Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 107 (2007) 159–171
Remote Sensing of Environment
www.elsevier.com/locate/rse

Selection of the automated thresholding algorithm for the Multi-angle Imaging SpectroRadiometer Radiometric Camera-by-Camera Cloud Mask over land

Yuekui Yang [a,*], Larry Di Girolamo [a], Dominic Mazzoni [b]

[a] Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States
[b] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States

Received 28 February 2006; received in revised form 18 May 2006; accepted 31 May 2006

Abstract

The Radiometric Camera-by-Camera Cloud Mask (RCCM) is archived at the NASA Langley Distributed Active Archive Center as one of the standard products from the Multi-angle Imaging SpectroRadiometer (MISR) mission. The RCCM algorithm applied over land surfaces uses an Automated Threshold Selection Algorithm (ATSA) to derive thresholds that are applied to a cloud masking test to determine whether a given image pixel is clear or contains cloud. In this article, we established a framework for the selection of ATSA and the cloud masking tests, which is not only suitable for the RCCM over land, but cloud detection for other satellite missions. Using this framework, we have undertaken the largest comparison of existing histogram-based ATSAs (16 in total) and applied them to four cloud masking tests that can be constructed from the MISR radiances, namely the red channel bidirectional reflectance function (BRF), the standard deviation (STDV) of the red channel BRF, the normalized difference vegetation index (NDVI), and a parameter D that is constructed by optimizing the information from NDVI and red channel BRF for cloud detection. The cloud masking tests and ATSAs are applied to 35 MISR scenes from six snow-free land surface types. To evaluate their performance, reference cloud masks are constructed for the 35 scenes using interactive, supervised learning, visualization software. Independent of the ATSA and as a single cloud masking test, D performed the best in terms of having the lowest misclassification rate using the best possible threshold, the highest bimodal rate in the shape of the histograms derived from the 35 scenes, and the least sensitivity to errors in the choice of threshold. Of the 16 ATSAs, the methods of Li and Lee [Li, C.H., and Lee, C.K., (1993). Minimum cross-entropy thresholding. Pattern Recognition, 26(4), 617-625.] and Pal and Bhandari [Pal, N. R., and Bhandari, D., (1993). Image thresholding: some new techniques. Signal Processing, 33, 139–158.] performed the best when applied to D, with essentially unbiased performance and a root mean square of 15% when compared to cloud masks using the best possible thresholds. It is recommended that increased performance of the RCCM-land algorithm can be had through an increase in the space–time sampling used to generate histograms of D and the addition of a STDV cloud masking test to improve the detection of small cumulus clouds.

Keywords: MISR; Cloud detection; Automated threshold selection algorithm; Daytime cloud mask

1. Introduction

The Multi-angle Imaging SpectroRadiometer (MISR) instrument was launched into a 705-km earth orbit on December 21, 1999, on board NASA's Terra spacecraft. It views the Earth with radiometers placed at nine different angles ranging from −70.5° to 70.5° along the orbital track, making measurements of scattered solar radiation from the Earth at four spectral bands centered at 446 nm, 558 nm, 672 nm and 866 nm at a spatial resolution of 275 m–1100 m (Diner et al., 2002). The measured radiances are converted to surface, cloud and aerosol properties that are used to improve our understanding of the climate system (e.g., Diner et al., 1998). Like most other meteorological satellite instruments, a cloud mask is required to label pixels as either clear or cloudy before other geophysical products could be derived (e.g., Rossow, 1989). Given that a cloud mask depends on line-of-sight (e.g., Zhao & Di Girolamo, 2004),

* Corresponding author. Tel.: +1 217 333 6349; fax: +1 217 244 4393.
E-mail address: yuekui@atmos.uiuc.edu (Y. Yang).

0034-4257/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2006.05.020

MISR requires nine separate cloud masks, one for each camera. To meet this requirement, the NASA Langley Distributed Active Archive Center (DAAC) produces a cloud mask product for MISR called the Radiometric Camera-by-Camera Cloud Mask (RCCM).

The MISR RCCM algorithm is described in its Algorithm Theoretical Basis Document (Diner et al., 1999). The algorithm is divided into ocean and land components, depending on the surface being viewed by a MISR camera. As is customary in most cloud detection algorithms, a series of thresholding tests are employed (e.g., Ackerman et al., 1998; Rossow & Garder, 1993; Saunders & Kriebel, 1988; Stowe et al., 1999). For example, if the pixel value X > Threshold, then the outcome of this test is one class (e.g., "cloudy"); otherwise it is the other (e.g., "clear"). We will refer to X throughout this article as the "observable". Thresholds for the ocean tests are static and depend only on the sun-view geometry at the time of observation. Further detail on the RCCM-ocean algorithm and performance can be found in Zhao and Di Girolamo (2004). Over land, however, thresholds will also depend on location and time of year because of the variability of the surface. It was concluded before launch that static thresholds with the few observables that can be constructed from the MISR channels would lead to a poor RCCM over land (Di Girolamo, 1996). To overcome this problem, a dynamic thresholding approach was proposed where the thresholds would be a function of time in addition to surface type, solar zenith angle, relative azimuth angle and view angle. Specifically, the RCCM-land algorithm implemented at the Langley DAAC and described in Diner et al. (1999) divides the surface into 1580 contiguous surface types within the Cloud Screening Surface Classifications (CSSC) dataset, which is solely based on the WE1.4D version of Olson's global ecosystem database (NOAA-EPA Global Ecosystems Database Project, 1992). The solar zenith angle and relative azimuth angle are divided into 10 and 12 bins, respectively. A histogram of an observable is constructed for each camera, season, CSSC, solar zenith angle and relative azimuth angle bin. Thresholds are derived from these histograms using an automated threshold selection algorithm (ATSA) that is suitably fast for satellite data processing on a global scale. The thresholds are then used during the reprocessing of the MISR datasets. For the RCCM-land algorithm to work well, it is crucial to select a good observable used for the cloud detection test and a good ATSA.

The purpose of this article is to justify the choice of the observables and the ATSA used by the RCCM-land algorithm. In so doing, we undertake the largest comparison of ATSAs applied to cloud detection on satellite imagery. The choice of observable and ATSA may also be used to improve cloud detection for other satellite missions carrying similar channels to MISR. We emphasize that this article is not a validation of the RCCM-land algorithm. Validation is currently underway with updates available through the Langley DAAC's website.

This article is organized as follows. In Section 2, information is given on the data used in this research. Section 3 investigates the suitability of four candidate observables constructed from MISR data for cloud detection. Sixteen automated thresholding algorithms are selected from the literature and a comparison of their cloud detection performances as applied to the four candidate observables is done in Section 4. The conclusions are summarized in Section 5.

2. Data

2.1. MISR scenes

To test candidate observables and ATSAs on MISR data, the following criteria for scene selection are employed: (1) the selected scenes should represent typical snow-free surface types since the RCCM-land algorithm is meant to detect clouds over snow-free land surfaces, and (2) all chosen scenes must be partially cloudy (0% < cloud fraction < 100%), otherwise thresholds can not be uniquely determined. A total of 35 scenes were selected over the following six representative surface types: (1) Eastern Saharan Desert, (2) Saharan Grass, (3) Brazilian Tropical Rain Forest, (4) Southeast US Mixed Trees and Crops, (5) East China Farmland and Settlements, and (6) Argentinian Farmland and Settlements. The chosen scenes are listed in Table 1. All 35 scenes are 256 × 256 pixels in size to ensure the scene is of an uniform surface type as defined in the CSSC dataset, while at the same time providing enough data to form a statistically representative histogram. In order to match the 1.1 km resolution of the MISR cloud mask, the selected scenes have a pixel size of 1.1 km. All scenes were derived from MISR's nadir camera.

2.2. Reference cloud mask

A major problem facing satellite cloud detection is the lack of truth to compare against (e.g., Di Girolamo & Davies, 1997). According to the *Glossary of Meteorology* (American Meteorological Society, 2000), cloud is defined as "a visible aggregate of minute water droplets and/or ice particles in the atmosphere above the earth's surface". What constitutes "visible" depends on the observing instruments and many other observing conditions, such as sun-view geometry and surface type. An accepted practice in evaluating the performance of an automated cloud mask is to use an expert analyst to interpret the satellite imagery in terms of clear and cloudy regions (e.g., Berendes et al., 2004; Stowe, 1984). Although it is not practical to visually inspect the satellite images and label each pixel as clear or cloudy for the entire MISR mission, it is possible to do so on the limited number of scenes given in Table 1. In this study, we used visualization software developed at the Jet Propulsion Laboratory (JPL) called PixelLearn as a means of deriving a supervised reference cloud mask for assessing the quality of candidate observables and ATSAs.

