based on the search for representative objects among the many objects of the data set called medoids [6]. The medoids are calculated such that the total dissimilarity of all objects to their nearest medoid is minimal. Each object of the data set then becomes a member of the cluster corresponding to the nearest medoid. A range of values for the number of clusters (k) desired is required as input for the program. The natural number of clusters can be obtained from analysis of a quality index calculated for each cluster. This index provides an indication of the relationship between the objects of a cluster. A high index implies a well defined cluster while a low index implies a poorly defined cluster. The quality indexes for all k clusters can then be averaged to yield a silhouette coefficient. Through experience an interpretation of this silhouette coefficient has been developed [6] shown in Table 1.

Table 1. Interpretation of the silhouette coefficients of the cluster analysis.

Silhouette Coefficient	Interpretation
0.71-1.00	A strong structure has been found.
0.51-0.71	A reasonable structure has been found.
0.26-0.50	The structure is weak and could be artificial.
≤ 0.25	No substantial structure has been found.

The value of k that yields the highest silhouette coefficient can then be selected as the natural number of clusters for the data set. The groups should be such that the degree of association is strong between members of the same cluster and weak between members of different clusters. Currently, the number of PSC classes are not well understood. There have been reports of up to five distinct groups detected [7]. These studies may yield additional groups with relatively distinct optical and physical properties.

III. RESULTS

Each object used in the clustering algorithm to classify the PSCs consists of backscatter ratio at 532 nm and the corresponding backscatter ratio at 1064 nm, depolarization at 532 nm, aerosol backscatter at 532 nm, aerosol backscatter at 1064 nm, and color ratio such that the backscatter ratio at 532 nm >1.1. For these preliminary study, 1000 measurements of the SOLVE data was used. The number of clusters ranges from two to six and their corresponding silhouette coefficients are shown in Table 2.

Table 2. Silhouette coefficients for the natural number of clusters

Number of Clusters	Silhouette Coefficient
2	0.70
3	0.79
4	0.81
5	0.82
6	0.76

Partitioning the data into five clusters yields the highest silhouette coefficient, thus can be considered as the most natural number of clusters for the data set. The medoids of the clusters provide an indication of the characteristics of the objects within that cluster. Table 3 shows the number of objects in each cluster and the seven medoids of each cluster. To illustrate the division of the clusters, we have plotted the 532 nm and 1064 nm backscatter coefficients in Fig. 5. The plot shows that the five clusters are quite distinct from each other.

Table 3. Number of objects in each cluster and the medoids of each cluster

Cluster	No.of objects	Rs (532 nm)	Rs (1064 nm)	δ
1	722	1.13	2.58	0.041
2	135	47.68	1.95	0.009
3	47	88.96	17.55	0.160
4	65	11.90	62.98	0.153
5	31	71.07	143.25	0.227

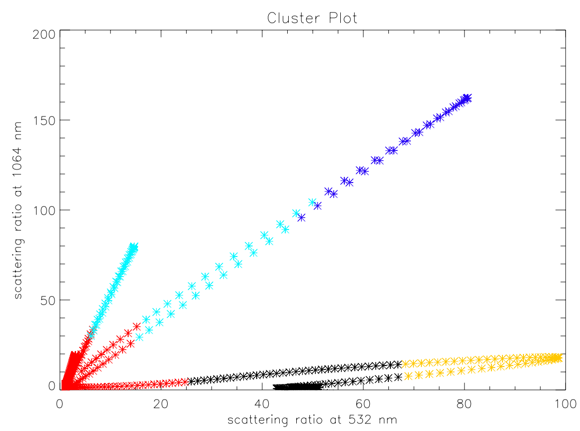

Fig. 5 The 532 nm backscatter ratio plotted against the 1064 nm backscatter ratio for the five clusters: Red=Cluster1, black=Cluster2, yellow=Cluster3, aqua=Cluster4, blue=Cluster5

IV. CONCLUSION

The clustering algorithm successfully classified the limited data set used for this study into five groups. The silhouette coefficient of 0.82 shows that these five groups are quite distinct from each other. This study shows that cluster analysis is an unbiased method that can be used to group PSCs. There still remains significant steps to more accurately perform the analysis. For instance, temperature will be included among the variables. Following the success of this preliminary study, the whole SOLVE data set along with ambient temperature will be used to obtain a more definitive classification.

V. REFERENCES

[1] Biele, J., Tsias, A., Luo, B.P., Carslaw, K.S., Neuber, R., Beyerle, G., and Peter, T., 2001: Nonequilibrium coexistence of solid and liquid particles in Arctic stratospheric clouds. Journal of Geophysical research, VOL. 106, NO. D19, 22,991-23,007.

[2] Gobbi, G.P., 1998: Parameterization of stratospheric aerosol physical properties on the basis of Nd:YAG lidar observations. Appl. Opt., VOL. 37, 4712-4720.

[3] MacKenzie, R.A., Kulmala, M., Laaksonen, A., and Vesala, T., 1995: On the theories of type 1 polar stratospheric cloud formation. Journal of Geophysical Research, VOL. 100, 11,275-11,288.

[4] McCormick, M.P., Steele, H.M., Hamill, P., Chu, W.P., Swissler, T.J., 1982: Polar Stratospheric Cloud Sightings by SAM II. American Meteorological Society, Vol. 39, 1,387- 1,397.

[5] Omar, A.H., Gardner, C.S., 2001: Observations by the Lidar In-Space Technology Experiment (LITE) of high-altitude cirrus clouds over the equator in regions exhibiting extremely cold temperatures. Journal of Geophysical Research, VOL. 106, 1,227-1,236.

[6] Struyf, A., Hubert, M., and Rousseeuw, P., 1996, Clustering in an Object-Oriented Environment, Journal of Statistical Software, 1(Issue 4), 1-30.

[7] Toon, O.B., Tabazadeh, A., Browell, E.V., and Jordan, J., 2000: Analysis of lidar observations of Arctic polar stratospheric clouds during January 1989. Journal of Geophysical Research, VOL. 105, 20,589-20,615.