Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

Application of High-Dimensional Fuzzy K-means Cluster Analysis to CALIOP/CALIPSO Version 4.1 Cloud-Aerosol Discrimination

Shan Zeng[1,2], Mark Vaughan[2], Zhaoyan Liu[2], Charles, Trepte[2], Jayanta Kar[1,2], Ali Omar[2], David Winker[2], Patricia Lucker[1,2], Yongxiang Hu[2], Brian Getzewich[1,2], and Melody Avery[2]

[1] Science Systems and Applications, Inc., Hampton, 23666, USA
[2] NASA Langley Research Centre, Hampton, 23666, USA

Correspondence to: Shan Zeng (shan.zeng@ssaihq.com)

Abstract. This study applies Fuzzy K-Means (FKM) cluster analyses to a subset of the parameters reported in the CALIPSO lidar level 2 data products and compares the clustering results with cloud-aerosol discrimination (CAD) scores reported in the version 4.1 release of the CALIPSO data products. The selection of samples, data training, performance measurements, fuzzy linear discriminants, defuzzification, error propagation, and key parameter analyses in feature type discrimination are discussed. Statistical results show that the FKM classification agrees with the CAD algorithm classification for more than 94% of the cases in troposphere. This is because the lidar-measured signatures of most clouds and aerosols are naturally different. Based on their different natures, objective methods can effectively separate clouds from aerosols in most cases. In addition to validating the current CAD algorithm, the FKM clustering can also provide new insights and supplemental information to help better understand the driving parameters in the scene classification process.

1 Introduction

The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission has been developed through a close and on-going collaboration between NASA Langley Research Center (LaRC) and the French space agency, Centre National D'Etudes Spatial (CNES) (Winker et al., 2010). This mission provides unique measurements to improve our understanding of global radiative effects of clouds and aerosols in the Earth's climate system. The CALIPSO satellite was launched in April 2006, as a part of the A-Train constellation (Stephens and Vane, 2007). The availability of continuous, vertically resolved measurements of the Earth's atmosphere at global scale leads to great improvements in understanding both atmospheric observations and climate models (Konsta et al. 2013; Chepfer et al. 2008).

The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), on-board CALIPSO, is the first satellite-borne polarization-sensitive lidar that specifically measures the vertical distribution of clouds and aerosols along with their microphysical and optical properties. The level 1 CALIOP data products report vertically-resolved total atmospheric backscatter intensity at both

532 nm and 1064 nm, and the component of the 532 nm backscatter that is polarized perpendicular to the laser polarization plane. The level 2 cloud and aerosol products are retrieved from the level 1 data and separately stored into two different file types: the layer product files (Clay, Alay, and MLay) and the profile product files (CPro and APro). The profile data are generated at 5 km horizontal resolution for both clouds and aerosols, with vertical resolutions of 60 m from -0.5 km to 20.2 km, and 180 m from 20.2 km to 30 km. The layer data are generated at 5 km horizontal resolution for aerosols and at three different horizontal resolutions for clouds (1/3 km, 1 km and 5 km). The layer products consist of a sequence of column descriptors (e.g., latitude, longitude, time, etc.) that provide information about the vertical column of atmosphere being evaluated. Each set of column descriptors is associated with a variable number of layer descriptors that report the spatial and optical properties of each layer detected in the column.

The CALIOP level 2 processing system is composed of three modules, which have the general functions of detecting layers, classifying the layers, and performing extinction retrievals. These three modules are the Selective Iterated BoundarY Locator (SIBYL), the Scene Classifier Algorithms (SCA), and the Hybrid Extinction Retrieval Algorithms (HERA) (Winker et al. 2009). The level 2 lidar processing begins with the SIBYL module that operates on a sequence of scenes consisting of segments of level 1 data covering 80 km in along-track distance. The module averages these profiles to horizontal resolutions of 5, 20 and 80 km respectively, and detects features at each of these resolutions (Vaughan et al., 2009). The SCA is composed of three main sub-modules: the cloud and aerosol discrimination (CAD) algorithm (Liu et al., 2004, 2009, 2018), the aerosol subtyping algorithm (Omar et al., 2009; Kim et al., 2018), and the cloud ice-water phase discrimination algorithm (Hu et al., 2009; Avery et al., 2018).

The CAD algorithm uses multi-dimensional probability density functions (PDFs) to distinguish between two classes, cloud and aerosol (Liu et al., 2004, 2009, 2018). The CAD algorithm has been improved over many years, from using three independent measurements (layer mean attenuated backscatter at 532 nm, the layer-integrated 1064 nm to 532 nm volume color ratio, and the mid-layer altitude) in version 1 (V1) to five independent measurements (adding layer-integrated 532 nm volume depolarization and latitude) in version 4.1 (V4). For the operational CAD PDF method, using more measured independent information (i.e., higher dimension PDFs) is expected to yield increasingly accurate cloud and aerosol discrimination.

In addition to the CALIPSO team, scientists over the globe also use the CALIOP data and products, and their work greatly contributes to the evaluation of the CAD products and helps to better understand lidar techniques for distinguishing clouds from aerosols (Chen et al. 2010; Jin et al., 2014; Di Pierro et al. 2011). Using a standard set of lidar measurements ($X_1, X_2, \ldots X_m$), separate multidimensional PDFs are constructed for clouds ($P_{cloud}(X_1, X_2, \ldots X_m)$) and aerosols ($P_{aerosol}(X_1, X_2, \ldots X_m)$). Discrimination between clouds and aerosols for previously unclassified layers is then determined using

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

$$f(X_1, X_2, \ldots, X_m) = \frac{P_{cloud}(X_1,X_2,\ldots,X_m) - P_{aerosol}(X_1,X_2,\ldots,X_m)}{P_{cloud}(X_1,X_2,\ldots,X_m) + P_{aerosol}(X_1,X_2,\ldots,X_m)} k \quad . \qquad (1)$$

The function *f* is a normalized differential probability, which value ranges from -1 to 1, and *k* is a scaling factor that is related to the ratio of the numbers of aerosol layers and cloud layers used to develop the PDFs (Liu et al. 2009). Within the CALIOP level 2 layer products, a percentile (integer) value of 100 × *f*, ranging from -100 to 100, is reported as the "CAD score" characterizing each feature. Aerosol CAD scores range from -100 to 0, and cloud CAD scores range from 0 to 100. The algorithm has been applied to Cloud Physics Lidar (CPL) Data from 2003 THORPEX-PTOST campaign and to desert dust data acquired during the Lidar In-space Technology Experiment (LITE; Winker et al., 1996), and was shown to work well with both data sets (Liu et al., 2004). Because the nature of clouds is quite different from aerosols, distinguishing between the two should generally be straightforward. Transition regions where clouds are embedded in aerosols, volcanic ash injected into the upper troposphere, and optically thick, strongly-scattering aerosols (e.g., haboobs) can still present significant discrimination challenges, but these cases occur relatively infrequently.

The operational CAD algorithm uses manually-derived, multi-dimensional PDFs to distinguish clouds from aerosols using a statistical discrimination function. In this paper, we introduce the Fuzzy K-Means (FKM) method, an unsupervised clustering algorithm, and use it for differentiating clouds from aerosols. The purpose of the study is twofold. First, by using an unsupervised clustering algorithm that is quite different from and independent of the operational method, we can validate the results of V4 CAD algorithm. Second, an unsupervised algorithm can help us better understand classifications made by the operational CAD algorithm. For example, the FKM approach can help determine which individual parameter is most crucial in the discrimination of clouds and aerosols. It can also evaluate the degree of improvement to be expected when adding observational dimensions to the PDFs and help estimate the biases that are coming from background noise.

A description of cluster analysis and the FKM method is given in section 2. In section 3, individual cases and statistical classification results from FKM clustering are produced and compared with classifications made by the operational V3 and V4 CAD algorithms. Error analyses are performed in the following section, including error propagation, key parameter analysis, fuzzy discriminant analysis and principle component analysis. Conclusions and perspectives are given in the last section.

2 Cluster analysis

Cluster analysis is a useful statistical tool to group data into several categories and can be applied to satellite observations to discriminate among different features of interest (Key et al., 1989; Kubat et al. 1998; Omar et al., 2005; Zhang et al., 2007; Usman, 2013; Luo et al., 2017; Gharibzadeh et al., 2018). In our study, we apply the FKM clustering algorithm to CALIOP

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

level 2 observations to distinguish between clouds and aerosols. CALIOP directly observes three quantities: height-resolved profiles of parallel and perpendicular backscatter intensity at 532 nm and total backscatter intensity at 1064 nm. The V4 CALIOP CAD algorithm distinguishes clouds from aerosols using five parameters reported in the level 2 layer products: layer mean attenuated backscatter at 532 nm, layer-integrated depolarization ratio at 532 nm, layer-mean total attenuated color ratio (1064 nm / 532 nm) and layer altitudes and latitudes (Liu et al., 2004, 2009). We use four of them (excluding latitude) as our set of observations and incorporate them directly into a neural network cluster analysis.

There are many different types of neural network clustering methods, such as K-means (KM), fuzzy K-means (FKM), Expectation Maximization (EM) and k-harmonic means (Nock and Nielsen, 2006). In this paper, we focus on the FKM method because the memberships belong to [0 1], and thus are comparable to operational CAD score and more logical for the cloud-aerosol discrimination task. The FKM algorithm is one of a popular class of center-based clustering algorithms. Given a set of observations X (X_1, X_2, …, X_n), where each observation is a p-dimensional real vector, logical clustering aims to partition the n observations into k ($\leq n$) sets $S = \{S_1, S_2, \ldots, S_k\}$ so as to "minimize the within-cluster sum of squares (WCSS) and maximize the between cluster sum of squares (BCSS)" (Hartigan and Wang 1979).

Different from the KM method, FKM clustering gives every point a degree of membership in all categories rather than belonging completely to just one category. Points on the edge of a cluster may be in the cluster to a lesser degree than points in the center of the cluster. The clustering results (i.e., fuzzy memberships organized into a matrix, M , with elements m_{ij}, i=1,…,n; j=1,...,k) all between 0 and 1 (Eq. 2). When elements of the membership matrix are equal to 1, an individual i belongs to a single class j and has a class membership of 0 in all other classes. Note also that m_{ij} can be only 1 or 0 in the KM method, but that intermediate values are permitted in the FKM method. The sum of the fuzzy memberships for an individual over all classes is equal to one (Eq. 3), and there will be at least one individual with some non-zero membership belonging to each class (Eq. 4), written as

$$m_{ij} \in [0,1],\ i = 1, \ldots n; j = 1, \ldots, k, \quad (2)$$

$$\sum_{j=1}^{k} m_{ij} = 1, i = 1, \ldots, n, \text{ and} \quad (3)$$

$$\sum_{i=1}^{n} m_{ij} > 0, j = 1, \ldots k. \quad (4)$$

To determine the best solution, based on minimization of the WCSS, a classic objective function, J, is built so that the best solution is the one that minimizes J (Bezdek, 1981; Bezdek, 1984; McBratney and Moore, 1985). The functional form of J is

$$J(M, C) = \sum_{i=1}^{n} \sum_{j=1}^{k} m_{ij}{}^{\phi} d_{ij}^2(x_{il},)\, c_{jl} \quad (5)$$

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

where C (c_{jl}; j=1,.., k; l =1,..., p) is a matrix of class centers, and $d^2(x_{il}, c_{jl})$ is the squared distance between individual x_{il} and class center c_{jl} according to a chosen definition of distance (i.e., the Mahalanobis distance; see section 2.3). The objective function is the squared error from class centers weighted by the ϕ^{th} power (fuzzy weighting exponent) of the membership values. For the least meaningful value ϕ= 1, J minimizes only at crisp partitions (the memberships converge to either 0 or 1), with no overlap between cluster boundaries. Increasing the value of ϕ tends to degrade memberships towards the fuzziest state where there are more overlaps between the boundaries of clusters. Minimization of objective function J provides the solutions for the best membership matrix M and centroid matrix C (Bezdek, 1981; McBratney and deGruijter, 1992; Minasny and McBratney, 2002). Class centers are the averages of the individual samples weighted by their class membership values raised to the ϕ^{th} function (Eq. 6). The membership (m_{ij}) of an individual belonging to a class is the distance between the individual and the class center divided by the sum of the distances between the individual and the centers of all classes (Eq. 7); i.e.,

$$c_{jl} = \frac{\sum_{i=1}^{n} m_{ij}^{\phi} x_{il}}{\sum_{i=1}^{n} m_{ij}^{\phi}} \quad j = 1,2,\dots k; l = 1,2,\dots p, \text{and} \qquad (6)$$

$$m_{ij} = \frac{d_{ij}^{-2/(\phi-1)}}{\sum_{j=1}^{k} d_{ij}^{-2/(\phi-1)}} \quad i = 1,2,\dots,n; \; j = 1,\dots,k. \qquad (7)$$

To obtain centroid (Eq. 6) and membership (Eq. 7) solutions, Picard iterations (Bezdek et al., 1984) are applied until the centers or memberships are constant to within some small value (Figure 1, the algorithm flowchart). We first initialize the memberships as random values using a uniform distribution that satisfies all conditions given by equations 1, 2 and 3. After that, we calculate class centers and recalculate memberships according to the new centers. If the new memberships do not change compared to the old ones (or change only within a small difference ε), the clustering process ends. Otherwise we recalculate the new centers and new memberships. This process is repeated until the relative change in the objective function (calculated from Eq. 5, which quantifies the changes in both the memberships and centers) is less than ε (0.001).