PixelLearn is an interactive tool designed to assist the user in quickly creating discrete labels of all pixels in a multispectral image. The supervised classification algorithm used is a highly optimized implementation of Support Vector Machines (Cortes & Vapnik, 1995). This software uses as many as hundreds of features (observables) from each pixel, incorporating multiple spectral channels and neighboring pixels for context, to distinguish cloudy from clear (Garay et al., 2005; Mazzoni, 2005).

Table 1
The list of scenes used in this study

Scene no.	Surface type[1]	Cloud fraction[2]	Central latitude[3]	Central longitude[4]	Orbit no.[5]	Date
1	Eastern Sahara Desert	0.15	23.1	28.6	13048	Jun. 01, 2002
2	Eastern Sahara Desert	0.16	28.1	28.3	12684	May 07, 2002
3	Eastern Sahara Desert	0.10	26.9	26.5	12320	Apr. 12, 2002
4	Eastern Sahara Desert	0.49	26.9	23.4	12291	Apr. 10, 2002
5	Saharan Grass	0.36	8.1	25.3	7689	May 29, 2001
6	Saharan Grass	0.30	9.4	24.0	7791	Jun. 05, 2001
7	Saharan Grass	0.62	9.4	22.5	8126	Jun. 28, 2001
8	Saharan Grass	0.57	8.1	16.0	8068	Jun. 24, 2001
9	Saharan Grass	0.52	8.1	14.5	8170	Jun. 01, 2001
10	Brazilian tropical rain forest	0.90	0.6	−64.6	8144	Jun. 29, 2001
11	Brazilian tropical rain forest	0.86	0.6	−67.7	8348	Jul. 13, 2001
12	Brazilian tropical rain forest	0.92	0.6	−69.2	7751	Jun. 02, 2001
13	Brazilian tropical rain forest	0.64	−4.4	−70.0	8217	Jul. 04, 2001
14	Brazilian tropical rain forest	0.79	−4.4	−71.5	8319	Jul. 11, 2001
15	Brazilian tropical rain forest	0.85	−5.6	−73.3	8188	Jul. 02, 2001
16	Brazilian tropical rain forest	0.89	−4.4	−73.1	8421	Jul. 18, 2001
17	Brazilian tropical rain forest	0.95	−5.6	−74.9	8290	Jul. 09, 2001
18	Brazilian tropical rain forest	0.91	−3.1	−75.9	7926	Jun. 14, 2001
19	Brazilian tropical rain forest	0.68	−3.1	−75.9	8159	Jun. 30, 2001
20	Brazilian tropical rain forest	0.72	−5.6	−78.0	8028	Jun. 21, 2001
21	Brazilian tropical rain forest	0.88	−4.4	−63.8	8042	Jun. 22, 2001
22	Southeast US mixed trees and crops	0.30	33.1	−90.7	8058	Jun. 23, 2001
23	Southeast US mixed trees and crops	0.96	34.4	−91.9	8160	Jun. 30, 2001
24	Southeast US mixed trees and crops	0.84	34.4	−93.4	8029	Jun. 21, 2001
25	Southeast US mixed trees and crops	0.53	35.6	−96.1	7767	Jun. 03 2001
26	Southeast US mixed trees and crops	0.12	34.4	−96.5	8466	Jul. 21, 2001
27	East China farmland and settlements	0.44	33.1	119.3	7700	May 30, 2001
28	East China farmland and settlements	0.78	34.4	115.1	8472	Jul. 22, 2001
29	East China farmland and settlements	0.92	35.6	113.9	7642	May 26, 2001
30	East China farmland and settlements	0.40	34.4	113.5	8108	Jun. 27, 2001
31	East China farmland and settlements	0.30	35.6	113.9	8341	Jul. 13, 2001
32	Argentina farmland and settlements	0.67	−31.9	−57.9	7925	Jun. 14, 2001
33	Argentina farmland and settlements	0.18	−31.9	−57.9	8158	Jun. 30, 2001
34	Argentina farmland and settlements	0.67	−35.6	−60.6	8493	Jul. 23, 2001
35	Argentina farmland and settlements	0.96	−31.9	−62.6	8464	Jul. 21, 2001

[1]Name designation based on Olsen (NOAA-EPA Global Ecosystems Database Project 1992) dataset from which the MISR CSSC dataset was derived.
[2]Based on the reference cloud masks (see Section 2.2).
[3]Positive latitude represents northern hemisphere and negative latitude represents southern hemisphere.
[4]Positive numbers represent eastern longitude and negative numbers represent western longitude.
[5]Orbit number designation used by EOS-Terra.

Because PixelLearn allows man–machine interaction at the pixel level, we can make the results generated by this software as close to visual interpretation as possible by correcting the misclassified pixels. We subjectively estimate that the classification error between visual inspection and the final reference mask is <5% for all 35 scenes. As an example, Fig. 1 shows the RGB image and its reference cloud mask for Scene 5. The scene contains small cumulus, thin cirrus, and patches of bright surfaces that are difficult to distinguish from the red channel alone, all of which are detected very well, at least through visual inspection, by the reference cloud mask.

3. Quality of observables

3.1. Observable construction

The observable and the algorithm used in a dynamic thresholding scheme are two fundamental factors that determine the quality of the cloud mask. The quality of an observable includes two aspects: (1) observable sharpness: a good observable would correctly reflect the cloud field, such that a single threshold exists to separate clear and cloudy regions; and (2) threshold sensitivity: for a good observable, a small shift from the best possible threshold should not change the cloud fraction dramatically.

The number of observables that can be constructed from MISR data is limited because of the fact that it only has four channels, none of which measures wavelengths over 1 μm. We compare four candidate observables that can be constructed from the MISR channels and have been used in other cloud detection research. They are:

(1) R_{670}, the bidirectional reflectance factor (BRF) of the red channel, is defined at the top of the atmosphere as:

$$R_{670} = \frac{\pi L_{670}}{\mu_0 F_0}$$

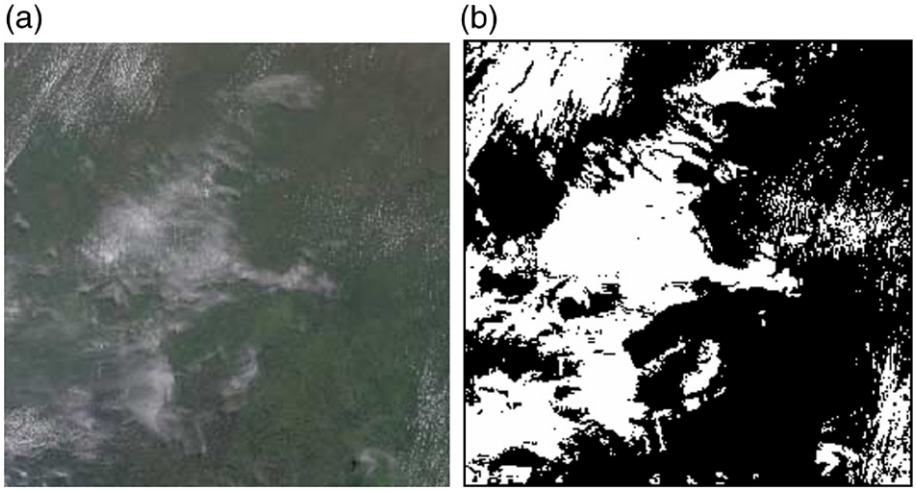

Fig. 1. (a) RGB image for Scene 5; (b) Reference cloud mask generated with PixelLearn.

where L_{670} is the spectral radiance recorded in the red band, μ_0 is the cosine of the solar zenith angle and F_0 is the solar spectral irradiance at the top of the atmosphere. Due to the low reflectance of the red channel over land, especially over vegetated land, the images from this channel usually have a relatively dark background when compared to clouds. For this reason, the BRF or other measures of reflectance in the red channel are widely adopted in cloud detection schemes (e.g., Ackerman et al., 1998; Baum & Trepte, 1999; Stowe et al., 1999).

(2) The Normalized Difference Vegetation Index (NDVI) is defined as:

$$NDVI = \frac{R_{865} - R_{670}}{R_{865} + R_{670}}$$

where R_{865} is the top of atmosphere BRF of the near infrared channel. Over most snow-free land surfaces, especially vegetated surfaces, $R_{865} > R_{670}$ (e.g., Liang, 2004), making NDVI > 0 over clear sky land. However, over clouds, R_{865} and R_{670} are very close to each other, making NDVI ~ 0. The NDVI or other similar vegetation indices that use R_{865} and R_{670} are popular choices for cloud detection (e.g., Ackerman et al., 1998; Lee et al., 2001; Stowe et al., 1999).

