# Cloud Detection over Snow and Ice Using MISR Data

Tao Shi [*], Bin Yu [*], Eugene E. Clothiaux [†] and Amy J. Braverman [‡]

## Abstract

Clouds play a major role in Earth's climate and cloud detection is a crucial step in the processing of satellite observations in support of radiation budget, numerical weather prediction and global climate model studies. To advance the observational capabilities of detecting clouds and retrieving their cloud-top altitudes, NASA launched the Multi-angle Imaging SpectroRadiometer (MISR) in 1999, which provides data in nine different views of the same scene using four spectral channels. Cloud detection is particularly difficult in the snow- and ice-covered polar regions and availability of the novel MISR angle-dependent radiances motivates the current study on cloud detection using statistical methods.

Three schemes using MISR data for polar cloud detection are investigated in this study. Using domain knowledge, three physical features are developed for detecting clouds in daylight polar regions. The features measure the correlations between MISR angle-dependent radiances, the smoothness of the reflecting surfaces, and the amount of forward scattering of radiances. The three features are the basis of the the first scheme, called Enhanced Linear Correlation Matching Classification (ELCMC). The ELCMC algorithm thresholds on three features and the thresholds are either fixed or found through the EM algorithm based on a mixture of two 1-dim Gaussians. The ELCMC algorithm results are subsequently used as training data in the development of two additional schemes, one Fisher's Quadratic Discriminate Analysis (ELCMC-QDA) and the other a Gaussian kernel Support Vector Machine (ELCMC-SVM). For both QDA- and SVM-based experiments two types of inputs are tested, the set of three physical features and the red radiances of the nine MISR cameras. All three schemes are applied to two polar regions where expert labels show that the MISR operational cloud detection algorithm does not work well, with a 53% misclassification rate in one region and a 92% nonretrieval rate in the other region.

The ELCMC algorithm produces misclassification rates of 6.05% and 6.28% relative to expert labelled regions across the two polar scenes. The misclassification rates are reduced to approximately 4% by ELMCM-QDA and ELCMC-SVM in one region and approximately 2% in the other. Overall, all three schemes provided significantly more accurate results and greater spatial coverage than the MISR operational stereo-based cloud detection algorithm. Compared with ELCMC-QDA, ELCMC-SVM is more robust against mislabels in the ELCMC results and provide slightly better results, but it is computationally slower.

KEY WORDS: Cloud detection; Feature selection; Ice and snow; Multi-angle Imaging Spectro-Radiometer (MISR); Quadratic Discriminate Analysis (QDA); Remote sensing; Support Vector Machine (SVM); Thresholding.

---

[*] Department of Statistics, University of California, Berkeley, CA 94720-3860. Email: taoshi@stat.berkeley.edu, binyu@stat.berkeley.edu

[†] Department of Meteorology, The Pennsylvania State University, University Park, PA 16802. Email: cloth@essc.psu.edu

[‡] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109-8099. Email: Amy.Braverman@jpl.nasa.gov

# 1 Introduction

Clouds play a major role in the Earth's climate because of their ubiquitous presence and their ability to interact with Sun- (i.e., solar) and Earth- (i.e., terrestrial) generated radiation. They may warm the Earth surface by absorbing radiation emitted by the surface and then, in turn, emitting some of the energy associated with this radiation back to the surface. They may also cool the Earth by reflecting incoming solar radiation back to space. Uncertainties about these cloud and radiation interactions and their feedback on the global climate are among the greatest obstacles in understanding and predicting the Earth's future climate.

In polar regions, where surfaces are usually covered by snow and ice, detecting clouds and assessing their impact on the radiation budget of Earth is particularly difficult (Diner et al., 1999a). Some studies indicate that clouds tend to warm the surface-atmosphere system over highly reflective surfaces, but others suggest that clouds have a net cooling effect. Moreover, the situation is complicated by the presence of aerosols: small radiation absorbing and/or scattering particles in the atmosphere that tend to cool the surface below them by reducing the amount of solar radiation absorbed at the surface. If clouds are present, they obscure the surface and hinder, or prevent, the detection of aerosols from satellite sensors. We can only begin to untangle these complex atmospheric particle detection problems by having reliable cloud detection methods for polar scenes.