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

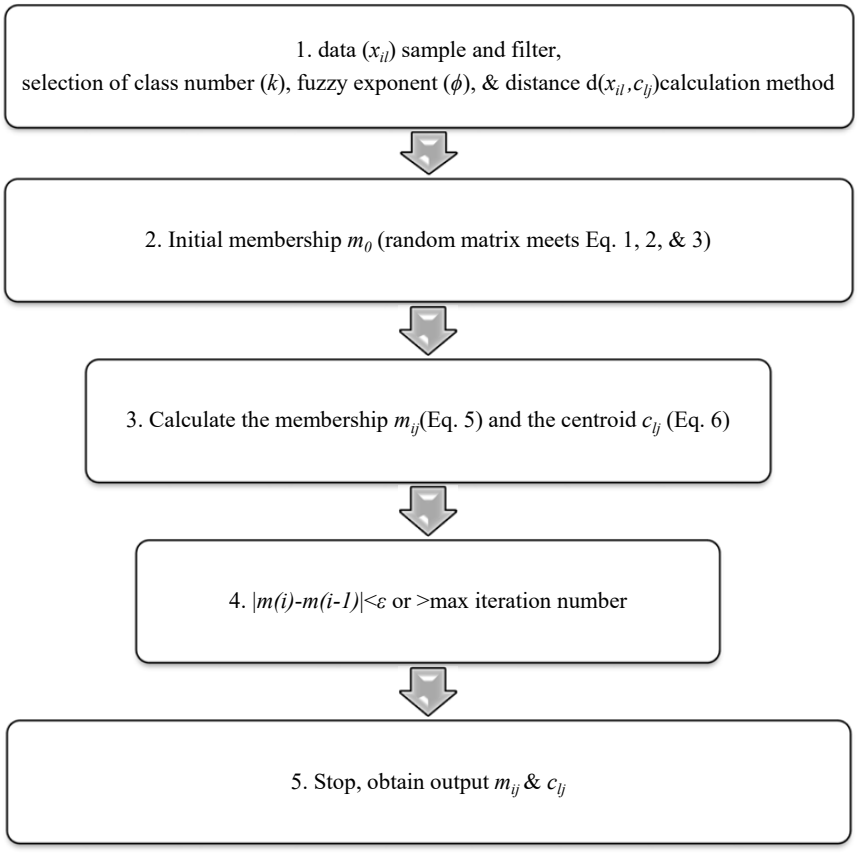


Figure 1: Flowchart of illustrating the operation of the Fuzzy K-means algorithm

Before running the FKM code (from Minasny and McBratney, 2002), we prepared our data by sampling, training and filtering (sections 2.1 and 2.2). We also selected optimal values for class number and fuzzy exponent (section 2.4) and chose a reasonable method to calculate the distance between individuals and centers (section 2.3).

2.1 Data sample and training

As mentioned above, four level 2 parameters are used for our cluster analysis: backscatter at 532 nm, depolarization ratio at 532 nm, total attenuated color ratio (mean 1064 nm divided by mean 532 nm), and mid-layer altitude. The selection of four dimensions is based on many previous studies (e.g., Liu et al., 2004, 2009; Hu et al., 2009; Omar et al., 2009; Burton et al.,

2013), which show that clouds, aerosols and their subtypes are quite different based on these observations. Latitude is not basic information to distinguish cloud from aerosol, but we can apply the FKM method in particular locations or seasons depending on the purpose of the study. In this paper, we apply FKM at a global scale. For any given region, results derived from a localized cluster analysis will likely give us better classifications compared to the results from a global scale analysis, but investigating these differences lies well beyond the scope of this study. The data sample size also strongly influences the clustering results. For example, clustering into two classes with a full complement of CALIPSO data could identify clear and "not clear" scenes. If clear scenes are excluded, clustering could separate clouds and aerosols. If only clear scenes are included, clustering could provide a means of identifying different surface types. With only cloudy data, clustering could be used to derive thermodynamic phase classification. With only aerosol data, clustering is actually aerosol subtyping. With only liquid cloud data, clustering could separate cumulus and stratocumulus. So, the size and composition of the dataset is very important for our analysis, which strongly depends on the objective of the classification.

To extrapolate the classification of identifiable elements using FKM from a small subset to a broader population, we identify an appropriate training data set from which the classifications can be derived (Burrough et al., 2000). This training data should be representative of the broader sample for which the classification will be implemented (i.e., similar domains). To select training data, a PDF for the smaller sample that closely matches the shape of global long-term dataset PDFs can be used to determine an appropriate sample. The training sample is used to determine the optimal number of classes (k) and fuzzy exponent (ϕ) required for classification and optimal values of the performance parameters, and to calculate class centroids for interpretation of similarities and differences between classes. To avoid errors due to small sample sizes, we used one-month of global observations (January 2008) as our 'standard full dataset' for the analysis.

Figure 2 (panels a, b, & c) shows the occurrence frequencies of different lidar observations for liquid water clouds, randomly- and horizontal-oriented ice clouds, and aerosols during all of January 2008. Liquid water clouds have the largest backscatter coefficients and color ratios compared with other species. Aerosols generally have the smallest color ratios, depolarization ratios, and backscatter coefficient, and ice clouds have the largest depolarization ratios compared with the other two species. There are overlaps between species, but these three parameters are still sufficient to separate aerosols and different phases of cloud in most cases. The three bottom panels (d, e, & f) in Figure 2 are from a single half orbit (2008-09-06T01-35-29ZN) of observations. The PDFs of one half orbit and one month of observations appear to agree very well, which means that feature clustering with the FKM method can be trained using a sample as small as one-half orbit of observations and not cause significant biases to the standard full dataset.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

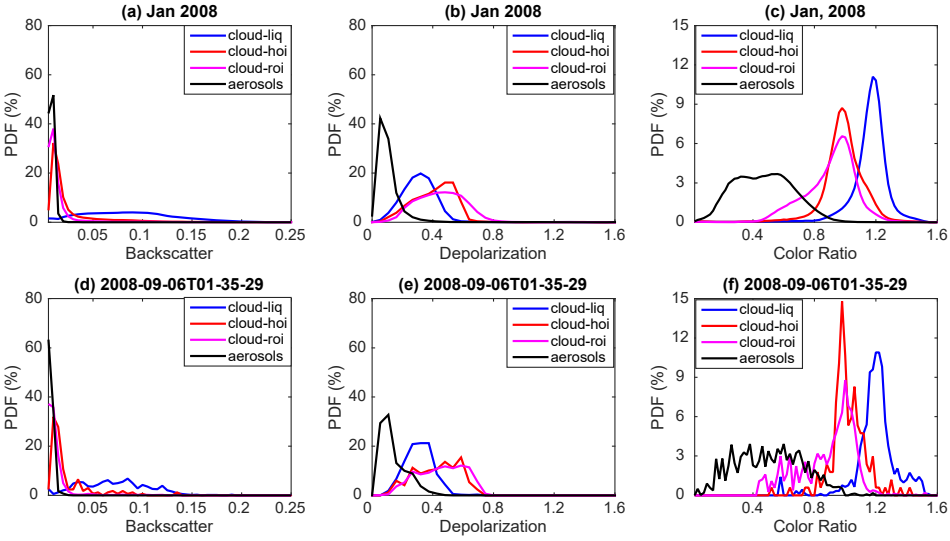


Figure 2: Comparisons of probability density functions; top row shows data from all of January 2008; bottom row shows data from a single half-orbit (06 Sep. 2008, 01:35:29 GMT). The left column (panels a and d) compares attenuated backscatter PDFs; the center column (panels b and e) compares depolarization ratio PDFs; and the left column (panels c and f) compares color ratio PDFs. Black lines represent aerosols, blue lines represent liquid water clouds, red lines represent ice clouds dominated by HOI, and magenta line represent ice clouds dominated by ROI.

2.2 Data filter

We filtered the training data to eliminate outliers in the backscatter, depolarization and color ratio measurements that were either too high or too low. The chosen filter thresholds are summarized in Table 1. The selection of these thresholds is based on the PDFs shown in Figure 2.

Dimension of the satellite observations	Filter criteria
1. Mean Layer Backscatter (AB)	$0 \leq AB \leq 0.2$
2. Layer Integrated Depolarization (DR)	$0 \leq DR \leq 2$
3. Layer Integrated Color ratio (CR)	$0 \leq CR \leq 2$

Table 1: Thresholds for individual satellite measurements.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

2.3 Distance calculation

The distances between attributes can be calculated in different ways (e.g., Euclidean distance, Diagonal distance and Mahalanobis distance). According to a study by Gorsevski et al. (2003), we should apply the Euclidean distance to uncorrelated variables on the same scale when attributes are independent and the clusters have spherical shape clouds. The Diagonal distance is also insensitive to statistically-dependent variables but clusters are not required to have spherically-shaped clouds. The Mahalanobis distance can be used for correlated variables on the same or different scales and is thus the best for FKM CAD analysis. The Mahalanobis distance (d_{ij}) of an observation i from a set of observations (x_{il}) with centers c_{jl} (x_{il} - c_{jl} is an l-dimensional vector) is defined in Eq. 8 (Mahalanobis, 1936).

$$d_{ij}^2 = \left(x_{il} - c_{jl}\right)^T S^{-1}\left(x_{il} - c_{jl}\right);\ \ i = 1, \dots n;\ \ j = 1, \dots k;\ l = 1, \dots, p\ . \qquad (8)$$

S^{-1} (an $l \times l$ matrix) is the inverse of the covariance matrix of the observations. Note superscript T indicates that the vector should be transposed. When the covariance matrix is the identity matrix, the Mahalanobis distance becomes the Euclidean distance. If covariance matrix is a diagonal matrix, the Mahalanobis distance calculation returns the normalized Euclidean distance.

2.4 The choice of class k and fuzzy exponent ϕ

The selection of an optimal number of classes k ($1 < k < n$) and degree of fuzziness ϕ ($\phi > 1$) has been discussed in many previous studies (Bezdek, 1981; Roubens, 1982; McBratney and Moore, 1985; Gorsevski, 2003). The number of classes specified should be meaningful in reality and the partitioning of each class should be stable. For each generated classification, analyses need to be performed and the results validated. Among different validation functions, the fuzzy performance index (FPI) and the modified partition entropy (MPE) are considered two of the most useful indexes among seven examined by Roubens (1982) to evaluate the effects of varying class number. The FPI is defined as in Eq. 9, where F is the partition coefficient calculated from Eq. 10. The MPE is defined as in Eq. 11, with the entropy function (H) calculated from Eq. 12.

The ideal number of continuous and structured classes (k) can be established by simultaneously minimizing both these two measures (FPI and MPE). For the fuzziness exponent, if the value ϕ is too low, classes are discrete and the membership values either approach 0 or 1. But if the value ϕ is too high, the classes will not provide useful discrimination among samples and classification calculations may fail to converge. McBratney and Moore (1985) suggested that the objective function (Eq. 13, Bezdek, 1981) decreases with increasing of both fuzzy exponent (ϕ) and the number of classes (k). They plotted a series of objective functions versus the fuzzy exponent (ϕ) for a given class where the best value of ϕ for that class is at the first maximum of objective function curves (Odeh et al. 1992a, McBratney and Moore 1985). Therefore, choosing an optimal combination of class number (k) and fuzzy exponent (ϕ) is established on the basis of minimizing both values of FPI and MPE and the least maximum of the objective function.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

$$FPI = 1 - \frac{k \times F - 1}{k - 1} \quad (9)$$

$$F = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} m_{ij}^2 \quad (10)$$

$$\mathrm{MPE} = \frac{H}{\log k} \quad (11)$$

$$H = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} m_{ij} \times \log(m_{ij}) \quad (12)$$

$$\frac{\delta J(M,C)}{\delta \phi} = \sum_{i=1}^{n}\sum_{j=1}^{k} m_{ij}^{\phi} \log(m_{ij}) d_{ij}^2 \quad (13)$$

Using one month of "no clear scenes" data, we created Figure 3 to determine optimal values for k and ϕ. From this figure, we concluded the ideal number classes for CALIOP layer classification is either 3 or 4, with corresponding fuzzy exponents equal to 1.4 or 1.6 (we use 1.4 for the analyses this paper). Before exploring the clustering results to see what each class represents, we can immediately confirm that using three classes would be physically meaningful (i.e., these 3 classes may be aerosols, liquid water clouds and ice clouds). Similarly, two classes could represent aerosols and clouds. In the following study, we will choose k equal to 3 or 2 and ϕ equal to 1.4.

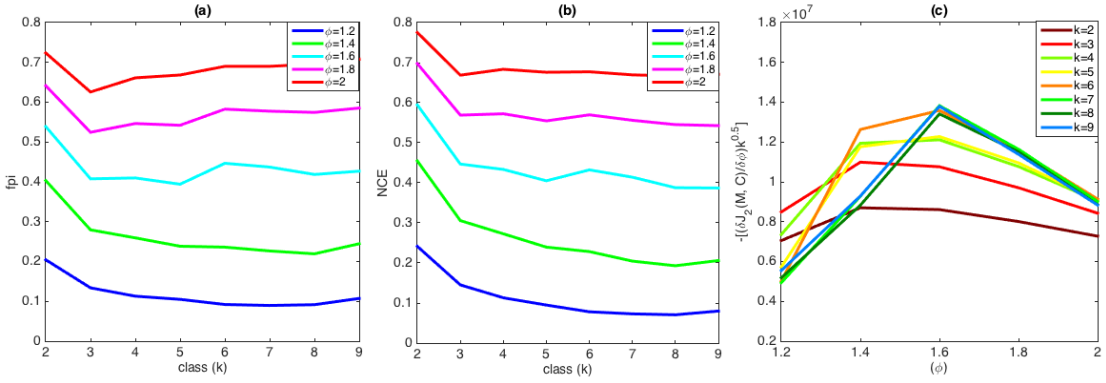


Figure 3: Determination of classes number k and fuzzy exponent ϕ for the FKM cloud and aerosol discrimination algorithm: (a) FPI versus class number k for different values of fuzzy exponent ϕ; (b) MPE versus class number k for different values of fuzzy exponent ϕ; and (c) objective function values versus the fuzzy exponent ϕ for various class numbers.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