(3) The D observable is defined as: $D = |\mathrm{NDVI}|^b / R_{670}^2$. This observable was introduced by Di Girolamo and Davies (1995) for cloud detection, where the parameter b is chosen to maximize the separation between clear and cloudy pixels based on principle components analysis. b depends on the underlying surface type, which are coarsely divided into sparsely vegetated surfaces (e.g., deserts) with $b = 2.0$ and vegetated surfaces (e.g., rain forests) with $b = 0.65$. In this study, $b = 2.0$ is used for Scene #1–4 (Table 1) and $b = 0.65$ is used for the rest.

(4) The standard deviation (STDV) of the sixteen 275 m-resolution red channel pixels within a 1.1 km area is also a candidate for cloud detection. Using STDV as an observable is based on the assumption that the spatial variations associated with cloudy pixels are higher than that associated with clear pixels (e.g., Stowe et al., 1999).

3.2. Observable sharpness

For a given scene having a total of N pixels, let $N_d(t)$ be the number of pixels that are labeled differently by the reference ("truth") cloud mask and the cloud mask generated by applying threshold, t, to the observable for the scene. The best threshold, t_{best} for the scene is the one that minimizes $N_d(t)$:

$$t_{best} = \min_t \{N_d(t)\} \quad \forall\, t = f_{min}, \cdots, f_{max}$$

where f_{min} to f_{max} are the minimum and maximum values of the observable in the image, respectively.

The minimum cloud mask error, E_{min}, is simply:

$$E_{min} = \frac{N_d(t_{best})}{N}$$

If the image of an observable for a given scene has perfect sharpness (i.e., a threshold would exist that could separate all the cloudy pixels from the clear pixels), then E_{min} would be 0, which hardly happens in practice.

Fig. 2 shows the average E_{min} of all 35 selected scenes, as well as the average E_{min} of the scenes grouped by surface type,

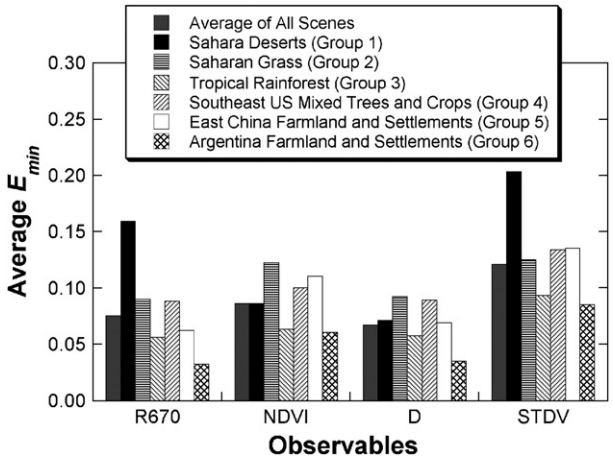

Fig. 2. Average minimum cloud mask error for the four observables.

for each of the four observables. Among the four observables, D gives the overall best result with an average $E_{min} = 6.7\%$, with all surfaces having an $E_{min} < 10\%$. The average E_{min} for R_{670}, NDVI and STDV are 7.5%, 8.6% and 12.1% respectively. The R_{670} makes a good observable over vegetated land (Groups 2 to 6), but performs poorly over the desert area (Group 1). The NDVI performs more poorly compared to R_{670}, except over desert surfaces. STDV is not good for cloud detection by itself because it gives large E_{min} for all surface types; however, we will see below that it has the potential for enhanced detection of small cumulus.

Contributions to E_{min} may come from three sources: (1) the reference masks themselves are not perfect as they are subjectively derived (see Section 2.2); (2) surface heterogeneity may cause clear pixels in one part of the image to be brighter (for R_{670} and STDV images) or darker (for D and NDVI images) than those thin or partially cloudy pixels in another part of the image (e.g., as discussed in Rossow & Garder, 1993), and (3) for STDV images, spatially uniform clouds and spatially heterogeneous surfaces exist in the same image. The latter two are shown in Fig. 3 for the scene displayed in Fig. 1, where we see that the misclassification comes from the thin cloud pixels, cloud edge pixels, small cumulus pixels and clear pixels over bright surface. However, unlike the other three observables, for which some small cumulus clouds go undetected, the thresholding result of STDV (Fig. 3(d)) does detect them. The undetected small cumulus clouds are shown as black dots on the middle and lower-right parts of Fig. 3(a), (b) and (c), while the misclassification is much smaller on the same parts of Fig. 3(d). This indicates that higher accuracy might be achieved by combining D and STDV.

3.3. Threshold sensitivity analysis

Threshold sensitivity refers to the cloud fraction change corresponding to a small shift to the best threshold t_{best}. Low sensitivity is desirable for a good observable. As the basis of sensitivity analysis and thresholding algorithm comparison, a 128-bin histogram of each observable is constructed for each scene in Table 1. To avoid the skewness caused by outliers, histograms are constructed by minimizing the dynamic range of the observable while keeping 98% of the data.

A sensitivity analysis is done on the 128-bin histogram by moving the thresholds away from the best threshold bin by bin

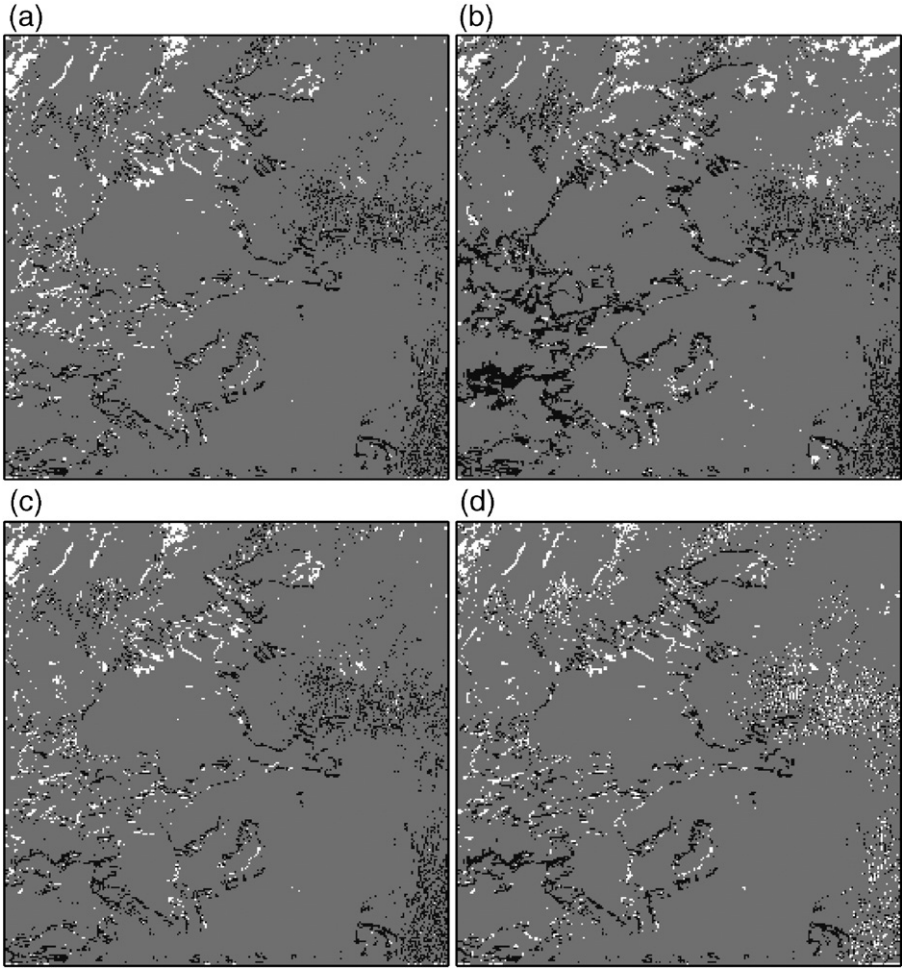

Fig. 3. Classification differences between the reference cloud mask and the cloud mask generated by applying t_{best} for Scene 5. The black points represent pixels where the reference cloud mask labels it cloudy while the thresholded image labels it clear. The white points are the other way around. Grey areas are where there is agreement. (a) R_{670}, (b) NDVI, (c) D, and (d) STDV.