Cloud detection is particularly challenging in polar regions because snow- and ice-covered surfaces can be both brighter and colder than the clouds above them. Under these conditions, the traditional assumptions about the properties of clouds and their underlying surfaces are violated and detection of cloud contributions to satellite radiances becomes problematic. Satellite data sets obtained from the NASA Earth Observing System (EOS) may provide significant new information, but new cloud detection algorithms are also required to obtain proportional scientific return.

As a sensor on the NASA EOS Terra satellite, the Multi-angle Imaging SpectroRadiometer (MISR) provides a new source of information for identifying clouds over ice and snow (Diner et al., 1999a). The MISR sensor has a set of nine cameras that view scenes of Earth and its atmosphere from nine different directions and in four spectral bands in the visible and near-infrared portion of the electromagnetic spectrum. Viewing the atmosphere from multiple angles endows MISR with the capability of measuring elevations and angular radiance signatures of objects. The MISR operational stereo-based algorithm retrieves the heights of reflecting surfaces and compares them with the altitudes of the underlying surfaces. Reflecting surfaces with retrieved heights significantly above the terrain height are labelled as clouds. However, this algorithm does not perform efficiently in polar regions because it is incapable of detecting clouds close to surface and such clouds are often present in polar regions.

In this study we begin by exploring a MISR data set that was collected over the Arctic Ocean, Greenland and Baffin Bay on June 20, 2001. This data set is a good one to investigate because it contains typical polar surface scenes, such sea ice, snow-covered mountains and plateaus, as well as a variety of cloud types at several different altitudes. Moreover, the MISR operational stereo-based algorithm has a poor performance for this scene. Relative to expert labels, which we discuss next, the MISR stereo-based operational algorithm has a misclassification rate of 53.25% and leaves 29.92% of the data unclassified.

Using the MISRLEARN software tool developed at the Jet Propulsion Laboratory (Dominic

Mazzoni, personal communication), we hand labelled 60,111 radiance values as "clear" or "cloud" contaminated by jointly investigating data from MISR and the Moderate Resolution Imaging Spectroradiometer (MODIS), which has 36 spectral bands across the visible, near-infrared and thermal infrared regions of the electromagnetic spectrum and is also onboard the NASA EOS Terra satellite. Hand labelling of satellite imagery is the best way to generate a large database for validation of polar cloud detection algorithms because only a limited number of ground-based measurements have been obtained in polar regions to date. We subsequently use the expert labels to develop and test three cloud-detection schemes. We also labelled another data set collected over Greenland on July 17, 2002, and these labels are reserved as an independent validation set for the three methods that we developed.

Improving upon Linear Correlation Matching Classification (LCMC; Shi et al., 2002) algorithm, we first developed three physical features for cloud detection that are measures of the correlations between radiances from the nine MISR camera view directions, the smoothness of reflecting surfaces and the amount of forward scattering of radiation by clouds. The features contain information to separate snow- and ice-covered surfaces from both high- and low-altitude clouds. The first feature, a measure of the linear correlation between radiances from different MISR cameras, separates high clouds from low clouds and the surface. The second feature, a measure of the smoothness of a reflecting surface, is introduced to identify the homogenous ground surfaces, such as smooth ice flows, where linear correlations in the radiances from different angles are not informative because they are dominated by instrument noise. The last feature, a measure of forward scattering of radiation by clouds, separates low-altitude clouds from the Earth surface.