3. Cluster results and comparison with V3 and V4 data

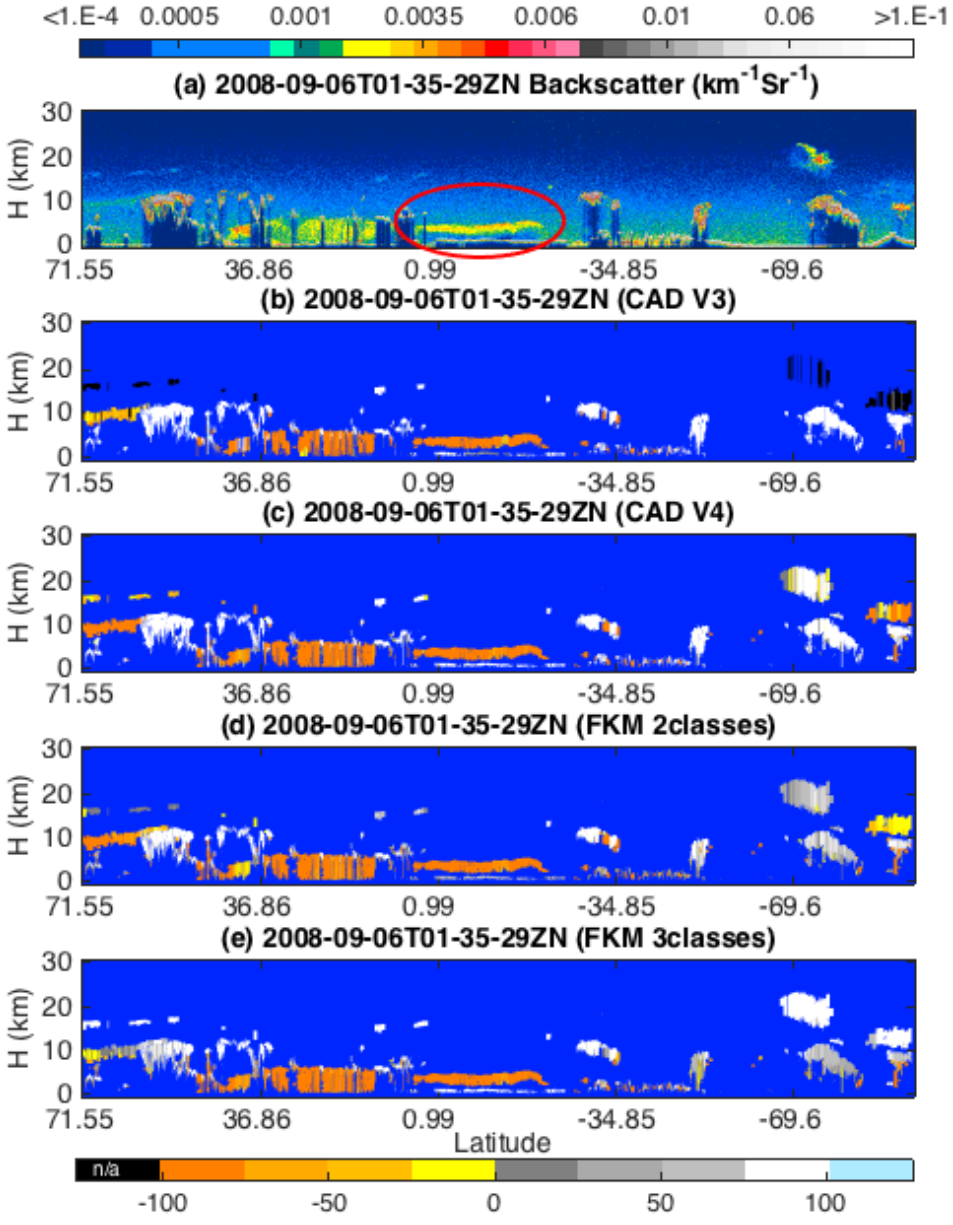
3.1 CAD from the Fuzzy K-Means algorithm

According to Liu et al. (2009), the CAD value for any layer is the difference between the probability of being a cloud and the probability of being an aerosol (Eq. 14). We calculate the fuzzy K-means CAD in a similar way, where the probability is derived from the membership values. For the 3-class FKM analyses, the cloud membership value is the sum of memberships of ice and water clouds (two classes). The fuzzy K-means CAD of clouds and aerosols is found using

$$\mathrm{CAD}_{FKM} = \frac{M_{cloud} - M_{aerosol}}{M_{cloud} + M_{aerosol}} \times 100. \qquad (14)$$

Figure 4 compares the operational V3 and V4 CAD products and our $\mathrm{CAD_{FKM}}$ classifications for a single nighttime orbit segment (06 September 2008, beginning at 01:35:29 GMT). Generally speaking, $\mathrm{CAD_{FKM}}$ from both the 2-class and 3-class analyses are quite similar to both the V3 and V4 operational CAD values. When CAD values are positive (namely clouds, shown in whitish colors in Fig. 4) in V3 and V4, the 2-class and 3-class $\mathrm{CAD_{FKM}}$ values are also positive. Likewise, when CAD values are negative (namely aerosols, yellowish colors in Fig. 4) in the operational data, the 2-class and 3-class $\mathrm{CAD_{FKM}}$ values are also negative. Furthermore, the particular orbit selected here includes some observations (latitudes between 0° and 20°S) of dense smoke over low water clouds. For these water clouds beneath dense smoke, both the V3 operational CAD and the 2-class $\mathrm{CAD_{FKM}}$ label them as clouds with low positive values. On the other hand, the V4 operational CAD and the 3-class $\mathrm{CAD_{FKM}}$ return higher values much closer to 100. The reasons for these differences will be discussed in section 3.2 and 3.4. Note too that weakly scattering edges of cirrus clouds (hereafter, cirrus fringes) around 74°S are misclassified as aerosols by both the 2-class and 3-class $\mathrm{CAD_{FKM}}$ (Figure 4c and d) but are correctly classified as cloud by operational algorithms.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018



Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

Figure 4: nighttime orbit segment from 6 September 2008, beginning at 01:35:29 UTC. The upper panel (a) shows 532 nm attenuated backscatter coefficients. The panels below show the CAD results as determined by (b) the V3 operational CAD algorithm, (c) the V4 operational CAD algorithm, (d) the 2-class FKM CAD algorithm, and (e) the 3-class FKM CAD algorithm. The red ellipse in the upper panel highlights a dense smoke layer lying above an opaque stratus deck. In the CAD images (panels b–e), stratospheric layers are shown in black and cirrus fringes are shown in pale blue.

3.2 Uncertainties: class overlap

The confusion index (CI) is a measure of the degree of class overlap or cluster uncertainty between classes (Burrough and McDonnell 1998). In effect, it measures how well each individual observation has been classified. CI values are calculated from Eq. 15, where m_{max} denotes the biggest membership value and $m_{max\text{-}1}$ is the second biggest membership value for each individual observation (i):

$$\mathrm{CI} = [1 - (m_{max_i} - m_{(max-1)_i})]. \qquad (15)$$

CI value approaches zero when m_{max} is much larger than $m_{max\text{-}1}$, indicating that the observation is more likely to belong to one dominant class. CI approaches one when m_{max} is almost equal to $m_{max\text{-}1}$. In such cases, the difference between the dominant and subdominant classes is negligible, which creates confusion in the classification of that particular observation. Note the value (1- CI) × 100 for the 2-class FKM algorithm is equivalent to $\mathrm{CAD_{FKM}}$ score.

Figure 5 shows CI values for 2-class and 3-class $\mathrm{CAD_{FKM}}$ calculated for all layers in the sample orbit. From the figure, we see that, in most cases, the CI values are low for both the 2-class and 3-class $\mathrm{CAD_{FKM}}$ classifications. The exceptions are stratospheric features (mostly near polar regions), cloud fringes, high altitude aerosols and, for 2-class $\mathrm{CAD_{FKM}}$ only, the liquid water clouds beneath dense smoke. Low CI values for the $\mathrm{CAD_{FKM}}$ classifications are analogous to high CAD scores assigned by the operational CAD algorithm: both indicate high confidence classifications. Similarly, $\mathrm{CAD_{FKM}}$ classifications with high CI values indicate low confidence classifications where the observation has roughly equal membership in two classes. For the liquid water clouds beneath dense smoke, the membership values determined by the 2-class $\mathrm{CAD_{FKM}}$ are larger than 0.5. However, the 3-class $\mathrm{CAD_{FKM}}$ results for these water clouds have low CI values, indicating high confidence classifications into one dominant class, and suggesting that the separation between the aerosols and low water clouds is better accomplished when 3 classes are used. For cloud fringes, the CI values are high for both the 2-class and 3-class $\mathrm{CAD_{FKM}}$. According to the $\mathrm{CAD_{FKM}}$ results, cirrus fringes are essentially different from cirrus and more closely resemble the dust particles that are the predominant sources of ice nuclei (DeMott et al., 2010).

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

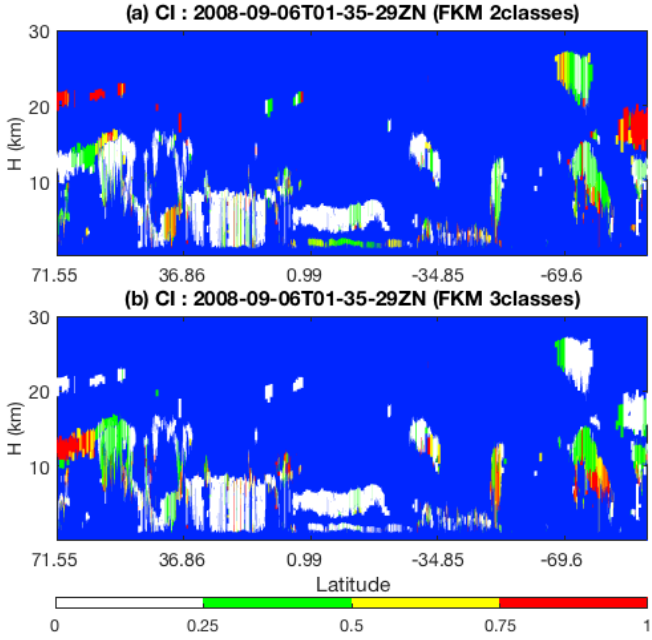


Figure 5: for the same data shown in Figure 4, the upper panel (a) shows the confusion index for 2-class CAD_{FKM}, and the lower panel (b) shows the confusion index for 3-class CAD_{FKM}

3.3 Statistic comparisons of clouds and aerosols

In this section, we first compare the PDFs of the different lidar parameters used in the 2-class and 3-class CAD_{FKM} classifications to the PDFs of those same parameters derived for the operational CAD classifications (Figure 6). It is evident that the PDFs of backscatter coefficient, depolarization ratio and color ratio of clouds and aerosols from the FKM classifications agree well with the PDFs from the V4 CAD classifications. Figures 6d, 6e, and 6f compare the 2-class CAD_{FKM} results (dashed lines) to the operational algorithm (solid lines). In these figures, the PDFs of backscatter (Fig. 6d), depolarization (Fig. 6e) and color ratio (Fig. 6f) of FKM class 1 (blue) agree well with those of V4 cloud PDFs (blue), while the PDFs of these different parameters of FKM class 2 (red) agree well with those of V4 aerosol (red) PDFs. Figures 6a, 6b, and 6c compare the 3-class CAD_{FKM} results to the operational algorithm. Once again, the comparisons are quite good: the PDFs of FKM class 1 (blue) agree well with the V4 water cloud (blue) PDFs, while the PDFs of FKM class 2 (red) and 3 (green) individually agree well with, respectively, the V4 ice cloud (red) and aerosol (green) PDFs.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

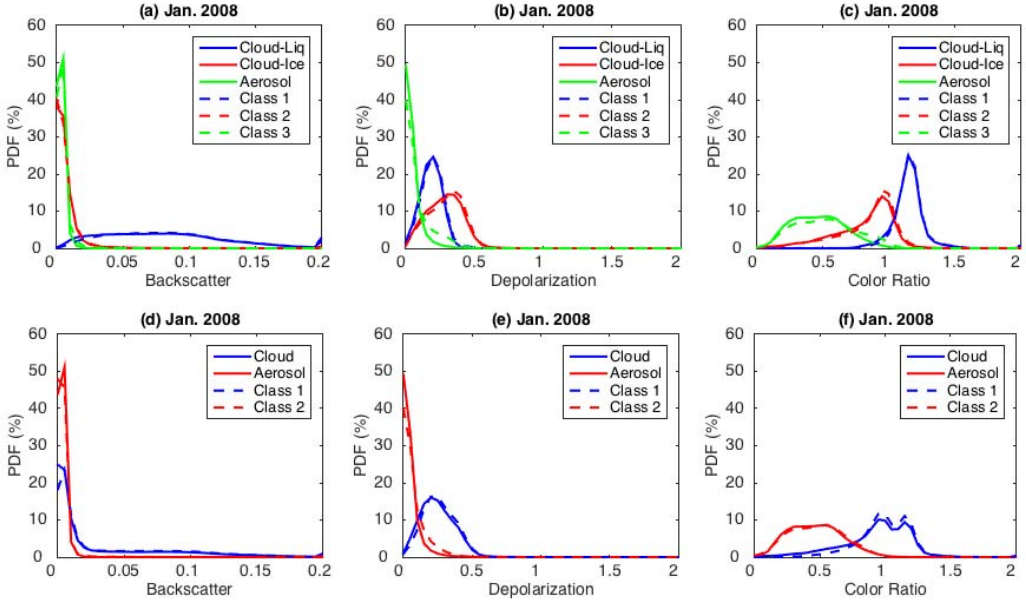


Figure 6: probability density functions derived from all data from January 2008. The top row compares V4 operational CAD PDFs to the PDFs derived from CAD_{FKM} 3-class results. V4 CAD PDFs for liquid water clouds, ice clouds, and aerosols are plotted in, respectively, solid blue, red and green lines. Similarly, CAD_{FKM} 3-class PDFs for classes 1, 2, and three are plotted in, respectively, dashed blue, red and green lines. The bottom row compares V4 operational CAD PDFs to the PDFs derived from CAD_{FKM} 2-class results, where once again the V4 CAD PDFs are shown in solid lines and the CAD_{FKM} 2-class PDFs are shown in dashed lines. Attenuated backscatter PDFs are shown in the left column (panels a and d), depolarization ratio PDFs in the center column (panels b and e), and color ratio PDFs in the right column (panels c and f).

Figure 7 compares one month (January 2008) of geographical (Fig. 7, panels a-f) and zonally-averaged (Fig. 7, panels g-1) distributions of 2-class CAD_{FKM} occurrence frequencies to the cloud and aerosol occurrence frequencies derived from the operational V4 CAD classifications, and the differences between V4 and FKM. The spatial distributions of clouds and aerosols are quite different. In January, clouds are mostly located in the storm tracks, to the east of continents, over the inter-tropical convergence zone (ITCZ) and in polar regions. Aerosols are more often found over the Sahara, over the subtropical oceans, and in south-central and east Asia (upper two rows of Figure 7). In the zonal mean plots (lower two rows of Figure 7), cloud tops are seen to extend up to the local tropopause, whereas aerosols are largely confined to the boundary layer. The geographical and vertical distributions of FKM class 1 are quite similar to the V4 CAD cloud distributions. Likewise, the

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

distributions of FKM class 2 closely resemble the V4 CAD aerosol distributions. Large differences appear in Polar Regions, where the composition of clouds and aerosols is notably different from other regions of the globe (last column of Figure 7). In summary, these geographic analyses show that the cloud-aerosol discrimination derived using an unsupervised FKM method is largely consistent with the classifications produced by the operational V4 CAD algorithm beyond Polar Regions.