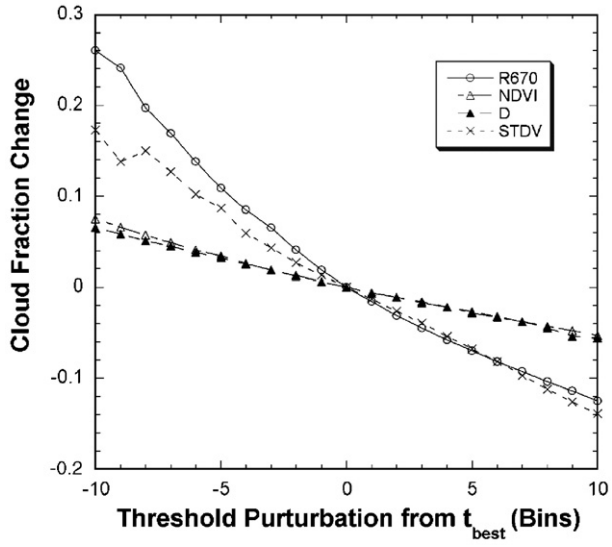

Fig. 4. Threshold sensitivity of the four observables averaged over all 35 selected.

in order to observe the corresponding change of cloud fraction. Fig. 4 gives the average result for all the scenes. The threshold sensitivity of each observable is represented by the slope of each individual line near t_{best}. A large slope represents a more dramatic change in cloud fraction accompanying a threshold shift. From Fig. 4, both D and NDVI show the quality of a good observable by having the lowest sensitivity among all the observables.

The sensitivity of the observables could be explained by the difference of their histogram shapes. Fig. 5 displays an example of how histogram shapes differ for different observables. It shows that D puts cloudy pixels tightly together on the left hand side of the histogram, which is the reason why it has the lowest threshold sensitivity. Physically, this property of D could be seen from its definition: $D=|\mathrm{NDVI}|^b/R_{670}^2$. For cloudy pixels, NDVI is small and R_{670} is large, which results in very small D values. On the other hand, for clear pixels NDVI is large and R_{670} is small, which results in large, variable D values. Thus the histogram of D is relatively flat around t_{best}, making it less sensitive to the choice of threshold.

Based on the analysis of E_{min} and threshold sensitivity, it would appear that D is better suited for cloud detection over snow-free land surfaces as compared to R_{670}, NDVI, and STDV. However, this is no guarantee that an automated threshold selection algorithm (ATSA) will work well with this observable, because the performance of an ATSA depends strongly on the data to which it is applied (e.g., Pal & Pal, 1993).

4. ATSA comparison

4.1. Histogram-based ATSAs

A desirable ATSA should select a threshold that minimizes the misclassification of clear and cloudy pixels. Theoretically, if the distributions of cloudy and clear pixels are known, then the optimal threshold could be determined according to Bayes' rule (e.g., Schowengerdt, 1997). In reality, however, only the combined distribution, the histogram, is known. Parametric techniques tend to fit the histogram with two distributions — one for clear and one for cloudy (e.g., Strahler, 1980). Non-parametric techniques, on the other hand, make no assumptions about the probability distribution and are often considered more robust (e.g., Schowengerdt, 1997). Although a huge number of algorithms exists in the published literature, few efforts have been spent on their evaluation (e.g., Zhang, 1996), and no research has been done on comparing their performance when applied to satellite cloud detection.

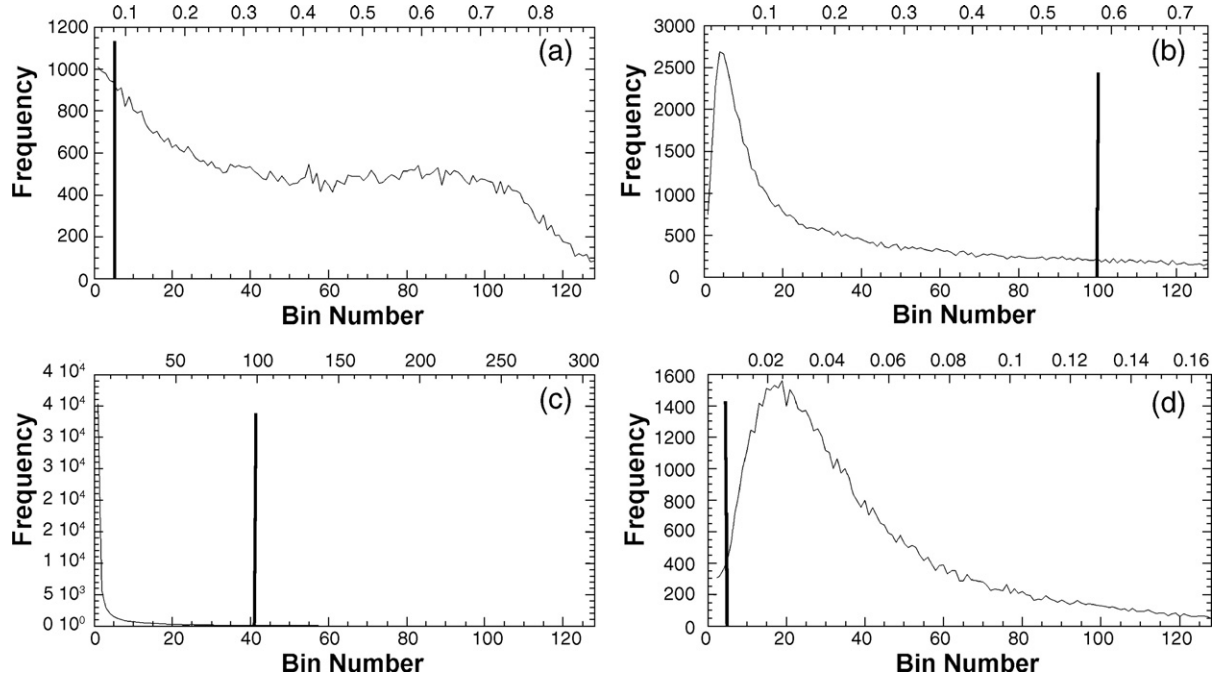

Fig. 5. Histograms of Scene 16 for (a) R_{670}, (b) NDVI, (c) D, and (d) STDV. The bold solid line in each frame is the best threshold for that observable.

Most of the popular ATSAs could be categorized into the following groups: minimum error method, entropy-based method, fuzziness theoretic method and discriminant analysis method (for a review, see Pal & Bhandari, 1993; Sahoo et al., 1988). The common feature of these methods is that all of them need to construct one or more criterion functions based on the histogram. Depending on the algorithm, the threshold is set in such a way that the criterion function(s) are minimized, maximized, or unchanged. Minimum error methods define the threshold as the point where the possibility of misclassification is minimum, which could be done by assuming the distributions of the background and the object or by finding the deepest concavity point of the histogram. Entropy-based methods originate from information theory. This group of methods set thresholds by either maximizing the total entropy of the background and the object or by minimizing the cross-entropy between the images before and after thresholding. Fuzziness theoretic techniques construct criterion functions based on fuzzy set theory, which could be done by minimizing the distance between the fuzzy image (before thresholding) and the crisp image (after thresholding). Discriminant analysis methods find thresholds by maximizing either the between class variance or the Euclidean distance of the two classes. There are some methods that do not fall clearly in any one of the above categories, such as the Tsai (1985) method, which finds thresholds by keeping the first three moments of the image unchanged, and the Yen et al. (1995) method, which selects thresholds by maximizing the total amount of correlation of the background and the object. Even though the criterion functions of these algorithms look very different, several of them are interrelated (Yan, 1996).

To find the appropriate automated thresholding procedure for MISR's RCCM scheme, we select representative algorithms from the literature for the comparison. The following criteria are employed in the selection: (1) the technique should be a histogram-based technique suitable for the MISR cloud detection approach and processing load, and (2) the technique should exist in the refereed literature within a full article. Based on this, 16 algorithms are chosen and they are listed in Table 2. The criterion functions used by these algorithms are summarized in the Appendix.

Table 2
Automated threshold selection algorithms selected for this study

Number	Method	Category
1	Otsu (1979)	Discriminant analysis
2	Rosenfeld and De La Torre (1983)	Minimum error
3	Kapur et al. (1985)	Entropy-based
4	Tsai (1985)	Moment preserving
5	Kittler and Illingworth (1986)	Minimum error
6	Whatmough (1991)	Minimum error
7	Pal and Dasgupta (1992)	Fuzziness theoretic
8	Li and Lee (1993)	Entropy-based
9	Pal and Bhandari (1993)	Minimum error
10	Huang and Wang (1995)	Fuzziness theoretic
11	Simpson and Gobat (1995)	Discriminant analysis
12	Yen et al. (1995)	Maximum correlation
13	Brink and Pendock (1996)	Entropy-based
14	Pal (1996)	Minimum error
15	Sahoo et al. (1997)	Entropy-based
16	Tobias and Seara (2002)	Fuzziness theoretic

4.2. ATSA performances

The ATSA and the observable to which the ATSA is applied are two key parts of a cloud detection scheme. An ATSA may work well on the images of certain observables and not work well on others (e.g., Pal & Pal, 1993). To compare their performances, each algorithm is applied to all four observables for each selected scene. The resulting cloud fraction of each automated threshold image is compared with the cloud fraction of their individual best threshold image. We did not compare the automated threshold image to the reference threshold image because we are testing the ability of each ATSA in putting thresholds at the best position for a given histogram.