Using these three features, three cloud-detection methods are proposed. The first method is a simple threshold algorithm on the three features with automatically chosen thresholds; we call this method the Enhanced Linear Correlation Matching (ELCMC) algorithm. The ELCMC algorithm uses fixed thresholds for two of the features, while for the measure of forward scattering the threshold is obtained automatically by modelling distribution of the feature as a mixture of two one-dimensional Gaussian models. Since the classification boundaries between "cloudy" and "clear" pixels in the three-dimensional feature space are not rectangles, as implicity assumed in the ELCMC algorithm, we develop nonlinear and smoother classifiers to improve upon the ELCMC algorithm classification results. To this end we test Quadratic Discriminate Analysis (QDA) and Gaussian kernel Support Vector Machine (SVM) classifiers are tested as post-processors for improving ELCMC algorithm results. Since both QDA and SVM classifiers require training data, which are usually unavailable when new observations become available, we use the ELCMC algorithm classification results as training data for the QDA and SVM classifiers. We call these coupled approaches the ELCMC-QDA and ELCMC-SVM methods. In addition to testing QDA and SVM methods on our set of three features, we also applied them directly to the red radiances from the nine MISR cameras.

The rest of this paper is organized as follows. In section 2 we briefly review the MISR instrument and its operational stereo-based algorithm for cloud detection, explaining why the operational algorithm does not perform well in polar regions. We then describe the data that we used in this study (section 3). Three physical features for cloud detection using MISR observations are presented in section 4. And in section 5 we discuss the three cloud-detection methods, their application to the set of three features and the red radiances of the nine MISR cameras, and the misclassification rates that we obtained from them. We conclude (section 6) with a summary of our results and our plans for extending this work into an operational setting.

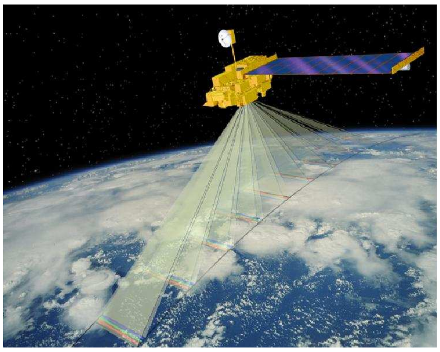

Figure 1: Cartoon illustration of both the NASA EOS Terra satellite and the view directions of the nine MISR cameras.

## 2 MISR and its Operational Algorithms for Cloud Detection

The MISR sensor (Diner et al, 1999a) is one of five instruments onboard the NASA EOS Terra satellite, which was launched into polar orbit on December 18, 1999. We briefly describe the MISR instrument and its operational cloud detection algorithms, highlighting the difficulties the operational algorithms have encountered in the polar regions.

### 2.1 Instrument Design

The MISR sensor consists of nine cameras, with each camera viewing Earth scenes at a different angle using four spectral bands (Figure 1). The view zenith angles of the nine cameras are $70.5^o$ (Df), $60^o$ (Cf), $45.6^o$ (Bf), and $26.1^o$ (Af) in the forward direction, $0.0^o$ (An) in the nadir direction and $26.1^o$ (Aa), $45.6^o$ (Ba), $60^o$ (Ca) and $70.5^o$ (Da) in the aft direction. The "f" in the letter designation of the cameras represents the "forward" direction, that is, in the direction of satellite flight, and the "a" represents the "aft" direction, that is, in the direction from which the satellite just came. The four spectral bands are at red, green, blue, and near-infrared (NIR) wavelengths. The nominal resolution of the MISR radiances is 275 m by 275 m at the surface of the Earth surface. To reduce MISR data rates the blue, green and NIR spectral radiances in the eight non-nadir cameras are averaged over $4 \times 4$ spatial arrays of pixels, for an effective spatial resolution of 1.1 km, before being transmitted to Earth.

The MISR cameras cover a swath at the Earth surface that is approximately 360 km wide and extends across the daylight side of the Earth from Arctic regions down to Antarctica. There are 233 geographically distinct, but overlapping, MISR swaths, which are also called paths, and MISR collects data from all of them on a repeat cycle of 16 days. For the MISR data products each path is subdivided into 180 blocks, with the block numbers increasing from the north to south pole. Each complete trip of the EOS Terra satellite around the Earth is given its own orbit number. The first orbit of the Terra satellite was labelled 1 and all orbits are numbered sequentially after this first orbit. Therefore, to identify the geographic location of MISR data products, we need to know the path number of the data products. To compute the date and time of a particular MISR radiance image along a certain path, we must know the EOS Terra orbit number.