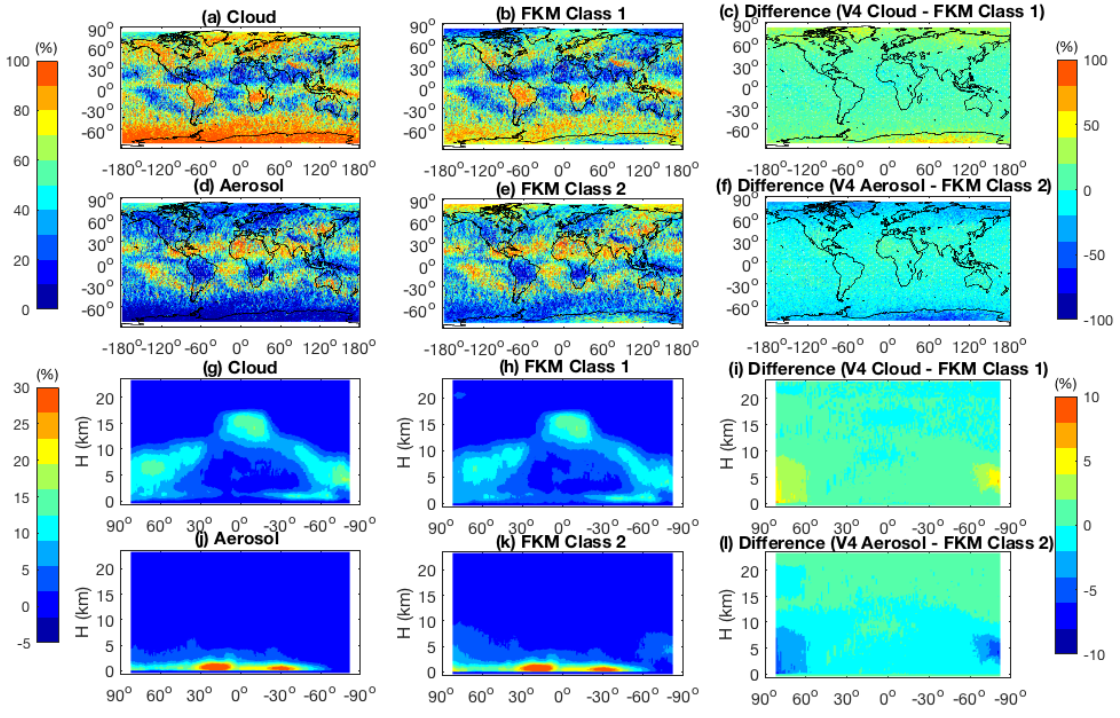


Figure 7: distributions of feature type occurrence frequencies during January 2008. Panels in the left column show V4 operational CAD results; panels in the center column show CAD_{FKM} 2-class results; and the panels in the right column show the absolute differences between the left and center columns. The top two rows show maps of occurrence frequencies as a function of latitude and longitude for clouds (panels a–c) and aerosols (panels d–f). The bottom two rows show zonal mean occurrence frequencies clouds (panels g–i) and aerosols (panels j–l).

To quantify the degree to which the different methods agree with each other, we used the January 2008 5-km Layer data between 60°S and 60°N to calculate the concurrent frequency of cloud and aerosol identifications made by the operational V4 5km CAD algorithm and CAD_{FKM} algorithms. We summarize the occurrence frequency statistics in Table 2. From the table we find that for our test month V3 CAD agrees with V4 CAD for 96.6% of the cases. The FKM 2-class and 3-class results

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

agree with both V3 and V4 for more than 93% of the cases. The FKM results agree slightly better with V3 than with V4. Compared to the 2-class CAD_{FKM}, results from the 3-class CAD_{FKM} agree somewhat better with the classifications from both the V3 and V4 CAD algorithms. Consistent with previous results in this paper, the 3-class CAD_{FKM} appears better able to separate clouds and aerosols than the 2-class CAD_{FKM}. Figure 4 provides an additional example. For those water clouds beneath dense smoke, the 3-class CAD_{FKM} scores are substantially higher than both the 2-class CAD_{FKM} scores and the operation V3 CAD scores, indicating that the 3-class CAD_{FKM} algorithm correctly identifies these features with much higher classification confidence. We also calculated the concurrent occurrence frequencies for only those features with CI values less than 0.75 (or 0.5). When the data are restricted to only relatively high confidence classifications, the FKM results agree with V3 and V4 for better than 96% (or 97%) of the samples tested.

Agreement (%)		**V3**			**V4**			**FKM (2-classes)**			**FKM (3-classes)**		
		C	A	T	C	A	T	C	A	T	C	A	T
V3	C				66.1	2.1		63.2	5.4		65.1	3.2	
	A		-		1.2	30.5		1.0	30.4		1.9	29.8	
	T						96.6			93.6			94.9
V4	C							58.9	5.3		60.9	3.2	
	A		-			-		1.5	34.4		2.5	33.4	
	T									93.2			94.3
FKM (2-classes)	C										63.2	5.1	
	A		-			-			-		1.4	30.4	
	T												93.6

Table 2: Statistical confusion matrix of a 1-month (Jan. 2008) CAD analysis that shows the agreement percentages (detected as clouds: C, aerosols: A, or total of clouds and aerosols: T for both algorithms) between different methods (V3: version 3, V4: version 4, FKM: fuzzy K-means).

3.4 Special cases study

We see from section 3.3 that, statistically, the CAD_{FKM} classifications agree well with the operational V3 and V4 CAD results. In this section, we investigate several challenging cases for which the classifications disagree between V3 and V4 CADs (Liu et al. 2018). In addition to the dense smoke over opaque water cloud case shown in Figure 4, the CAD_{FKM} algorithm, like the operational CAD algorithm, also has difficulty correctly identifying high altitude smoke, dense dust, lofted dust, cirrus fringes, polar stratosphere clouds (PSC) and stratospheric volcanic ash (Figure 8-10). We briefly review each of these cases below.

a. Dust

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

Two different dust cases are selected for this study (Figure 8). One dust case located in Taklamakan desert, is a night-time granule beginning at 20:15:32 UTC on 4 May 2008, shown in Figures 8 a-e. The second case is lofted Asian dust from a night-time granule at beginning at 18:28:54 UCT on the 1 March 2008, shown in Figure 8 f-j. These dust layers exhibit strong backscatter coefficient, high depolarization ratio and color ratios and are frequently misclassified as ice clouds in V3 but correctly classified as aerosol in V4. The CAD_{FKM} classifications for the 2-class and 3-class both agree with version 3 but with low confidence in the dense dust case, suggesting that FKM does not correctly identify the dense dust. But both FKM algorithms agree well with the version 4 CAD in the lofted Asian dust case.

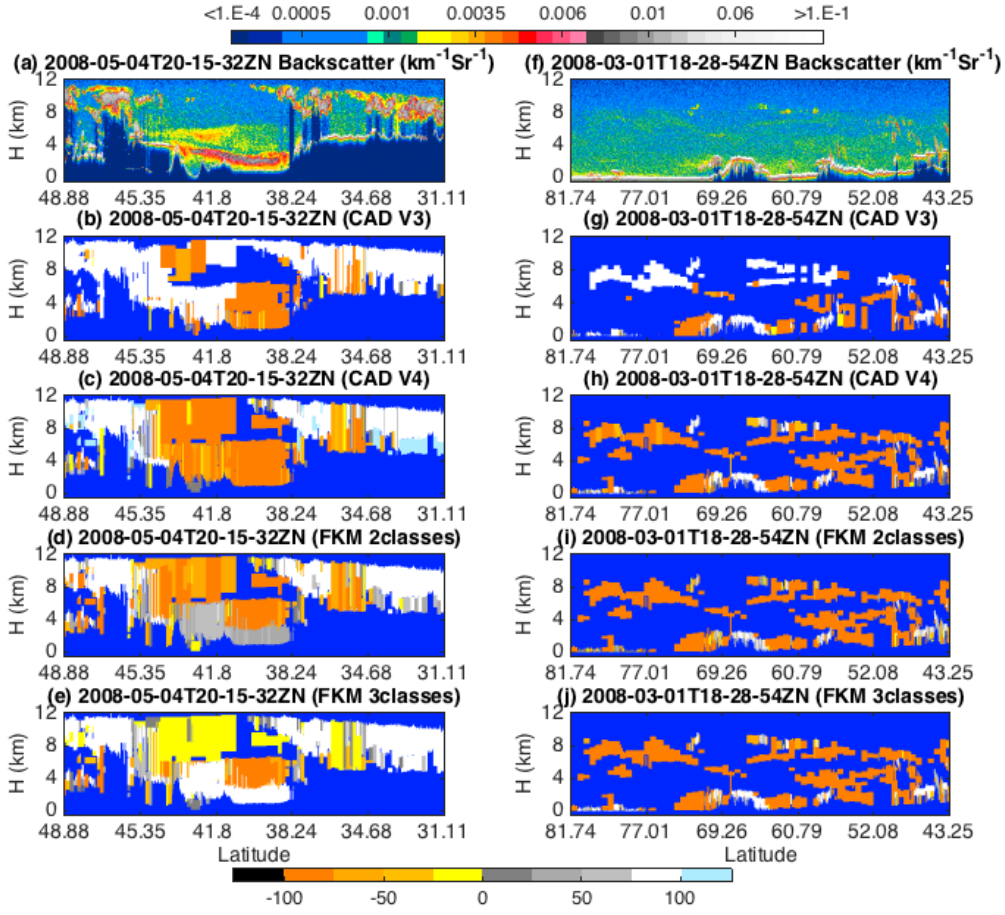


Figure 8: top row shows 532 nm attenuated backscatter coefficients for (a) dust in the Taklamakan basin on 4 May 2008 and (f) lofted Asian dust being transported into the Arctic on 1 March 2008. The rows below show the CAD results reported by four different algorithms; the V3 operational CAD (panels b and g), the V4 operational CAD (panels c and h), the 2-class CAD_{FKM} (panels d and i), and the 3-class CAD_{FKM} (panels e and j).

b. High altitude smoke

An example of high altitude smoke plume was observed by CALIPSO during the "Black Saturday" fires that started 7 February 2009 and quickly spread across the Australian state of Victoria. Figure 9 (panels f-j) shows extensive smoke layers at 10 km and higher on Monday 10 February between 20°S-40°S. The V3 CAD algorithm misclassifies these smoke layers as clouds (or stratospheric features), and so too does the FKM 3-class analysis. On the other hand, the V4 CAD and FKM 2-class analysis correctly identify them as aerosols (not shown here). Without altitude as inputs, both FKM 2-class and 3 class algorithms identify the layers as aerosols (Figure 9 d, e, also see Figure 11). This is because including altitude information can introduce unwanted classification uncertainties when attempting to distinguish between high altitude clouds and aerosols. The reasons for this are discussed in some detail in section 4.1.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

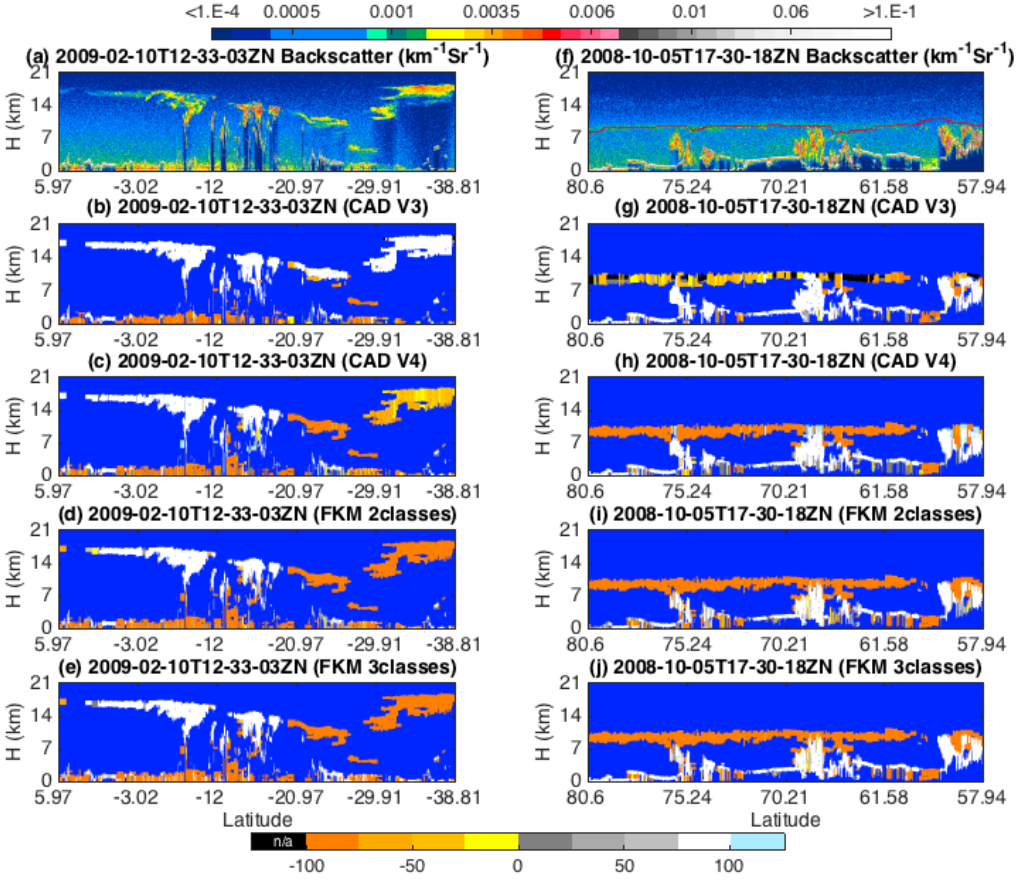


Figure 9: top row shows 532 nm attenuated backscatter coefficients for (a) measurements acquired on 10 February 2009 showing smoke injected into the upper troposphere and lower stratosphere by the Black Saturday fires in Australia and (f) measurements acquired on 5 October 2008 showing a layer of volcanic ash from the eruption of Kasatochi. The rows below show the CAD results reported by four different algorithms; the V3 operational CAD (panels b and g), the V4 operational CAD (panels c and h), the 2-class CAD_{FKM} (panels d and i), and the 3-class CAD_{FKM} (panels e and j).

c. Tropospheric Volcanic Ash

Figure 9 shows an example of ash from the Kasatochi volcano (52.2°N, 175.5°W), which erupted unexpectedly on 7–8 August 2008 in the central Aleutian Islands. Volcanic aerosols remained readily visible in the CALIOP images for over 3 months after the eruption (Prata, et al. 2017). On 5 October 2008, CALIOP observed the ejecta near the tropopause at ~17:30:18 UCT

(Figure 9, panels f-j). The V3 operational CAD algorithm misclassified a substantial portion of this ash plume as cloud, and those segments classified as aerosol were frequently assigned low CAD scores. In contrast, the V4 CAD algorithm and the CAD_{FKM} show greatly reduced cloud classifications, and the aerosols have high confidence CAD scores. Again, when altitude information is not included the FKM algorithm produces a better separation of clouds and aerosols at high altitudes.