Fig. 6 shows the average bias and the Root Mean Square (RMS) of the biases of the cloud fraction derived from the automated threshold when compared to the best threshold image of all 35 scenes. A positive bias means the algorithm tends to overestimate the cloud fraction while a negative one means the opposite. As shown in Fig. 6, all the algorithms tend to underestimate the cloud fraction when applied to the observables of R_{670}, NDVI and STDV when averaged over all 35 scenes. The results for D are generally better with the biases oscillating around zero. The best results would be the observable-ATSA combination that gives the smallest absolute bias with the smallest RMS values. Fig. 6 shows that ATSA No. 8 and 9, which are the Li and Lee (1993) method and the Pal and Bhandari (1993) method, applied to D give the best results with a mean bias $<1\%$ and an RMS $\sim 15\%$.

From the data studied here, the combination of either Li and Lee method (1993) or the Pal and Bhandari (1993) method with the D observable would be a desirable choice for an automated threshold selection scheme. However, the 15% RMS indicates that not every scene is equally satisfactory and other algorithms may perform better for certain scenes. The worst case for both the Li and Lee method and the Pal and Bhandaru method is Scene no. 26, which was taken over the surface type "Southeast US Mixed Trees and Crops" on July 21, 2001. Based on the reference cloud masks, this scene had the lowest cloud fraction of all scenes over vegetated surfaces. The cloud fraction bias is 53% for the Li and Lee method and 54% for the Pal and Bhandaru method. Fig. 7 gives its histogram and the positions of the best threshold and the threshold calculated from the Li and Lee method. Clearly, the Li and Lee method missed the signal of the first mode because of its narrowness, which caused a large overestimate of the cloud fraction of the scene. This low cloud fraction case may be indicating that an increase in cloudy samples over vegetated surfaces is required. Further discussion on this point is given in Section 5. We noticed that when applied to this case, ATSA #6 (Whatmough, 1991) and #14 (Pal, 1996) did a good job with cloud fraction biases 0% and 0.7% respectively, but the general performance of these two methods are not as good as the Li and Lee method and the Pal and Bhandaru method.

4.3. Separability analysis

To explain why ATSAs perform better when applied to images of D, a separability analysis is performed. There are

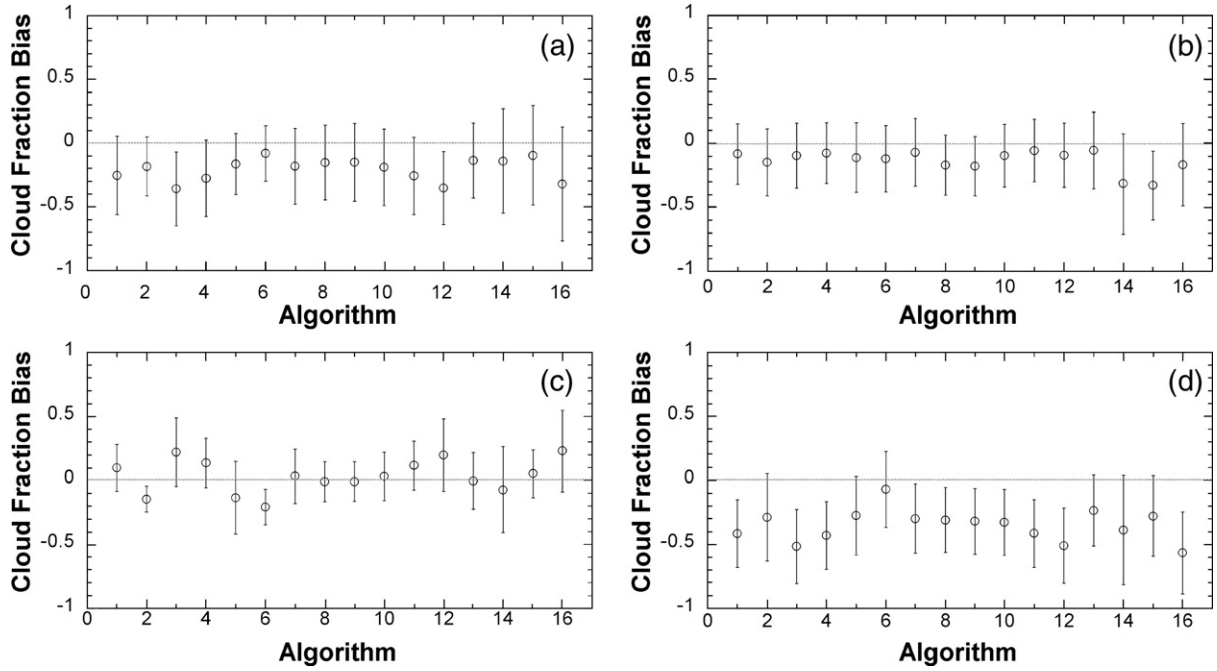

Fig. 6. Biases in cloud fraction for each ATSA averaged over the 35 scenes. The dotted horizontal line is the zero bias line. The error bars show the root mean square of the biases. Positive biases indicate that the ATSA tends to overestimate the cloud fraction while negative biases indicate that the ATSA tends to underestimate the cloud fraction. The results are for (a) R_{670}, (b) NDVI, (c) D, and (d) STDV.

many different ways to define separability (Schowengerdt, 1997). One of the simpler measures of separability, s, is,

$$s = |\mu_1 - \mu_2|$$

where μ_1 and μ_2 are the means of the two classes. In this study, μ_1 and μ_2 are the average bin numbers weighted by the frequency of data falling in each bin of the two classes, which are separated by the best threshold. Large values of s indicate the clear and cloudy pixels are on average well separated, which should allow for a good cloud mask using a single well placed threshold. Fig. 8 gives the results of the separability analysis of the four observables. The average of all scenes shows that D has the largest separability with a value of 65, followed by NDVI, R_{670}, and STDV. This result is consistent with the general performance of the ATSAs shown in Fig. 6.

Fig. 8 also gives the average separability of scenes over each individual surface type. Due to the low contrast between clear and cloudy pixels over desert areas for all observables, separability for desert scenes are generally lower than other groups. Again, D performs consistently better than other observables over all groups.

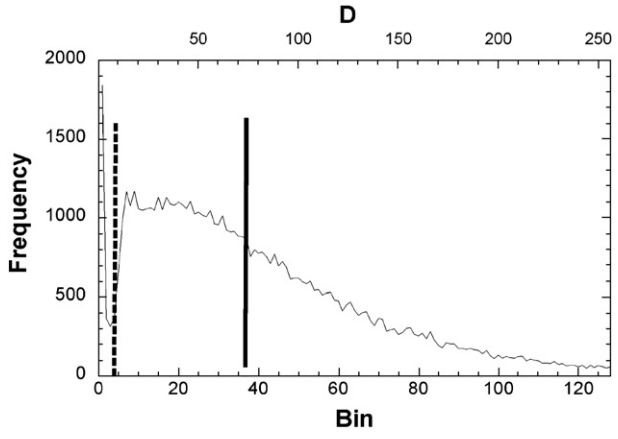

Fig. 7. The histogram of the D image for Scene no. 26, which was taken over the land surface of the Southeast US Mixed Trees and Crops on July 21, 2001. The bold dash line is the best threshold and the bold solid line is the threshold selected with the Li and Lee (1993) method.

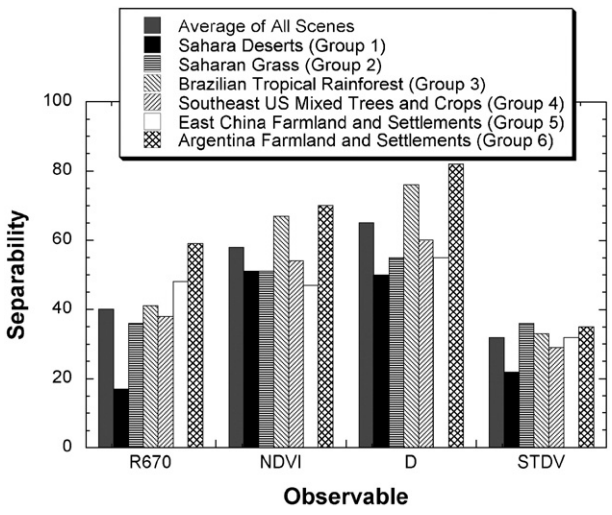

Fig. 8. Results of separability analysis.