## 2.2 MISR L2TC Cloud Height Retrieval Algorithm

Viewing the atmosphere from multiple angles, MISR has stereo capabilities that can be used to retrieve the elevation of objects on and above the surface of the Earth. The MISR Level 2 Top-of-Atmosphere Cloud (L2TC) algorithm uses this capability to detect clouds by comparing retrieved object elevations to the underlying terrain height, which is known from ancillary data. The rational of the L2TC algorithm is based on the registration of MISR measurements to a known reference ellipsoid. In the MISR registration process a MISR radiance can be registered to a surface ellipsoid, the terrain, or some other feature such as a cloud, as shown on the left side of Figure 2. The L2TC algorithm initially uses the ellipsoid projected data, as shown in the right side of Figure 2, in which clouds are registered at different locations in radiance maps from the different cameras. The L2TC algorithm matches the same object in the different angle-dependent images, allowing object heights and horizontal velocities to be retrieved through simple trigonometric relationships. Retrieved object heights are compared with the known terrain height and objects more than approximately 650 m above the terrain height are classified as clouds. The cloud mask derived from the L2TC algorithm is called the Stereoscopically Derived Cloud Mask (SDCM).

As we just mentioned, disparities in the location of an object in the different angle-dependent images can be produced by the object either being located above the reference ellipsoid or by movement of the object between the different camera views of the object. The L2TC algorithm uses the MISR Df-Bf-An camera triplet for cloud top motion retrieval (estimation) and the Af-An camera pair for cloud height retrieval (estimation). The L2TC algorithm estimates cloud motion and height in two separate stages, although they can be solved with two linear equations simultaneously after a match is found. We briefly review the algorithm here, as algorithm details can be found in Diner et al. (1999b).

High (275-m) resolution red radiances are used in the L2TC retrieval because these radiances are available from all nine cameras and they lead to the most accurate results. The radiances are transformed into Bi-directional Reflectance Factors (BRFs), i.e., the ratio of the observed radiance at a particular illumination and view geometry to the radiance that would be obtained under the same conditions with a target consisting of a perfect lambertian reflector (cf. Diner et al, 1999b). The first step in the L2TC algorithm is to retrieve cloud-top motion over non-overlapping regions of 70.4 km × 70.4 km. The coarse resolution of the resulting cloud-top motion vectors is a necessity of computational speed. To this end the L2TC algorithm uses a Nested Maxima (NM) stereo-matcher to find objects in An-Bf-Df camera triplet images. The NM matcher feature of interest is defined as a vertical column of five pixels such that the BRF decreases monotonically in value for two pixels on both sides of the center pixel. Since clouds are usually brighter than the underlying surface, the NM matcher features are usually from cloudy pixels if clouds are present in the 70.4 km × 70.4 km block. Note, however, that this assumption is not strictly valid for detecting cloud over snow- and ice-covered surfaces since these surfaces are also bright.

For each 70.4 km × 70.4 km block, the NM matcher searches for all nested maxima in the reference (Bf) angle image, as well as the An and Df camera images to be searched for matches to the reference image. In the next step nested maxima from the reference image and the search images are tested in pairs to determine if they are the same feature (Diner et al, 1999b). If no matches are found, a label of “no retrieval” is given to the block and hence no pixels in the block will have cloud-top motion, or cloud-top height, retrievals associated with them. If there are several matches in a block with inconsistent cloud-top motion estimates, a “no retrieval” label will again