d. Fringes

The improved calibration coefficients in V4 facilitated the detection of optically thinner layers than were detected in V3 (Kar et al., 2018; Getzewich et al., 2018). A side effect of this improvement is an increased occurrence of optically-thin layers detected along the tenuous edges of ice clouds. These layers, named "cirrus fringes", are detected at 20 km and 80 km horizontal resolutions, and occur along the sides of ice clouds or along their lower edges where overlying attenuation has substantially reduced the lidar signal. A "cirrus fringe amelioration" algorithm has been added in V4 as a CAD post-processor which detects cirrus fringes that have been misclassified as aerosol and reclassifies them as cloud (Liu et al., 2018). These layers are given a special CAD score of 106. The misclassification of cirrus fringes as aerosols is basically due to their special nature. Fringes are optically thin and weakly scattering layers that occupy the transition zone between cirrus and clear sky. They are characterized by cirrus-like depolarization ratios coupled with lower color ratios, suggesting a reduction in ice crystal sizes possibly due to sublimation.

e. PSC

Polar Stratospheric Cloud (PSC) is the generic name for a class of clouds of several different compositions that all form in the winter polar stratosphere (Höpfner et al., 2009; Pitts et al., 2013). Prior to the V4 data release, all layers having base altitudes above the local tropopause were assigned to a generic class of 'stratospheric feature', and no further subtyping was done for these layers. In V4, however, the operational CAD algorithm now distinguishes between clouds and aerosols in the stratosphere (Liu et al., 2018), and stratospheric aerosols are further evaluated to determine aerosol type (Kim et al., 2018). For PSCs, the CALIPSO project also produces a dedicated PSC data product that reports PSC presence in 5 km horizontal by 180 m vertical bins and classifies each bin according to a composition classification scheme described by Pitts et al. (2018). The compositions in the dedicated PSC product include water ice, supercooled ternary solutions (STS), and several mixtures of liquid droplets and nitric acid trihydrate (NAT). Figure 10 (panels f-j) shows an example of PSC measurements from ~15:25:28 UCT on 15th August 2008. The V4 classifications agree well with the composition classifications developed by Pitts et al. (2009). High confidence CAD scores are given to those PSCs containing water ice particles, while low confidence CAD scores are given to low concentrations of liquid/NAT mixtures. The FKM classifications of the stratospheric features in this case do not agree with the V4 CAD or Pitts composition results. While there are some cases that agree well, these are relatively few and are not shown here. Even when the altitude dimension is excluded from the input parameters, classification of PSCs remains challenging for the FKM algorithm.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

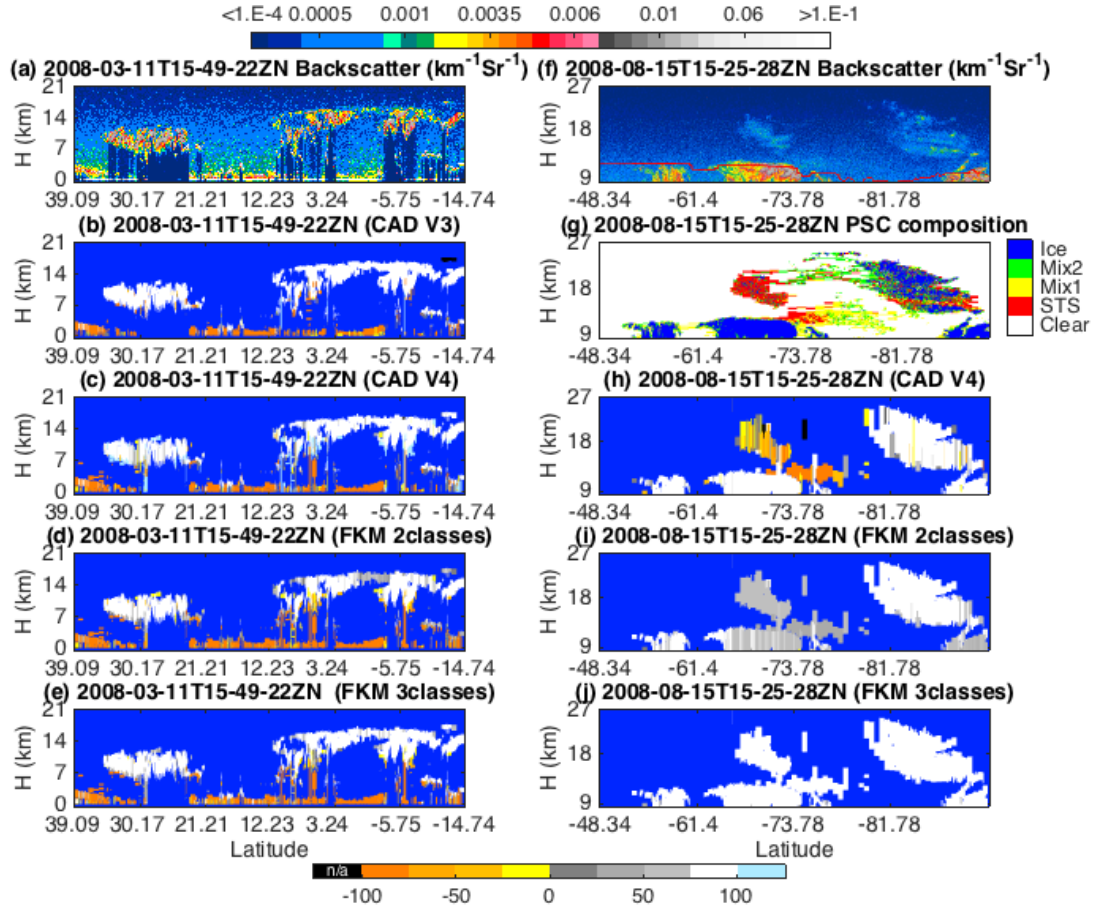


Figure 10: top row shows 532 nm attenuated backscatter coefficients for (panel a) extensive ice clouds measurements acquired on 10 July 2008 and (f) PSCs overlying tropospheric ice clouds on 15 August 2008. The second row shows (panel b) V3 operational CAD results for the 10 July 2008 data and (panel g) the PSC classifications reported in the dedicated CALIPSO PSC data product for the 15 August 2008 data. The three remaining rows show CAD results for the two scenes computed by the V4 operational CAD (panels c and h), the 2-class FKM algorithm (panels d and i), and the 3-class FKM algorithm (panels e and j). The 10 July 2008 results are always in the left column and the 15 August 2008 results are always in the right column. For the PSC classifications in panel g, mix 1 and mix 2 (shown in yellow and green, respectively) represent different mixing ratios of STS and NAT.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

4 Discussion

In section 3, we demonstrated that both the FKM and operational CAD algorithms generate very similar cloud and aerosol classifications. In this section, we discuss key parameter analysis, fuzzy linear discriminant analysis, principle component analysis and error propagation to explore several classification questions: how much improvement can be made by adding additional measurements as classification inputs (section 4.1), how well separated are the current classifications (section 4.2), what are the essential measurements required for accurately discriminating between cloud and aerosol (sections 4.1 and 4.3), and how do the measurement uncertainties (or noise) impact the classifications (section 4.4)?

4.1 Key parameter analysis

CALIOP's operational CAD algorithm was improved from V2 to V3 by adding two additional dimensions (latitude and volume depolarization ratio) to the cloud and aerosol CAD PDFs. Higher dimension PDFs should generally improve the classification accuracy so long as the additional dimensions provide some new information (i.e., they should be orthogonal, or at least semi-orthogonal, to the data already being used). It is therefore important to quantify how much improvement we can make by adding additional dimensions into the analysis. With the FKM method, it is easy to add or remove one or multiple observational dimensions and re-cluster without re-building new PDFs. If one dimension is added (or removed) and the new classifications are similar (or inferior) to the old ones, the added (or removed) dimension does not provide significant new information to the classification (i.e., it is not as important as other dimensions in the classification). If the CAD values are improved (or degraded) by adding (or removing) a dimension, this dimension actually adds (or removes) dispositive information in the determination of the classification, and hence is key to separating clouds from aerosols. Moreover, by using the FKM method we can determine how many dimensions are enough for the cloud and aerosol classification according to our required classification accuracy, either in general or for a particular class (e.g., dust), and how much improvement (or degradation) occurs additional dimensions.

Re-classifications that omit individual dimensions from one orbit of nighttime observations are shown in Figure 11 and a summary of the confusion matrices are shown in Table 3. From both the figure and the table we find that without any one dimension, most new clusters and their CAD_{FKM} values are unchanged compared to classification including the dimension (more than 75% of cases stay same). Note also from the figure we see that, for the low water clouds covered by a plume of heavy absorbing smoke, both V3 and the 2-class FKM have low CAD_{FKM} values. When either mid-layer altitude or backscatter is removed, the CAD_{FKM} values actually improve. Without color ratio or depolarization ratio, the CAD_{FKM} values get worse, which indicates that color ratios and depolarization ratios play a more important role in separating aerosols from low water clouds in this case, as also explained later in section 4.3. In this case, both color ratio and backscatter can bias CAD_{FKM} values due to uncertainties in the measurements related to the strong absorption at 532 nm from smoke above the layer. The attenuated

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

backscatter of these water clouds decreases and gets closer to the backscatter magnitudes expected from classic aerosols (Figure 6a) while color ratio increases and gets far away from those of classic aerosols (Figure 6c).

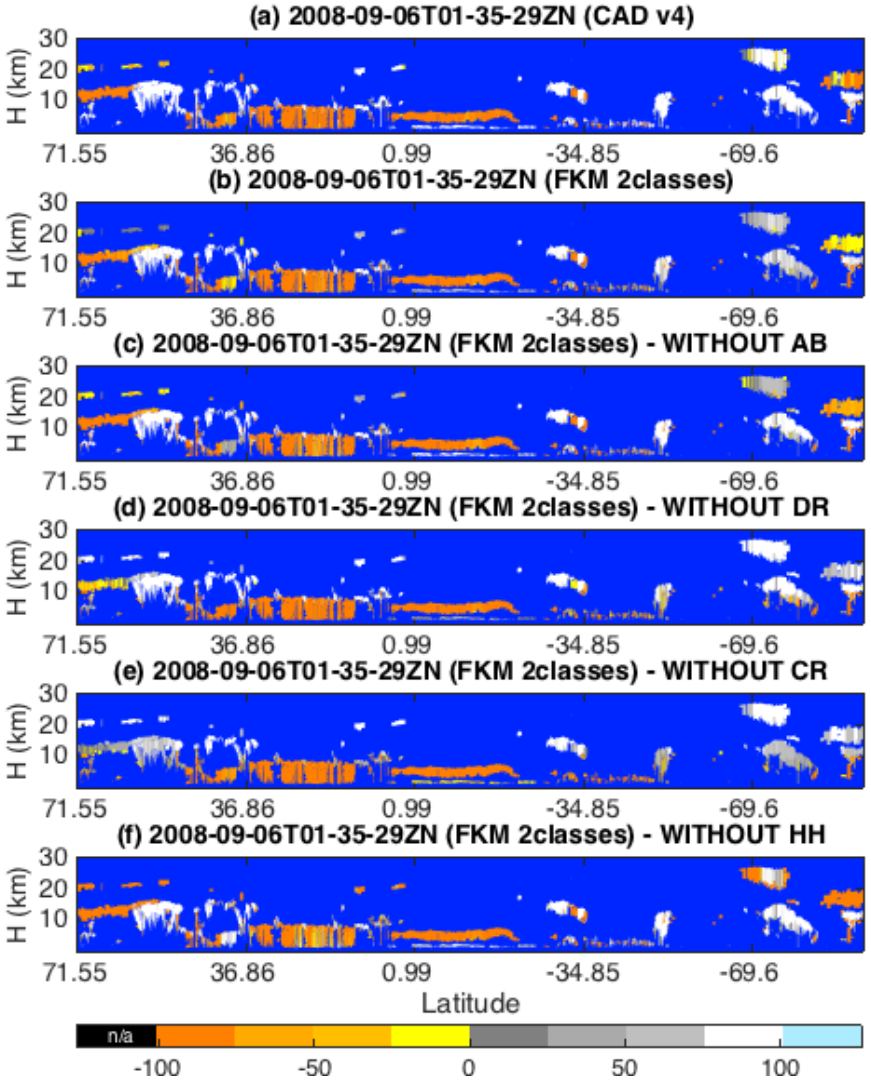


Figure 11: CAD scores calculated using various techniques for an orbit segment on 6 September 2008 beginning at 01:35:29 UTC. The upper two rows show results from (a) the V4 operational CAD algorithm and (b) the 2-class FKM algorithm using

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

all four standard inputs. The remaining rows show 2-class CAD_{FKM} results calculated when omitting one of the four standard inputs: (c) omits backscatter intensity, (d) omits depolarization ratios, (e) omits color ratios, and (f) omits mid-layer height.

(%)		CAD_{FKM}			CAD_{FKM} (no AB)			CAD_{FKM} (no DR)			CAD_{FKM} (no CR)			CAD_{FKM} (no HH)		
		C	A	T	C	A	T	C	A	T	C	A	T	C	A	T
CAD4	C	45.1	3.3		45.6	2.8		43.7	1.7		40.1	11.4		44.9	3.5	
	A	7.3	44.3		7.4	44.2		12.7	38.3		13.6	34.9		12.3	39.6	
	T			89.4			89.8			82.0			75.0			84.5
CAD_{FKM}	C				56.9	2.6		52.3	7.2		49.6	10.0		55.6	3.2	
	A				4.3	36.1		3.7	36.8		12.6	27.9		8.4	32.1	
	T						93.1			89.1			77.4			87.8

Table 3: confusion matrices comparing V4 CAD and the CAD_{FKM} results shown in Figure 11; abbreviations as follows: AB = attenuated backscatter intensity; DR = depolarization ratio; CR = color ratio; H = mid-layer altitude (height); C = cloud; A = aerosol; and T = total.