The separability is directly related to the distribution of the clear and cloudy pixels. Theoretically, if a remotely sensed image consists of a homogeneous background and a homogeneous object, the distribution of the pixels, the histogram, could be considered as a combination of two Gaussian or Poisson distributions (e.g., Schowengerdt, 1997). In this case, the shape of the histogram is bimodal, in which one mode represents the background and the other represents the object. Most parametric thresholding algorithms are based on this assumption. However, for cloud detection over land, the heterogeneity of the surface and the cloud makes this assumption largely untrue. Of the 35 scenes examined in this study, only five of the R_{670}, 11 of the NDVI, 15 of the D, and six of the STDV histograms are bimodal. The low bimodal rate of R_{670} histograms indicates that the clear and cloudy modes are not well separated, while the higher bimodal rate of D histograms suggests a better separation of clear and cloudy modes.

5. Conclusion

In this article, a framework for ATSA and observable selection was established and used to justify their choices employed by the MISR RCCM over land. Even though this research is done specifically for the MISR RCCM-land algorithm, the result may be valuable to other satellite missions, especially those with instruments having only a few spectral solar channels. In this study, a thorough investigation is performed on four observables and 16 ATSAs on 35 scenes chosen from the MISR data over the Sahara Deserts, the Saharan Grass, the Brazilian Tropical Rainforest, the Southeast US Mixed Trees and Crops, the East China Farmland and Settlements and the Argentinian Farmland and Settlements. To the authors' knowledge, this represents the largest comparison of ATSAs, not only for cloud detection, but in general. The major conclusions of this study are as follows:

(1) Even if the best threshold was achieved for a single observable, there would be a wide discrepancy between reference cloud masks and the best thresholded images. If we neglect the small amount of error that the reference cloud masks may have, the discrepancy is mainly a result of confusion between bright areas of the surface and the thin or partially cloudy pixels. Among all four candidate observables (R_{670}, NDVI, D and STDV), D gives the overall best result with an average minimum cloud mask error 6.7%. Sensitivity analysis also shows that D is the best choice in that it is least sensitive to small changes from the best threshold.

(2) 16 ATSAs are selected for thresholding algorithm comparison. Among them, the Li and Lee (1993) method and Pal and Bhandari (1993) method, when applied to the D observable, performed the best. The average differences between the automated thresholded cloud masks and the best thresholded images derived from these two methods are less than 1% with an RMS ~ 15%. The superiority of D over the other observables was explained by the separability analysis, which showed that D observable has the largest separability among the four observables.

(3) It has been widely assumed in image thresholding that histograms are bimodal, in which one mode is for the background and the other is the object. This assumption is largely untrue in our study for clouds over snow-free lands. In this study, only 14% of the R_{670} histograms are bimodal. However, 43% of the D histograms are bimodal, which suggests a more clearly defined threshold could be achieved through the combination of information from different spectral channels.

(4) Even though the overall performance of the standard deviation (STDV) observable is not desirable, its ability of detecting small cumulus cloud is superior compared to the other three observables, which indicates that higher cloud detection accuracy might be achieved by combining D and STDV.

Based on the above results, the authors recommend using the Li and Lee or Pal and Bhandari method with D as the observable for the MISR RCCM-land algorithm. STDV needs to be considered in order to improve the detection of cumulus clouds. To avoid the situation where ATSA misses a small cloudy mode (cf. Fig. 7) due to an insufficient number of cloudy pixels over vegetated land, an increase in sampling area or sampling over time is recommended. These recommendations have already been implemented in Version 3.3 of the RCCM-land algorithm at the Langley DAAC, with the Li and Lee (1993) method for the ATSA. The details on how these recommendations were implemented (e.g., how to combine D and STDV for final decision on labeling a pixel clear or cloudy) is left to a separate article that the authors are currently preparing on the full RCCM-land algorithm and the evaluation of its performance over the globe.

Acknowledgements

Partial support from the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration is gratefully acknowledged. The data used in this study were obtained from the NASA Langley Research Center's Atmospheric Sciences Data Center.

Appendix A

This appendix provides only a brief review of the 16 automated threshold selection algorithms (ATSAs) examined in this study. A common feature of these ATSAs is that they construct one or several criterion functions for selecting a threshold based on a histogram of the observable. Thresholds are found by minimizing, maximizing or keeping these criterion functions unchanged. Only the criterion functions are defined in this appendix, with their derivation and rationale left to their source literature.

Let the number of observations within a histogram bin be denoted by h_i, $i \in \{1, \ldots, n\}$, where n is the total number of bins

that make up the histogram. In this review, the histogram is normalized and regarded as a probability mass function, p_i:

$$p_i = \frac{h_i}{\sum_{i=1}^{n} h_i}$$

Note that $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$.

The following mth order cumulative moments of the histogram are used routinely throughout this appendix:

$$M_k^m = \sum_{i=1}^{k} i^m p_i$$

$$\bar{M}_k^m = \sum_{i=k+1}^{n} i^m p_i$$

The goal of an ATSA is to select a threshold, $t \in \{1, \ldots, n\}$, that best divides the histogram into clear and cloudy categories.

1) Otsu (1979)

The Otsu method is based on discriminant analysis. The optimal threshold is selected by maximizing the between class variance, σ_B^2, given by:

$$\sigma_B^2(k) = \frac{[M_n^1 M_k^0 - M_k^1]^2}{M_k^0[1 - M_k^0]}$$

The threshold t is chosen when:

$$\sigma_B^2(t) = \max_{1 \leq k \leq n} \sigma_B^2(k)$$

2) Rosenfeld and De La Torre (1983)

This method requires the construction of the convex hull of the histogram:

$$r_i = \max\left[p_i, \max_{l,m}\left(p_l + \frac{i-l}{m-l}(p_m - p_l)\right)\right]$$

where any available (m, l) combination with $1 \leq l < i$, $i \leq m < n$ is included.

The threshold t is selected at the deepest concavity point where:

$$s(t) = \max_k [r_k - p_k]$$

3) Kapur et al. (1985)

This method chooses the threshold t by maximizing the sum of the entropy of the background and the object:

$$\psi(t) = \max_k [H_b + H_w]$$

where H_b and H_w are the entropy of the two categories:

$$H_b = -\sum_{i=1}^{k} \frac{p_i}{M_k^0} \ln \frac{p_i}{M_k^0}$$

$$H_w = -\sum_{i=k+1}^{n} \frac{p_i}{\bar{M}_k^0} \ln \frac{p_i}{\bar{M}_k^0}$$

4) Tsai (1985)

The Tsai method is based on the moment preserving principle in which the threshold is selected in such a way that the first three moments of the image are kept unchanged before and after the thresholding. The threshold t is set when:

$$M_t^0 = \frac{(z - M_n^1)}{(c_1^2 - 4c_0)^{1/2}}$$

where

$$z = \frac{1}{2}[-c_1 + (c_1^2 - 4c_0)^{1/2}]$$

$$c_0 = \frac{M_n^1 M_n^3 - (M_n^2)^2}{M_n^2 M_n^0 - (M_n^1)^2}$$

$$c_1 = \frac{M_n^0 M_n^3 - M_n^1 M_n^2}{M_n^2 M_n^0 - (M_n^1)^2}$$

5) Kittler and Illingworth (1986)

Based on the assumption that both of the background and the object are normally distributed, this method constructs a criterion function $J(k)$, which reflects the amount of overlap between the distribution of the background and the object:

$$J(k) = M_k^0 \ln \frac{\sigma_1}{M_k^0} + \bar{M}_k^0 \ln \frac{\sigma_2}{\bar{M}_k^0}$$

where

$$\sigma_1^2(k) = \left[\sum_{i=1}^{k} \left\{i - M_k^1 / M_k^0\right\}^2 p_i\right] \Big/ M_k^0$$

$$\sigma_2^2(k) = \left[\sum_{i=k+1}^{n} \left\{i - \bar{M}_k^1 / \bar{M}_k^0\right\}^2 p_i\right] \Big/ \bar{M}_k^0$$

The threshold t is chosen when:

$$J(t) = \min_k J(k)$$

6) Whatmough (1991)

This method could be regarded as a revised version of the Rosenfeld and De La Torre (1983) method. Instead of constructing the convex hull of the histogram, the exponential hull is constructed:

$$H_i = \max\left[p_i, \max_{l,m}\left\{p_l \exp\left\{\frac{i-l}{m-l}(\ln p_m - \ln p_l)\right\}\right\}\right]$$

Then threshold t is selected when:

$$r(t) = \max_k [H_k - p_k]$$

7) Pal and Dasgupta (1992)

This method constructs several fuzziness memberships, which represent different opinions, and calculates the total amount of ambiguity. The threshold is selected when the total amount of ambiguity is minimized.