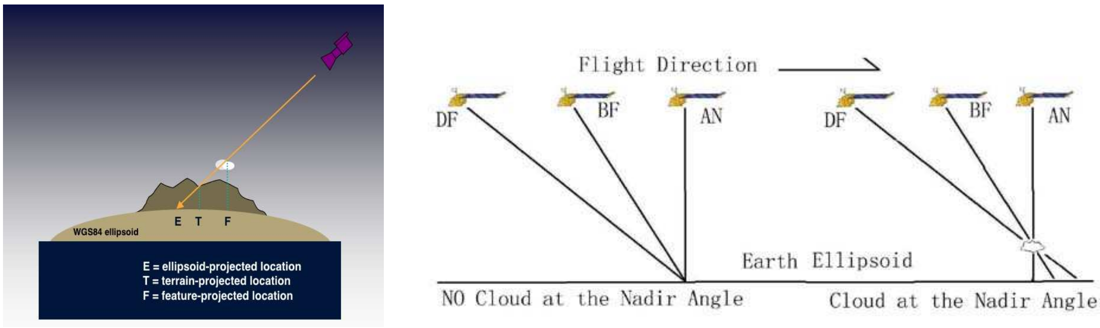

Figure 2: Left: Schematic of the three possible projections of MISR-camera radiance data, which include ellipsoid- (E), terrain- (T) and feature- (F) projected locations. Over the oceans, the ellipsoid and terrain projections are identical. Right: Registration of surface features and clouds to the reference ellipsoid. Note that only three of the nine MISR cameras (i.e., the Df, Bf and Af cameras) are illustrated and that surface objects in areas of low, smooth topography are mapped to the same location on the reference ellipsoid while clouds are mapped to different locations on it.

be given to the block. If a match, or self-consistent set of matches, is found between the reference and search images, the L2TC algorithm uses the following two equations to determine cloud-top motion, together with a preliminary estimate of cloud-top height:

$$v_c(t_{\mathrm{An}} - t_{\mathrm{Bf}}) - h(tan(\theta_{\mathrm{Bf}}) - tan(\theta_{\mathrm{An}})) = (x_{\mathrm{An}} - x_{\mathrm{Bf}}) \tag{1}$$

$$v_c(t_{\mathrm{Df}} - t_{\mathrm{Bf}}) - h(tan(\theta_{\mathrm{Bf}}) - tan(\theta_{\mathrm{Df}})) = (x_{\mathrm{Df}} - x_{\mathrm{Bf}}) \tag{2}$$

where $v_{\mathrm{c}}$ is the along-track cloud-top motion speed, $h$ is the cloud-top height, $t$ is the time when the given camera looks at that location, $\theta$ is the view zenith angle, $x_{\mathrm{An}} - x_{\mathrm{Bf}}$ and $x_{\mathrm{Df}} - x_{\mathrm{Bf}}$ are the along-track location shifts of the best matched features in the An/Bf and Df /Bf camera pairs.

In the second stage of L2TC processing cloud-top motion, $v_{\mathrm{c}}$, for the block is assumed known and An/Af image pairs are used to obtain cloud-top height on a pixel-by-pixel basis. To this end the L2TC algorithm matches features in the image pairs and incorporates their disparities, as well as $v_{\mathrm{c}}$, into equation (1) above to retrieve cloud-top height. In the feature matching between the Af/An image pairs the L2TC algorithm matches small patches around each pixel in the An image with the same sized patches around each pixel in a search window within the Af image. The tests for identifying matches are the same as those used in the first stage (Diner et al, 1999b).

There are two characteristics of the L2TC algorithm that reduce its effectiveness in the polar regions. First, low clouds are often present in daytime Arctic regions and the L2TC stereo matching algorithms together with the MISR data do not have sufficient accuracy to separate these clouds from the underlying surface. Second, the NM matcher has difficulties matching cloudy pixels between the images over snow- and ice-covered surfaces, which can be as bright, if not brighter, than the overlying clouds. When the NM matcher fails, the L2TC retrievals for the entire block fail. Another problem for the NM matcher, as well as the matching procedures in the cloud-top height retrievals, is lack of texture in some snow- and ice-covered surfaces as well as in some Arctic stratus clouds. These problems with the L2TC algorithm for cloud detection in polar regions motivated our current study, in which we investigate the information content in the MISR angular radiances for separating clouds from snow- and ice-covered surfaces.

# 3 Data

The data that we analyzed in this study consist of 23 orbits collected in 2001 and 2002, all from MISR path 26. Blocks 13 through 33 of this path cover the Arctic Ocean to the north and east of Greenland, northern Greenland, Baffin Bay and Baffin Island. Because each block contains approximately 30,000 valid 1.1-km radiance values, the total study area is about 700,000 km$^2$. We begin the analysis with MISR radiance values from blocks 16 through 33 of orbit 7898, which were obtained on June 20, 2001. This data set is a good one to initiate the study because it contains sea ice, snow-covered mountains and plateaus, and a variety of cloud types at several different altitudes (e.g., Figure 4).