In addition to the single half-orbit tests, we also analyzed a full month (January 2008) of CALIOP level 2 data acquired between 60°S and 60°N. To better focus on the troposphere, where the vast majority of detectable atmospheric layers occur, data from the polar regions were omitted in this test. We assessed the relative importance of various observational parameters by computing CAD_{FKM} classifications using only a limited number of inputs (i.e., either 1, 2, or 3 of the CALIOP layer descriptors used in the standard CAD_{FKM} classifications). The left panel of figure 12 shows the joint occurrence frequencies of these classifications and the V4 operational CAD classifications. Similarly, the right panel of figure 12 shows joint occurrence frequencies of the limited input classifications and the standard CAD_{FKM} using 4-dimensional (4-D) observations. Two distinct classes are considered, so that Figure 12 is, in effect, a linearized 2 x 2 confusion matrix, with values along the x-axis representing each of the four different types of comparisons. Values plotted for x = 1 indicate the fraction of cases where both the limited input FKM and the V4 operational CAD algorithm (or the 4-D CAD_{FKM}) identified features as being in class 1. The results are color-coded according to the number and type of the dimensions used in the limited input FKM method. Similarly, values plotted for x = 4 indicate the fraction of cases where both algorithms identified features as being in class 2. x =1 and x = 4 correspond to the diagonal elements of the confusion matrix (i.e., M[1,1] and M[2,2]). The off-diagonal elements – i.e., M[1,2] and M[2,1] – are represented by, respectively, x = 2 and x = 3. For x = 2, the limited input FKM identifies the feature as belonging to class 1, while the V4 CAD (or 4-D CAD_{FKM}) identifies it as belonging to class 2. For x = 3 the assignments

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

are reversed: the limited input FKM identifies the feature as belonging to class 2, whereas the V4 CAD (or 4-D CAD_{FKM}) identifies it as class 1.

From Figure 12b it is obvious that, even without any one observational dimension, more than 83% of classifications agree with results from four observational dimensions. Also, we notice that without color ratio, the agreement (only about 83%) between 3-D and 4-D FKM classifications is somewhat worse compared to results when omitting any of the other observations (agreement > 94%). We also find that when using only a single dimension of observations (e.g. backscatter or depolarization ratio or color ratio), FKM classification can only correctly separate the clouds and aerosols for about 60% of the cases compared to using 4-D CAD_{FKM} and/or the V4 operational CAD. This means that the additional 3 dimensions improve the cloud and aerosol discriminations by 30~40%. The more independent measurements that are used, the more accurate the classification can be. From Figure 12a, we again see that with 4-dimensional observations, the agreement between the V4 CAD and the FKM method is ~94%. This 6% differences may come from, for example, very thin, broken clouds, cloud fringes and dense aerosols that are inherently difficult to separate. When different combinations of observations are used in the classification, the disagreement between the two methods are different. V2 CAD algorithm used backscatter intensity, color ratio and mid-layer altitude, which is the best combination of three independent parameters, showing only 0.4% fewer agreements compared to using four parameters. To further improve the CALIOP CAD algorithm, multiple investigations into additional combinations of observations or the use of weighted observations could be pursued in the future.

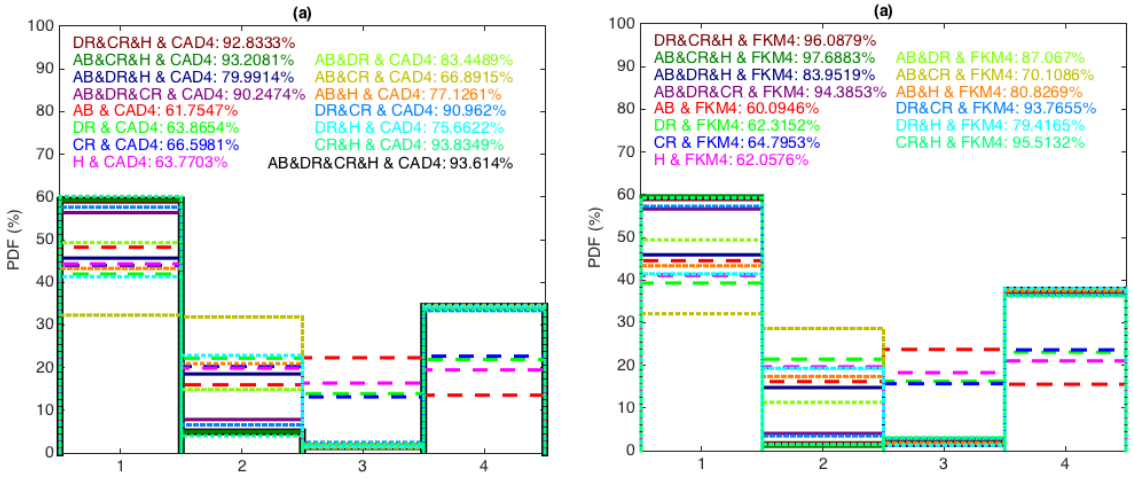


Figure 12: statistics of occurrence frequency during January 2008, between 60°S and 60°N. The left panel (a) shows the joint occurrence of the V4 operational CAD classifications and the FKM classifications based on limited input parameter sets (i.e.,

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

1, 2, or 3 CALIOP measurements). The right panel (b) shows the joint occurrence of the 4-D CAD_{FKM} classifications and the limited input FKM classifications.

4.2 Fuzzy linear discriminant analysis

Linear discriminant analysis (Fisher, 1936) is usually performed to investigate differences among multivariate classes, to validate the classification quality, and to determine which attributes most efficiently contribute to the classifications. Here we introduce Wilks' lambda, which is the ratio of within-class variance (to evaluate the dispersion within class) and between-class variance (to examine the differences between the classes). Considering a data matrix X (elements x_{il}, i=1,..,n; l = 1,..p), the FKM classification returns a membership matrix M (elements m_{ij}, i = 1,..,n; j = 1,..,k) and centroid matrix C (elements c_{jl}, j = 1,…,k; l = 1,…,p) where n is the number of samples, p is the number of attributes, and k is the number of classes. The sums of squares and products (SSP) within-classes covariance matrix W_{lm}, also called the within-classes fuzzy scatter matrix (Bezdek, 1981), is given as

$$W_{lm} = \sum_{j=1}^{k} \sum_{i=1}^{n} m_{ij}^{\phi} \left(x_{il} - c_{jl}\right)\left(x_{im} - c_{jm}\right), \ \forall(l,m), l,m = 1,\dots,p \ . \quad (16)$$

The SSP between-classes covariance matrix B_{lm} are given as

$$B_{lm} = \sum_{j=1}^{k} \left(\sum_{i=1}^{n} m_{ij}^{\phi}\right) \left(c_{jl} - \bar{\bar{x}}_l\right)\left(c_{jm} - \bar{\bar{x}}_m\right), \ \forall(l,m), l,m = 1,\dots,p \ . \quad (17)$$

The ratio of within-classes to the total SSP matrix is the Wilks' lambda (Eq. 18, Wilks, 1932; Oh et al. 2005). Wilks' lambda for multi-dimensional observations is a $p \times p$ matrix and the determinant of the matrix represents the geometric volume of this object in p dimensions, written as

$$\Lambda = \frac{|W_{lm}|}{|W_{lm}+B_{lm}|} \ . \quad (18)$$

Here we use Wilks' lambda (Λ) as a measure of the difference between classes, although a scalar cannot replace a vector for investigating different aspects of Wilks' lambda. The value Λ varies from 0 to 1, where 0 suggests that classes differ (within-classes SSP is smaller compared to between-classes SSP), and 1 suggests that all classes are the same.

For the January 2008 data, Wilks' Λ for different dimensional observations are calculated and summarized in Table 4. For 4 dimensional (p=4) observations, Wilks' Λ could be as small as 0.21 for 2-class FKM and even smaller (0.05) for 3-class FKM. This means that the classes generated by the FKM method are well separated, with clusters quite different from each other, particularly in 3-class FKM. For FKM 2-class, the value of Wilks' Λ is largest for the mid-layer altitude dimension, indicating a more overlaid classification compared to classification using any other observational dimension. The reason for the large

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

value of Wilks' Λ for the dimension altitude is because clouds have two distinct altitude centers, one for low water clouds and the other for high ice clouds. (Mid-level clouds occur too infrequently to form a third dominant altitude center). The center altitude of water clouds is comparable to that of aerosols. These distinct altitude centers induce large within-classes SSP and hence large values of Wilks' Λ. Following the values of Wilks' Λ from the dimension of altitude are the ones from color ratio and depolarization ratio, with the Wilks' Λ from backscatter intensity being the smallest. The value of Wilks' Λ from any combination observations is between the maximum and minimum values of those dimensions. For FKM 3-class analyses, large values of Wilks' Λ are produced by depolarization, and followed by altitude, color ratio and backscatter.

Input parameters	Λ, 2 classes	Λ, 3 classes
Backscatter intensity, depolarization ratio, color ratio, altitude	0.21	0.048
Depolarization ratio, color ratio, altitude	0.20	0.060
Backscatter intensity, color ratio, altitude	0.20	0.060
Backscatter intensity, depolarization ratio, altitude	0.17	0.035
Backscatter intensity, depolarization ratio, color ratio	0.14	0.030
Backscatter intensity and depolarization ratio	0.12	0.025
Backscatter intensity and color ratio	0.14	0.039
Backscatter intensity and altitude	0.14	0.052
Depolarization ratio and color ratio	0.13	0.043
Depolarization ratio and altitude	0.20	0.056
Color ratio and altitude	0.23	0.077
Backscatter intensity	0.08	0.030
Depolarization ratio	0.16	0.136
Color ratio	0.18	0.053
Altitude	0.28	0.121

Table 4: Wilks' lambda (Λ) for 2-class (center column) and 3-class (rightmost column) FKM classifications using different dimensional observations (left column)

4.3 Principal Component Analysis

In this section we apply principal component analysis (PCA, Wold et al. 1987) to the FKM classifications to determine which of the input parameters account for the greatest variability in the outputs. The first function (PCA-1) maximizes the differences between the classes and represents the dominant contribution to the classifications. Successive functions (PCA-2) will be orthogonal to, or independent of, the other functions and hence their contributions to the discrimination between classes will not overlap. These functions, or canonical variants, are calculated from the eigenvalues and eigenvectors of matrix B_f/W_f (the ratio of within-class variance and between-class variance). Using this method will help us understand how independent the input parameters are and how they individually contribute to the classifications.

The scatter plots of PCA-1 and PCA-2 for FKM 2 classes and 3 classes are shown in Figure 13. The projection of vector lengths on PCA-1 and PCA-2 of different measurements (i.e., backscatter intensity, depolarization ratio, color ratio, and mid-

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

layer altitude) indicate how much each individual dimension contributes to the classifications. Longer projections mean stronger contributions. From the figure, we clearly see that water clouds, ice clouds and aerosols are quite different (i.e., their cluster centers are located in different positions). Class 1 (cloud) of 2-class FKM breaks into 2 classes (ice cloud and water cloud) when applying 3-class FKM. We can also see that color ratio and depolarization ratio contribute the most on PCA-1 (longer projections on the axis of PCA-1 in both subpanels) while backscatter and height contribute more on PCA-2. Hence, color ratio and depolarization are the driving components for the cloud and aerosol classification. From figure 13b, we could also argue that the backscatter and depolarization are the driving factors in classifying water and ice clouds (projections of the vectors on C1-C2, namely combined projection of PCA-1 and PCA-2, are longer), while color ratio and altitude also contribute to the classification. Altitude, and, to a greater extent, color ratio and depolarization ratio are the driving factors that allow aerosols to be separated from ice clouds (projections of the vectors on C1-C3 are longer), whereas backscatter intensity and color ratio are the driving factors that separate water clouds from aerosols (projections of the vectors on C2-C3 are longer). Comparing contributions of individual measurements to different classes, mid-layer altitude is most useful in helping discriminate aerosols and ice clouds, while simultaneously being the least useful in aerosols and water clouds classification. Backscatter intensity is the most useful parameter in the aerosols vs. water clouds classification and the water clouds vs. ice clouds classification, and the least useful in the aerosols vs. ice clouds classification. Depolarization is most useful in distinguishing between water clouds and ice clouds and between aerosols and ice clouds, and the least useful in the aerosols vs. water clouds classifications. These observations agree very well with earlier findings in figures 6 and 12 and tables 2 and 4.

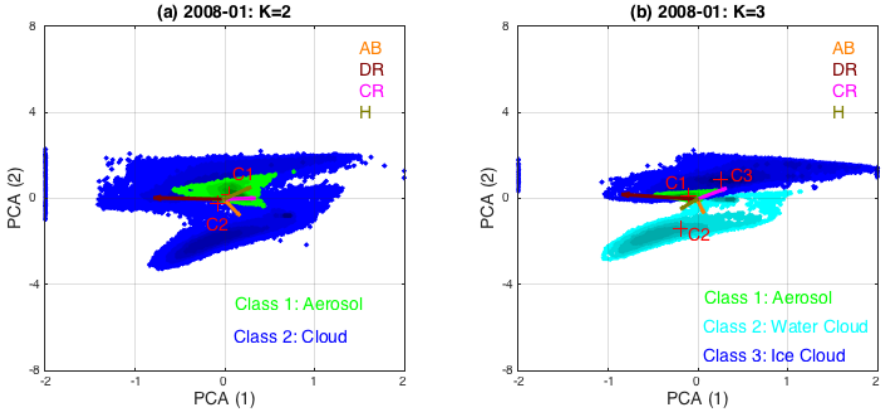


Figure 13: principle component analysis of the FKM classifications for the January 2008 test data. PCA results for the 2-class CAD_{FKM} classifications are shown on the left (panel a), and the 3-class CAD_{FKM} classifications are shown on the right (panel

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

b). In both figures, the green points are projections of aerosol data onto the PCA axes with their center located at red crosses labeled C1. Similarly, the blue and cyan (left panel only) points are projections of cloud data onto the PCA axes. In panel a, the blue points represent all clouds, while in panel b the blue points represent ice clouds and the cyan points represent water clouds. Also shown in both panels are color-coded vectors representing each of the classification variables: backscatter intensity (AB, in orange), depolarization ratio (DR, in brown), color ratio (CR, in magenta), and altitude (H, in olive). The projections of the variable vectors along the principal component axes indicate the degree to which each variable contributes to PCA1 and PCA2. Variable vectors that are parallel to either PCA1 or PCA2 contribute essential information to that component, while vectors that are perpendicular do not contribute at all.