In our calculation, the following fuzziness memberships are used:

$$\mu_1 = \begin{cases} 0 & i<a \\ \left(\frac{i-a}{c-a}\right)^2 & a\le i\le c \\ 1 & i>c \end{cases}$$

$$\mu_2 = \begin{cases} 0 & i<a \\ 1-\left(1-\frac{i-a}{c-a}\right)^2 & a\le i\le c \\ 1 & i>c \end{cases}$$

$$\mu_3 = \begin{cases} 0 & i<a \\ 2\left(\frac{i-a}{c-a}\right)^2 & a\le i\le k \\ 1-2\left(\frac{i-c}{c-a}\right)^2 & k<i\le c \\ 1 & i>c \end{cases}$$

$c-k = k-a = 4$ is used in the calculation

Let

$$\delta_{mn} = |\mu_m - \mu_n|$$

$$T_j = \min[\mu_j, 1-\mu_j]$$

The total amount of ambiguity is written as:

$$d(k) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\delta_{12}+\delta_{13}+\delta_{23}}{4} + \frac{T_1+T_2+T_3}{3}\right]p_i$$

The threshold t is chosen when:

$$d(t) = \min_k d(k)$$

8) Li and Lee (1993)

In this method, the expression of the cross-entropy between the original image and the thresholded image is:

$$\eta(k) = \sum_{i=1}^{k} i p_i \ln\left(\frac{i}{M_k^1/M_k^0}\right) + \sum_{i=k+1}^{n} i p_i \ln\left(\frac{i}{\bar{M}_k^1/\bar{M}_k^0}\right)$$

The threshold is selected by minimizing the cross-entropy:

$$\eta(t) = \min_k \eta(k)$$

9) Pal and Bhandari (1993)

This method is a revision of the Kittler and Illingworth (1986) method. Instead of assuming that the background and the object follow normal distributions, Pal and Bhandari assume they follow Poisson distributions. The criterion function is constructed as:

$$\begin{aligned} J(k) = {} & M_n^1/M_n^0 - M_k^0[\ln M_k^0 \\ & + (M_k^1/M_k^0)\ln(M_k^1/M_k^0)] - \bar{M}_k^0[\ln \bar{M}_k^0 \\ & + (\bar{M}_k^1/\bar{M}_k^0)\ln(\bar{M}_k^1/\bar{M}_k^0)] \end{aligned}$$

The threshold t is chosen when:

$$J(t) = \min_k J(k)$$

10) Huang and Wang (1995)

The Huang and Wang method defines threshold by minimizing the fuzziness index:

$$\begin{aligned} E(k) = {} & \sum_{i=1}^{k}[-\mu_A(i)\ln(\mu_A(i)) - (1-\mu_A(i))\ln(1-\mu_A(i))]p_i \\ & + \sum_{i=k+1}^{n}[-\mu_B(i)\ln(\mu_B(i)) - (1-\mu_B(i))\ln(1-\mu_B(i))]p_i \end{aligned}$$

$$\mu_A(i) = \frac{1}{1+|i-M_k^1/M_k^0|/C}$$

$$\mu_B(i) = \frac{1}{1+|i-\bar{M}_k^1/\bar{M}_k^0|/C}$$

where C is a constant value such that

$$\frac{1}{2}\le\mu_A(i)\le 1$$

$$\frac{1}{2}\le\mu_B(i)\le 1$$

In our study, we choose $C=n-1$.

11) Simpson and Gobat (1995)

Simpson and Gobat used a simple adaptive threshold selection procedure in their cloud detection effort. First, the average of the histogram is calculated as the initial threshold,

$$t = M_n^1/M_n^0$$

The means of the two groups of data created using this threshold are calculated. The average of the means is used as a new threshold:

$$t^{\text{new}} = [M_t^1/M_t^0 + \bar{M}_t^1/\bar{M}_t^0]/2$$

This process continues until the threshold remains constant.

12) Yen et al. (1995)

This method finds the threshold by maximizing the total amount of correlation provided by the thresholded image, which is defined as:

$$TC(k) = -\ln \sum_{i=1}^{k} \left(\frac{p_i}{M_k^0}\right)^2 - \ln \sum_{i=k+1}^{n} \left(\frac{p_i}{\bar{M}_k^0}\right)^2$$

The threshold is selected when:

$$TC(t) = \max_k TC(k)$$

13) Brink and Pendock (1996)

This is another method that uses the principle of minimizing cross-entropy to select threshold. They defined cross-entropy as:

$$h(k) = \sum_{i=1}^{k} p_i M_k^1 \ln\left(\frac{M_k^1}{i}\right) + \sum_{i=k+1}^{n} p_i \bar{M}_k^1 \ln\left(\frac{\bar{M}_k^1}{i}\right)$$

The threshold t is chosen when:

$$h(t) = \min_k h(k)$$

14) Pal (1996)

Pal used the assumption that both the background and the object follow the Poisson distribution. These distributions could be written as:

$$q_i^A = \frac{e^{-M_k^1/M_k^0}(M_k^1/M_k^0)^i}{i!} \quad i = 1, 2, \ldots, k$$

$$q_i^B = \frac{e^{-\bar{M}_k^1/\bar{M}_k^0}(\bar{M}_k^1/\bar{M}_k^0)^i}{i!} \quad i = k+1, k+2, \ldots, n$$

Different with Li and Lee's definition, Pal defines the cross-entropy as:

$$D(k) = \sum_{i=1}^{k} \frac{p_i}{M_i^0} \ln\left(\frac{p_i}{M_i^0 q_i^A}\right) + \sum_{i=1}^{k} q_i^A \ln\left(\frac{M_i^0 q_i^A}{p_i}\right) + \sum_{i=k+1}^{n} \frac{p_i}{\bar{M}_i^0} \log\left(\frac{p_i}{\bar{M}_i^0 q_i^B}\right) + \sum_{i=k+1}^{n} q_i^B \log\left(\frac{\bar{M}_i^0 q_i^B}{p_i}\right)$$

The threshold is chosen when:

$$D(t) = \min_k D(k)$$

15) Sahoo et al. (1997)

This method maximizes the total entropy of the background and the object, which is defined as:

$$H(\alpha, k) = \frac{1}{1-\alpha} \ln \sum_{i=1}^{k} \left(\frac{p_i}{M_k^0}\right)^{\alpha} + \frac{1}{1-\alpha} \ln \sum_{i=k+1}^{n} \left(\frac{p_i}{\bar{M}_k^0}\right)^{\alpha}$$

Different α values would result in different thresholds. In our calculation, three thresholds, t_1, t_2, and t_3, are computed using α values as 0.1, 0.9999 and 10 respectively. The optimal threshold t is set by combining t_1, t_2, and t_3 together with the following formula:

$$t = t_1\left[M_{t_1}^0 + \frac{\beta_1\omega}{4}\right] + t_2\frac{\beta_2\omega}{4} + t_3\left[1 - M_{t_3}^0 + \frac{\beta_3\omega}{4}\right]$$

$$\omega = M_{t_3}^0 - M_{t_1}^0$$

where

$$(\beta_1, \beta_2, \beta_3) = \begin{cases} (1,2,1) \text{ if } (|t_1-t_2| \le 5) \text{ and } (|t_2-t_3| \le 5) \\ (1,2,1) \text{ if } (|t_1-t_2| > 5) \text{ and } (|t_2-t_3| > 5) \\ (0,1,3) \text{ if } (|t_1-t_2| \le 5) \text{ and } (|t_2-t_3| > 5) \\ (3,1,0) \text{ if } (|t_1-t_2| > 5) \text{ and } (|t_2-t_3| \le 5) \end{cases}$$

16) Tobias and Seara (2002)

This method creates two fuzzy subsets, A and B, on each end of the histogram, $[1,j]$ and $[r,n]$. In our calculation, we use $j=10$ and $r=n-9$. The membership functions, μ_A and μ_B are:

$$\mu_A = \begin{cases} 0 & i < a \\ 2\left(\frac{i-a}{c-a}\right)^2 & a \le i \le b \\ 1 - 2\{i-c)/(c-a)\}^2 & b < i \le c \\ 1 & i > c \end{cases}$$

$$\mu_B = 1 - \mu_A$$

where

$$b = \frac{\sum_{i=i_{\min}}^{i_{\max}} i p_i}{\sum_{i=i_{\min}}^{i_{\max}} p_i}$$

$$c = b + \max\{|b - i_{\max}|, |b - i_{\min}|\}$$

$$a = 2b - c$$

and $i_{\max}$ and $i_{\min}$ are the upper and lower bound of the subset.