4.4 Error propagation

In this section, we assess the impact of instrument noise and measurement uncertainties on the FKM classifications. The observations from 6 September 2008 at ~01:35:29 UTC are used to investigate how noise in the lidar measurements affects the accuracy of the clustering results and what, if any, biases are introduced into the cloud and aerosol classifications. To simulate the measurement uncertainties, two different methods are used. In the first method, pseudo-random variables were drawn from Gaussian distributions having means equal to the various measured values and standard deviations between 10% and 200% of the means. As illustrated in Figure 14, using this method allows us to quantify the effects of varying measurement errors on the FKM classification algorithm results. A sequence of Monte Carlos tests was constructed in which one of the four classification variables was randomly perturbed (i.e., drawn from the aforementioned Gaussian distributions) while the other three remained unchanged. For each of the four tests, 100 realizations of simulated input were created. To estimate the propagation of measurement uncertainties, we calculated the shifts in classification, confusion index (CI, see section 3.2) and cluster centers between new clusters with added noise and the original clusters derived using unperturbed inputs. The shifts in cluster center are the mean distances between the centers of the new clusters (C_n, obtained from perturbed dataset) and the old ones (C_o, obtained from error free dataset) for both clouds and aerosols, calculated using Eq. 19 (Omar et al., 2005). These distances are normalized by the standard deviation of the distributions (C_{std}) of individual record distances from unperturbed center as

.

$$\delta d_e = \frac{|\overline{c_n - c_o}|}{C_{std}}. \qquad (19)$$

Figure 14 plots the shifts in cluster centers or (a) each class, (b) the fraction of correct classifications, and (c) the revised confusion index as a function of relative uncertainties ranging from 10% to 200%. From Figure 14a we see that shifts in cluster centers between perturbed and unperturbed data are very small when the uncertainties are small. The largest shift comes from

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

color ratio perturbations and the smallest shift comes from backscatter perturbations. Perturbations on class-2 (aerosol) are more important compared to class-1 (cloud). Figures 14b and 14c show that when the uncertainties in the measurements are small (i.e. less than 10%), the errors in the classifications are also small (e.g., less than 2% in Figure 14b) with less overlaps between classes (e.g., small values of CI from 0.3 to 0.305 seen in Figure 14c). When the uncertainties increase, the classification accuracies slightly decrease and the shifts in cluster center and CI slightly increase. The rates of change in the accuracy and confusion index are rapid at first (i.e., between relative uncertainties between 10% and 100%), but tend to be stable for larger uncertainties. Large measurement uncertainties (i.e., 200%) in color ratio can introduce biases of 20% in the classification results, with CI values less than 0.335. This suggests that uncertainties in the measurements can cause misclassification, but that most of the classifications (~80%) are still robust. This is because cloud and aerosol properties are largely distinct and the misclassifications that do occur may come from features such as the few very thin clouds and dense aerosols in the transitional zone in Figure 6.

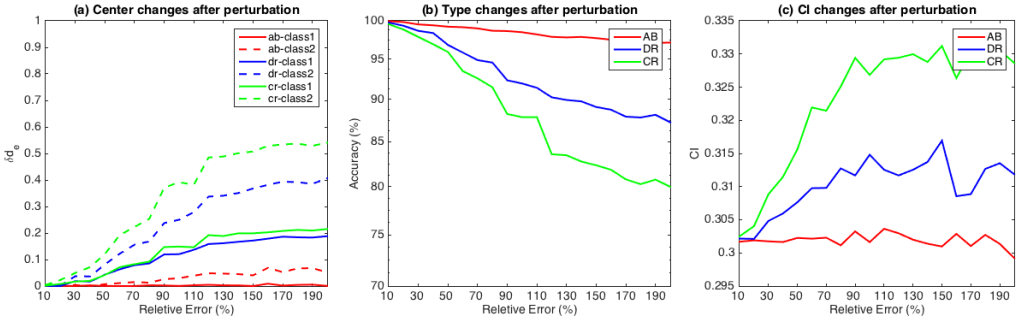


Figure 14: classification changes as a function of errors in the input parameters. The left panel (a) shows shifts in cluster centers for each class; the center panel (b) shows the relative accuracy of the FKM classifications; and the right panel (c) shows changes in the cluster confusion indexes. Panels b and c show perturbations in the classifications due to uncertainties in attenuated backscatter intensity (AB, in red), depolarization ratio (DR, in blue), and color ratio (CR, in green).

Our first error propagation test used arbitrarily assigned relative uncertainties between 10% and 200% of the parameter mean values. In our second test we used the measured uncertainties reported in the CALIOP layer products to construct the Gaussian distributions from which pseudo-random variables were generated. By using this method, we can assess the actual impacts on the classifications due to noise in the CALIPSO measurements. To isolate the influence of the individual inputs, three test cases were constructed in which only one parameter was varied in each case. Figure 15 shows the results. Figure 15a shows the unperturbed results, while Figures 15b–15d show CAD_{FKM} scores averaged over 100 perturbations of the test parameter. Figure 15b shows the results when the attenuated backscatter intensities are varied, Figure 15c shows the results when the depolarization ratios are varied, and Figure 15d shows the results when the color ratios are varied.

Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-168
Manuscript under review for journal Atmos. Meas. Tech.
Discussion started: 28 June 2018

Atmospheric Measurement Techniques Discussions

From Figure 15 we find that the averaged CAD_{FKM} scores from the perturbed datasets do not differ markedly from the CAD_{FKM} scores in the unperturbed dataset. In more than 88% of the cases, clouds are still classified as clouds and aerosols are still classified as aerosols. When examining perturbations to backscatter intensity alone (Figure 15b), we find that the perturbed and unperturbed classification results are identical more than 98% of the time. However, the CAD_{FKM} differences arising from perturbations to depolarization ratio and color ratio (Figures 15c and 15d, respectively) can be much larger. This finding is consistent with results shown earlier in Figure 14.

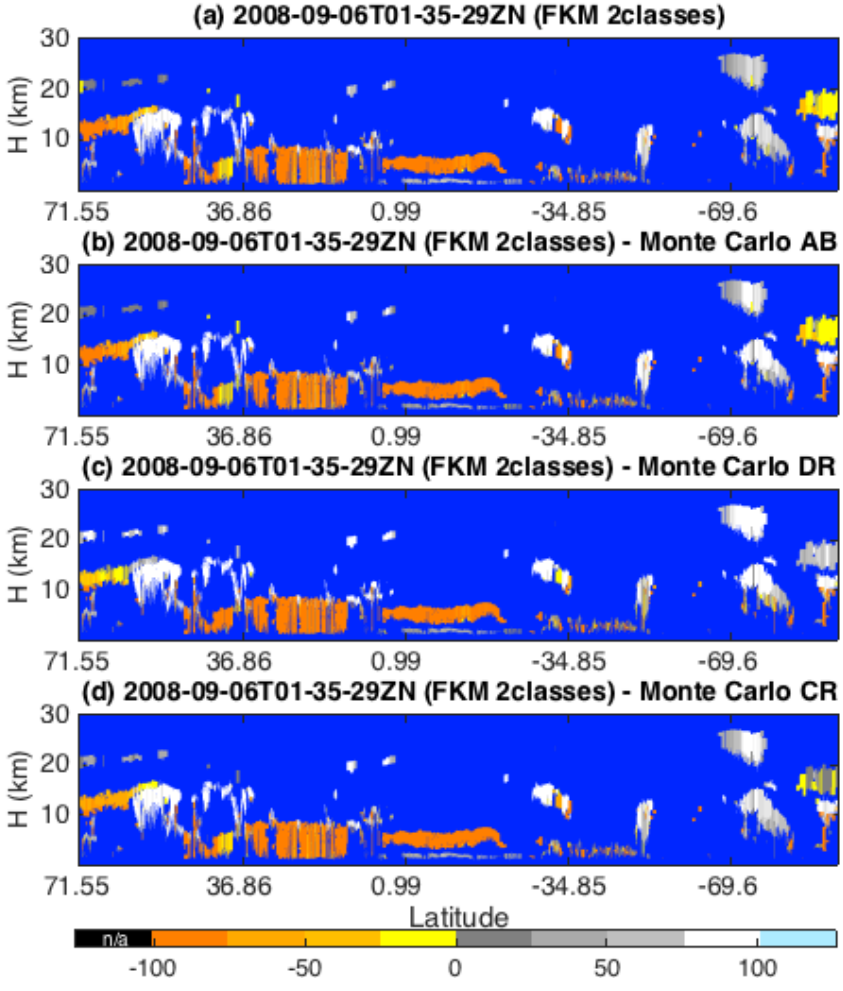


Figure 15: CAD scores for the same orbit shown in Fig. 4 (06 Sep. 2008, 01:35:29 GMT). The uppermost panel (a) shows 2-class FKM results derived using unperturbed measurements. Panels b, c, and d show, respectively, 2-class FKM results derived using perturbed measurements of attenuated backscatter intensity (AB), depolarization ratio (DR), and color ratio (CR).

Most often, the perturbed measurements only induce CAD_{FKM} changes for features that were originally classified with low confidence and for those challenging features such as water clouds beneath smoke, high altitude aerosols, and PSCs, whose input parameters frequently lie in the transition zone between clouds and aerosols. Water clouds beneath thick smoke layers are an especially difficult case, as the uncertainties introduced by the absorption of smoke at 532 nm can significantly reduce the confidence of the water cloud classification. Looking at Figure 15, together with Figures 2, 11 and 13, we find that this is a reasonable and even expected result. From Figures 11 and 15 we know that the most effective measurements for separating water clouds and aerosols are color ratio and backscatter intensity. But relative to measurements of water clouds in otherwise clear skies, the color ratios for water clouds lying under absorbing smoke layers have large positive biases while the backscatter intensities have large negative biases, and these biases will produce low confidence CAD scores, both for the FKM method and the V4 operational method (Liu et al., 2018). A somewhat similar scenario can occur in the classification of high altitude aerosols, where high biases (i.e., measurement errors) in depolarization ratios and color ratios can lead to the misclassification of aerosols as ice clouds.

5. Conclusions

In this paper we use the Fuzzy K-Means (FKM) algorithm to validate the performance of the cloud-aerosol discrimination (CAD) algorithm used in the standard processing of the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) measurements. Being able to accurately separate clouds from aerosols is an essential task in the analysis of the elastic backscatter lidar measurements being continuously acquired by Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission. When coupled to a well-validated CAD algorithm, the data products delivered by CALIOP can be used to reliably map the vertical distributions of clouds and aerosols on global and regional scales throughout the full 12 years of the CALIPSO mission.

To assess the performance of the operational CAD algorithm, we have compared its outputs to those derived from the same scenes using an implementation of the FKM technique. The CALIOP operational CAD algorithm is a supervised learning technique, in which classification decisions are tuned to match externally provided expert human judgements. Unlike the operational CAD, the FKM is an unsupervised learning scheme, which assigns class memberships based on similarities discovered in the inherent characteristics of the input data. While the two algorithms use largely identical inputs, the underlying mathematical formulations are entirely different, as is the framework for expressing class membership values. The flexibility of the FKM technique allows us to investigate the relative importance of various inputs in deriving the final classifications and

to explore classification biases arising from current lidar measurement techniques. Establishing these performance metrics should enable the development of enhanced classification schemes for use with future space-based lidars.

The key finding of this study is that the feature classifications assigned by CALIOP operational CAD algorithm are very closely replicated by the FKM method. Our assessment of parameter PDFs shows that, in general, one half-orbit of data is sufficient to represent a much larger data set in the cloud-aerosol classification process. The classifications obtained from our independently derived FKM analyses compare well with the classifications determined by CALIOP's operational V4 CAD algorithm and reported in the CALIOP V4 data products. Using a one-month test set, the 2-class and 3-class FKM classifications agreed with the V3 and V4 operational data products over 93 % of the time, and the 3-class FKM results agreed with the V4 operational CAD in 94~95 % of all cases. This strong agreement between two independent methods provides convincing evidence that V4 CAD operational algorithm is delivering robust and accurate classifications.

Those instances where the two methods fail agree (5~6 % of all cases) are typically highly ambiguous scenes in which the observables lie in the overlap regions between the peaks of the cloud and aerosol PDFs. In particular, in scenes containing Taklamakan dust (or lofted Asian dust in general), high altitude smoke plumes, cirrus fringes, and/or volcanic ash, both the V4 operational CAD and the FKM algorithm struggle to make accurate classifications. The Taklamakan dust cases provide an instructive example that illustrates the classification conundrum. Over the Taklamakan, lofted dust layers and cirrus clouds occur in similar temperature regimes, and frequently have similar backscatter intensities and depolarization ratios. The most critical criterion for distinguishing clouds from aerosols is color ratio, and the characteristic color ratios of dust and cirrus are reasonably distinct (~0.75 vs. ~1.01). However, the natural variability within each feature type is quite broad (e.g., ±0.25 for cirrus), and the measurements are very noisy, especially during daytime.

To characterize the CAD improvements made in the most recent CALIOP data release, we used the FKM method to explore the capabilities of both the V3 and V4 operational CAD algorithms. As expected, the V4 operational algorithm was more effective than the V3 version, but the overall differences were not large. The primary differences are found by examining the results obtained for specific feature classes. The FKM classifications agree well with both the V3 and V4 CAD results in most cirrus fringe and dense aerosol cases and agree well with V4 CAD results for lofted Asian dust, high altitude smoke, and volcanic ash. FKM classifications of stratospheric features and polar region features had the largest uncertainties. More studies are needed to better understand why these specific types of features are proving so resistant to confident classification, irrespective of the algorithmic approach applied.

Our investigation of error propagation in the FKM shows that while measurement uncertainties on the order of the CALIPSO measured noise will introduce biases into the cloud and aerosol classifications, more than 80% of the classifications stay unchanged. For the rest of classifications (which are low confidence clouds or aerosols), as the uncertainties increase, the

classification confidence decreases, as indicated by higher confusion indexes, and the classification accuracies decrease as well. The dependence and the number of measurements can also impact the classification efficiency. Key parameter analysis shows that higher classification accuracies are achieved by increasing the number of independent observational parameters used in the analyses. FKM classifications using only a single input yield the same results as the operational CAD in only ~60 % of all classifications, a rate only marginally better than would be expected from random choice. While using three parameters achieved an agreement between the FKM and the V4 operational CAD in the neighborhood of ~80 %, raising the agreement to ~95 % required four parameters. When only three inputs were used, removing color ratio from the FKM caused the largest classification disparities between the two methods.

Certain parameters are especially significant for the classification of particular feature types, and thus optimizing the number of successful classifications across all features requires the inclusion of all measurements that effectively contributed to any species-specific classification. Principal component analysis and key parameter analysis together show that the most important dimensions for distinguishing between clouds and aerosols are depolarization ratios and color ratios; that backscatter intensity and depolarization are the driving factors in classifying water and ice clouds; and that altitude, color ratio and depolarization ratio are the key inputs that allow aerosols to be separated from ice clouds, while backscatter intensity is the critical factor for separating aerosols and water clouds. Moreover, from fuzzy linear discriminant analysis we found the values of Wilks' lambda are close to 0, confirming that the FKM classification technique reliably separates clouds from aerosols.

While the FKM and official CAD classification methods both provide reliable discrimination between clouds and aerosols in the CALIOP data set, the FKM method is much more time consuming than the operational algorithm. On the other hand, the flexibility of FKM method offers opportunities to explore the effectiveness of future classification schemes that potentially incorporate measurements from multiple sensors, perhaps even from multiple satellites. While the input data used by our implementation of the FKM technique is essentially identical to that required by the CALIOP V4 operational algorithm, the two decision-making frameworks are independently derived and rely on very different mathematics (i.e., probabilities vs. fuzzy logic). The very close similarity between the results produced by the two independent approaches argues strongly that the V4 operational classifications are essentially correct at the 94% level.

References

Avery, M. A., Ryan, R., Getzewich, B., Vaughan, M., Winker, D., Hu, Y., and Trepte, C.: Impact of Near-Nadir Viewing Angles on CALIOP V4.1 Cloud Thermodynamic Phase Assignments, in preparation, 2018.

Bankert, R. L. and Solbrig, J. E.: Cluster Analysis of A-Train Data: Approximating the Vertical Cloud Structure of Oceanic Cloud Regimes, J. Appl. Meteor. Climatol., 54, 996–1008. doi:http://dx/doi.org/10.1175/JAMC-D-14-0227.1, 2015.

Bezdek, J. C.: Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

Bezdek, J. C., Ehrlich, R., and Full, W.: FCM: the fuzzy c-means clustering algorithm, Comput. Geosci., 10, 191-203, 1984.

Burton, S. P., Ferrare, R. A., Hostetler, C. A., Hair, J. W., Rogers, R. R., Obland, M. D., Butler, C. F., Cook, A. L., Harper, D. B., and Froyd, K. D.: Aerosol classification using airborne High Spectral Resolution Lidar measurements - methodology and examples, Atmos. Meas. Tech., 5, 73-98, doi:10.5194/amt-5-73-2012, 2013.

Burrough, P. A., Van Gaans, P. F. M., and MacMillan, R. A.: High-resolution landform classification using fuzzy K-means. Fuzzy Set, Syst., 113, 37–52, 2000.

Burrough P. A. and McDonnell R. A.: Principles of Geographic Information Systems, Oxford University Press, Oxford, 1998.

Chen, B., Huang, J., Minnis, P., Hu, Y., Yi, Y., Liu, Z., Zhang, D., and Wang, X.: Detection of dust aerosol by combining CALIPSO active lidar and passive IIR measurements, Atmos. Chem. Phys., 10, 4241-4251, https://doi.org/10.5194/acp-10-4241-2010, 2010.

Chepfer, H., Bony, S., Winker, D., Chiriaco, M., Dufresne, J.-L., Sèze, G.: Use of CALIPSO lidar observations to evaluate the cloudiness simulated by a climate model, Geophys. Res. Lett. 35, L15704, DOI: 10.1029/2008GL034207, 2008.

DeMott, P. J., Prenni, A. J., Liu, X., Kreidenweis, S. M., Petters, M. D., Twohy, C. H., Richardson, M. S., Eidhammer, T., and Rogers, D. C.: Predicting global atmospheric ice nuclei distributions and their impacts on climate, PNAS, 107, 11217-11222, doi:10.1073/pnas.0910818107, 2010.

Di Pierro, M., Jaeglé, L., and Anderson, T. L.: Satellite observations of aerosol transport from East Asia to the Arctic: three case studies, Atmos. Chem. Phys., 11, 2225-2243, doi:10.5194/acp-11-2225-2011, 2011.

Fisher, R. A: The use of multiple measurements in taxonomic problems, Ann. Eugen. (Lond), 7, 179-188, 1936.

Gharibzadeh, M., Alam, K., Abedini, Y., Bidokhti, A. A., Masoumi, A., and Bibi, H.: Characterization of aerosol optical properties using multiple clustering techniques over Zanjan, Iran, during 2010–2013, Appl. Opt., 57, 2881–2889, doi:10.1364/AO.57.002881, 2018.

Getzewich, B., Vaughan, M. A., Hunt, W. H., Avery, M. A., Tackett, J. L., Kar, J., Lee, K.-P., Toth, T., and Powell, K.: CALIPSO Lidar Calibration at 532 nm: Version 4 Daytime Algorithm, in preparation, 2018.

Gorsevski , P. V., Gessler, P. E. and Jankowski P.: Integrating a fuzzy K-means classification and a Bayesian approach for spatial prediction of landslide hazard, J. Geograph. Syst., 5, 223-251, 2003.

Hartigan, J. A. and Wong, M. A.: Algorithm AS 136: A K-means Clustering Algorithm, J. Roy. Statist. Soc. Ser. B, 28, 100–108, 1979.

Höpfner, M., M. Pitts, and L. Poole, 2009: "Comparison between CALIPSO and MIPAS observations of polar stratospheric clouds", J. Geophys. Res., 114, D00H05, doi:10.1029/2009JD012114.

Hu, Y., Winker, D., Vaughan, M., Lin, B., Omar, A., Trepte, C., Flittner, D., Yang, P., Nasiri, S. L., Baum, B., Holz, R., Sun, W., Liu, Z., Wang, Z., Young, S., Stamnes, K., Huang, J., Kuehn, R.: CALIPSO/CALIOP cloud phase discrimination algorithm, J. Atmos. Oceanic Technol., 26, 2293-2309, 2009.

Jin, Y., Kai, K., Okamoto, H., and Hagihara, Y.: Improvement of CALIOP cloud masking algorithms for better estimation of dust extinction profiles, J. Meteorol. Soc., 92, 433-455, 2014.

Kar, J., Vaughan, M. A., Lee, K.-P., Tackett, J. L., Avery, M. A., Garnier, A., Getzewich, B. J., Hunt, W. H., Josset, D., Liu, Z., Lucker, P. L., Magill, B., Omar, A. H., Pelon, J., Rogers, R. R., Toth, T. D., Trepte, C. R., Vernier, J.-P., Winker, D. M., and Young, S. A.: CALIPSO lidar calibration at 532 nm: version 4 nighttime algorithm, Atmos. Meas. Tech., 11, 1459–1479, doi:10.5194/amt-11-1459-2018, 2018.

Key, J. R., Maslanik, J. A. and Barry, R. G.: Cloud classification from satellite data using a fuzzy sets algorithm: a polar example. International Journal of Remote Sensing, 10, 1823-1842, 1989.

Kim, M.-H., Omar[2], A. H., Tackett, J. L., Vaughan, M. A., Winker, D. M., Trepte, C. R., Hu, Y., Liu, Z., Poole, L. R., Pitts, M. C., Kar, J., and Magill, B. E.: Updates of the Version 4 CALIPSO Level 2 Aerosol Products, in preparation, 2018.

Konsta, D., Chepfer, H., and Dufresne, J. L.: Evaluation of Cloud Description in General Circulation Models Using A-Train Observations, Advances in Meteorology, Climatology and Atmospheric Physics. Springer Atmospheric Sciences. Springer, Berlin, Heidelberg, 2013.

Kubat, M., Holte, R.C. and Matwin, S.: Machine Learning for the Detection of Oil Spills in Satellite Radar Images, Machine Learning, vol. 30, nos. 2/3, pp. 195-215, 1998.

Liu, Z., Vaughan, M. A., Winker, D. M., Hostetler, C. A., Poole, L. R., Hlavka, D., Hart, W. and McGill, M.: Use of probability distribution functions for discriminating between cloud and aerosol in lidar backscatter data, J. Geophys. Res., 109, D15202, doi:10.1029/2004JD004732, 2004.

Liu, Z., Vaughan, M., Winker, D., Kittaka, C., Getzewich, B., Kuehn, R., Omar, A., Powell, K., Trepte, C. and Hostetler C.: The CALIPSO Lidar Cloud and Aerosol Discrimination: Version 2 Algorithm and Initial Assessment of Performance, J. Atmos. Oceanic Technol., 26, 1198–1213, doi:10.1175/2009JTECHA1229.1, 2009.

Liu, Z., Kar, J., Zeng, S., Tackett, J., Vaughan, M., Getzewich, B., Lee, K.-P., Magill, B., Omar, A., Lucker, P., Trepte, C., and Winker, D.: Discriminating Between Clouds and Aerosols in the CALIOP Version 4.1 Data Products, in preparation, 2018.

Luo, Z. J., Anderson, R. C., Rossow, W. B., and Takahashi, H.: Tropical cloud and precipitation regimes as seen from near-simultaneous TRMM, CloudSat, and CALIPSO observations and comparison with ISCCP, J. Geophys. Res., 122, 5988–6003, doi:10.1002/2017JD026569, 2017.

Mahalanobis, P. C.: On the generalised distance in statistics, Proceedings of the National Institute of Sciences of India, 2, 49–55, 1936.

Minasny, B., and McBratney, A. B.: FuzME version 3.0, Australian Centre for Precision Agriculture, The University of Sydney, Australia, 2002.

McBratney, A. B., and de Gruijter, J. J.: A continuum approach to soil classification by modified fuzzy k-means with extragrades, European Journal of Soil Science, 43, 159–175, 1992.

McBratney, A.B., and Moore, A.W.: Application of fuzzy sets to climatic classification. Agricultural and Forest Meteorology, 35, 165–185, 1985

Nam, C. C. W. and Quaas, J.: Evaluation of Clouds and Precipitation in the ECHAM5 General Circulation Model Using CALIPSO and CloudSat Satellite Data, J. Climate, 2012.

Nock, R. and Nielsen, F.: On Weighting Clustering, IEEE Trans. Pattern Anal. Mach. Intell., 28, 1-13, 2006.

Odeh, I. O. A., McBratney, A. B., and Chittleborough, D. J.: Soil pattern recognition with fuzzy c-means: application to classification and soil landform interrelationships, Soil Sci. Soc. Am. J., 56, 505–516, 1992a.

Odeh, I. O. A., McBratney, A. B., and Chittleborough, D. J.: Fuzzy c-means and kriging for mapping soil as a continuous system. Soil Sci. Soc. Am. J., 56, 1848–1854, 1992b.

Oh, C., Tok, A., and Ritchie, S. G.: Real-Time Freeway Level of Service Using Inductive-Signature-Based Vehicle Reidentification System, IEEE Trans. Intell. Transp. Syst., 6, 138-146, 2005.

Omar, A. H., Winker, D. M., Vaughan, M. A., Hu, Y., Trepte, C. R., Ferrare, R. A., Lee, K., Hostetler, C. A., Kittaka, C., Rogers, R. R., Kuehn, R. E., and Liu, Z.: The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm, J. Atmos. Oceanic Technol., 26, 1994 - 2014, doi:10.1175/2009JTECHA1231.1, 2009.

Omar, A. H., Won, J.-G., Winker, D. M., Yoon, S.-C., Dubovik, O., McCormick, M. P.: Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements, J. Geophys. Res. Atmos., 110(10D), doi:10.1029/2004JD004874, 2005.

Pitts, M. C., Poole, L. R., Lambert, A., Thomason, L. W.: An assessment of CALIOP polar stratospheric cloud composition classification, Atmos. Chem. Phys., 13, 2975-2988, doi:10.5194/acp-13-2975-2013, 2013.

Pitts, M. C., Poole, L. R.: Polar stratospheric cloud climatology based on CALIPSO spaceborne lidar measurements from 2006-2017, Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2018-234, in review, 2018.

Prata, A. T., Young, S. T., Siems, S. T., and Manton, M. J.: Lidar ratios of stratospheric volcanic ash and sulfate aerosols retrieved from CALIOP measurements, Atmos. Chem. Phys., 17, 8599-8618, 2017.

Roubens, M.: Fuzzy clustering algorithms and their cluster validity, Eur. J. Oper. Res., 10, 294-301, 1982.

Stephens, G., D. Winker, J. Pelon, C. Trepte, D. Vane, C. Yuhas, T. L'Ecuyer and M. Lebsock, 2017: "CloudSat and CALIPSO within the A-Train: Ten years of actively observing the Earth system", B. Am. Meteorol. Soc., doi:10.1175/BAMS-D-16-0324.1, in press.

Triantafilis, J., Odeh, I. O. A., Minasny, B., and McBratney, A. B.: Elucidation of physiographic and hydrogeological features of the lower Namoi valley using fuzzy K-means classification of EM34 data, Environ. Model. Softw.,18, 667–680, 2003.

Usman, B.: Satellite Imagery Land Cover Classification using K-Means Clustering Algorithm Computer Vision for Environmental Information Extraction. Elixir Comp. Sci. & Engg,. 63, 2013, 18671-18675, 2013

Vaughan, M., K. Powell, R. Kuehn, S. Young, D. Winker, C. Hostetler, W. Hunt, Z. Liu, M. McGill, B. Getzewich: Fully

Automated Detection of Cloud and Aerosol Layers in the CALIPSO Lidar Measurements, J. Atmos. Oceanic Technol., 26, 2034–2050, doi:10.1175/2009JTECHA1228.1, 2009.

Vaughan, M., Garnier, A., Josset, D., Avery, M., Lee, K.-P., Liu, Z., Hunt, B., Getzewich, B., and Burton, S.: CALIPSO Lidar Calibration at 1064 nm: Version 4 Algorithm, in preparation, 2018.

Wilks, S. S.: Certain generalizations in the analysis of variance, Biometrika, 24, 471-494, 1932.

Winker, D. M., R. H. Couch, and M. P. McCormick: An overview of LITE: NASA's Lidar In-space Technology Experiment, Proc. IEEE, 84, 164–180, doi:10.1109/5.482227, 1996.

Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., Powell, K. A., Liu, Z., Hunt, W. H. and Young, S. A.: Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms, J. Atmos. Oceanic Technol., 26, 2310–2323, doi:10.1175/2009JTECHA1281.1, 2009.

Winker, D. M., J. Pelon, J. A. Coakley, Jr., S. A. Ackerman, R. J. Charlson, P. R. Colarco, P. Flamant, Q. Fu, R. Hoff, C. Kittaka, T. L. Kubar, H. LeTreut, M. P. McCormick, G. Megie, L. Poole, K. Powell, C. Trepte, M. A. Vaughan, B. A. Wielicki: "The CALIPSO Mission: A Global 3D View Of Aerosols And Clouds", Bull. Am. Meteorol. Soc., 91, 1211–1229, doi:10.1175/2010BAMS3009.1, 2010.

Wold, S., Esbensen, K., Geladi, P.: Principal component analysis, Chem. Intell. Lab. Syst., 2 37-52, doi:10.1016/0169-7439(87)80084-9, 1987.

Zhang, Y., Klein, S., Mace, G. G., and Boyle, J.: Cluster analysis of tropical clouds using CloudSat data, Geophys. Res. Lett., 34, L12813, doi:10.1029/2007GL029336, 2007.