For a fuzzy set (X) with membership μ that has m supports, the index of fuzziness is defined as:

$$\psi(X) = \frac{2}{m} \sum_{i=1}^{m} |\mu_i - \bar{\mu}_i| p_i$$

$$\bar{\mu}_i = \begin{cases} 0, \text{if } \mu_i < 0.5 \\ 1, \text{if } \mu_i \geq 0.5 \end{cases}$$

For each $k \in [j+1, r-1]$, the threshold is selected when:

$$x(t) = \min_k [\psi(A \cup k) - \alpha\psi(B \cup k)]$$

where

$$\alpha = \frac{\psi(A)}{\psi(B)}$$

References

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C., & Gumley, L. E. (1998). Discriminating clear sky from clouds with MODIS. *Journal of Geophysical Research*, *103*, 32,141−32,157.

American Meteorological Society (2000). *Glossary of meteorology.* Allen Press 855pp.

Baum, B. A., & Trepte, Q. (1999). A grouped threshold approach for scene identification in AVHRR imagery. *Journal of Atmospheric and Oceanic Technology*, *16*, 793−800.

Berendes, T. A., Berendes, D. A., Welch, R. M., Dutton, E. G., Uttal, T., & Clothiaux, E. E. (2004). Cloud cover comparisons of the MODIS daytime cloud mask with surface instruments at the North Slope of Alaska ARM site. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(11), 2584−2593.

Brink, A. D., & Pendock, N. E. (1996). Minimum cross-entropy threshold selection. *Pattern Recognition*, *29*(1), 179−188.

Cortes, C., & Vapnik, V. (1995). Support–vector networks. *Machine Learning*, *20*, 273−297.

Di Girolamo, L., (1996). Detecting and interpreting clouds from satellite radiometric measurements with application to the Multi-angle Imaging SpectroRadiometer (MISR). PhD Thesis, McGill University, Montreal, Canada, 123pp.

Di Girolamo, L., & Davies, R. (1995). The image navigation cloud mask for the Multiangle Imaging SpectroRadiometer (MISR). *Journal of Atmospheric and Oceanic Technology*, *12*(6), 1215−1228.

Di Girolamo, L., & Davies, R. (1997). Cloud fraction errors caused by finite resolution measurements. *Journal of Geophysical Research*, *102*(D2), 1739−1756.

Diner, D. J., Beckert, J. C., Bothwell, G. W., & Rodriguez, J. I. (2002). Performance of the MISR instrument during its first 20 months in Earth orbit. *IEEE Transactions on Geoscience and Remote Sensing*, *40*, 1449−1466.

Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R. A., et al. (1998). Multi-angle Imaging SpectroRadiometer (MISR) — Instrument description and experiment overview. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(4), 1072−1087.

Diner, D. J., Di Girolamo, L., & Clothiaux, E. E. (1999). *MISR Level 1 Cloud Detection Algorithm Theoretical Basis. JPL D-11399, Rev. D.*

Garay, M., Mazzoni, D., Davies, R., & Diner, D. (2005). The application of support vector machines to the analysis of global datasets from MISR. *Proc. of the 4th Conf. on Artificial Intelligence Applications to Environmental Sciences, American Meteorological Society, San Diego, CA, January 2005*.

Huang, L. -K., & Wang, M. -J. J. (1995). Image thresholding by minimizing the measures of fuzziness. *Pattern Recognition*, *28*(1), 41−51.

Kapur, J. N., Sahoo, P. K., & Wong, A. K. C. (1985). A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, *29*, 273−285.

Kittler, J., & Illingworth, J. (1986). Minimum error thresholding. *Pattern Recognition*, *19*(1), 41−47.

Lee, W. -H., Kudoh, J. -I., & Makino, S. (2001). Cloud detection for the far east region using NOAA AVHRR images. *International Journal of Remote Sensing*, *22*(7), 1349−1360.

Li, C. H., & Lee, C. K. (1993). Minimum cross entropy thresholding. *Pattern Recognition*, *26*(4), 617−625.

Liang, S. (2004). *Quantitative remote sensing of land surfaces*.John Wiley and Sons 534pp.

Mazzoni, D. (2005). LibFeature: A software library for quickly generating feature vectors on the fly from structured data. *Proc. of the 8th Workshop on Mining Scientific and Engineering Datasets, 2005 SIAM Intl. Conf. on Data Mining, April 2005*.

NOAA-EPA Global Ecosystems Database Project (1992). Global Ecosystems Database Version 1.0 User's Guide, Documentation, Reprints, and Digital Data on CD-ROM. *U.S. DOC/NOAA National Geophysical Data Center, Boulder, CO* 36 pp.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics, SMC*, *9*(1), 62−66.

Pal, N. R. (1996). On minimum cross-entropy thresholding. *Pattern Recognition*, *29*(4), 575−580.

Pal, N. R., & Bhandari, D. (1993). Image thresholding: Some new techniques. *Signal Processing*, *33*, 139−158.

Pal, S. K., & Dasgupta, A. (1992). Spectral fuzzy sets and soft thresholding. *Information Sciences*, *65*, 65−97.

Pal, N. R., & Pal, S. K. (1993). A review on image segmentation techniques. *Pattern Recognition*, *26*, 1277−1294.

Rosenfeld, A., & De La Torre, P. (1983). Histogram concavity analysis as an aid in threshold selection. *IEEE Transactions on Systems, Man, and Cybernetics, SMC*, *13*(3), 231−235.

Rossow, W. B. (1989). Measuring cloud properties from space: A review. *Journal of Climate*, *2*(3), 201−213.

Rossow, W. B., & Garder, L. C. (1993). Cloud detection using satellite measurements of infrared and visible radiances for ISCCP. *Journal of Climate*, *6*, 2341−2369.

Sahoo, P., Soltani, S., Wong, A. K. C., & Chen, Y. C. (1988). A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, *41*, 233−260.

Sahoo, P., Wilkins, C., & Yeager, J. (1997). Threshold selection using Renyi's entropy. *Pattern Recognition*, *30*(1), 71−84.

Saunders, R. W., & Kriebel, K. T. (1988). An improved method for detecting clear sky and cloudy radiances from AVHRR data. *International Journal of Remote Sensing*, *9*(1), 123−150.

Schowengerdt, R. A. (1997). *Remote sensing models and methods for image processing.* Academic Press 522pp.

Simpson, J. J., & Gobat, J. I. (1995). Improved cloud detection in GOES scenes over the oceans. *Remote Sensing of Environment*, *52*, 79−94.

Stowe, L. L. (1984). Evaluation of NIMBUS-7 THIR CLE and air-force 3-dimensional nephanalysis estimates of cloud amount. *Journal of Geophysical Research*, *89*(ND4), 5370−5380.

Stowe, L. L., Davis, P. A., & McClain, E. P. (1999). Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for the Advanced Very High Resolution Radiometer. *Journal of Atmospheric and Oceanic Technology*, *16*(6), 656−681.

Strahler, A. H. (1980). The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment*, *10*, 135−163.

Tobias, O. J., & Seara, R. (2002). Image segmentation by histogram thresholding using fuzzy sets. *IEEE Transactions on Image Processing*, *11* (12), 1457−1465.

Tsai, W. -H. (1985). Moment-preserving thresholding: A new approach. *Computer Vision, Graphics, and Image Processing*, *29*, 377−393.

Whatmough, R. J. (1991). Automatic threshold selection from a histogram using the "exponential hull". *CVGIP: Graphical Models and Image Processing*, *53*(6), 592−600.

Yan, H. (1996). Unified formulation of a class of image thresholding techniques. *Pattern Recognition*, *29*(12), 2025−2032.

Yen, J. -C., Chang, F. -J., & Chang, S. (1995). A new criterion for automatic multilevel thresholding. *IEEE Transactions on Image Processing*, *4*(3), 370−378.

Zhang, Y. J. (1996). A survey on evaluation methods for image segmentation. *Pattern Recognition*, *29*(8), 1335−1346.

Zhao, G., & Di Girolamo, L. (2004). A cloud fraction versus view angle technique for automatic in-scene evaluation of the MISR cloud mask. *Journal of Applied Meteorology*, *43*(6), 860−869.