To initiate our studies we analyzed in detail MISR blocks 19, 20 and 21 of orbit 7898 (referred as DATA-20JUNE2001 in this paper), labelling 60,111 1.1-km pixels as either "cloudy" or "clear" with high confidence. Assignment of the labels to the pixels was based not only on the MISR multi-angle radiances but also on the multi-spectral data from the Moderate Resolution Imaging Spectroradiometer (MODIS), which has 36 spectral channels and is also part of the EOS Terra satellite platform. Because high quality ground-based measurements are sparse in the polar regions, expert labelling of clear and cloudy scenes in satellite data is the best way to produce validation data for assessing automated polar cloud detection algorithms. In addition to orbit 7898, we also labelled data from MISR blocks 25, 26 and 27 of orbit 13723 which occurred on July 17, 2002 (referred as DATA-17JULY2002 as shown in Figure 12). We use this second set of labelled data to validate independently the performance of our proposed methods for cloud detection in polar regions. We use the remaining 21 orbits to better understand the surface types for path 26 and to assess visually the performance of our new methods when applied to a variety of scene types.

The dominant scene types for DATA-20JUNE2001 are snow-covered mountains, sea ice and three different cloud decks, one cloud deck being quite close to the surface, the other containing intermediate level clouds in a compact shape, and the third containing relatively high clouds. As illustrated in Figure 4, the snow-covered mountains occur at the bottom of the image with sea ice apparent in the upper two-thirds of the image. Note that the elevations of the coastal mountains decrease to sea level and that the two flows of ice in the bottom third of the image, which join as they approach the sea, are visibly brighter than the snow-covered mountains. While clouds and snow- and ice-covered surfaces have approximately the same brightness in the MISR An-camera image, clouds are noticeably brighter than the snow- and ice-covered surfaces in the Df-camera image. There are three types of clouds in the scene. The high clouds are to the right and bottom of the images and form a solid bright area in the Df-camera image. At the top of the images, towards the right, is the intermediate altitude cloud deck. A patch of low, thin clouds, almost transparent in the An-camera image, covers a significant portion of the sea ice on the left hand side of the image. These clouds are readily apparent in the Df-camera image.

We illustrate the expert labels for DATA-20JUNE2001 in Figure 9, where the blue colored areas represent "cloudy" pixels and red areas represent "clear" pixels. There are 60,111 pixels labelled with high confidence for this image, of which 38,161 are "cloud" pixels and 21,950 are "clear" ones. The black areas between the red and blue areas represent regions where no labels were assigned. Compared to the complicated scene in DATA-20JUNE2001, the area covered by DATA-17JULY2002 is much simpler (Figure 12), with several well-defined patches of thin and thick cloud cover over the Greenland plateau.

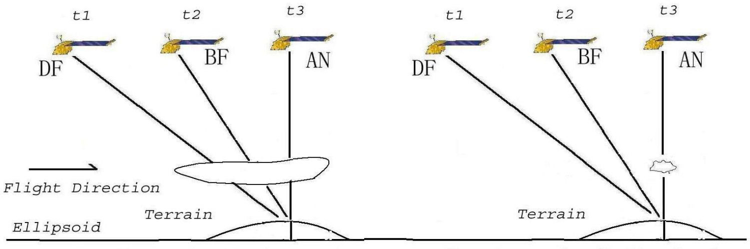

Figure 3: Left: Projection of MISR radiances to the surface terrain for a cloud with large spatial extent; note that radiances from the different parts of the cloud are mapped to the same spot on the terrain. Right: Projection of MISR radiances to the surface terrain for a cloud with small spatial extent; note that radiances from different objects are now mapped to the same spot on the terrain.

The MISR L2TC algorithm does not work well on these two scenes, with a 53.25% misclassification rate against the expert labels and a 29.92% non-classification rate for DATA-20JUNE2001 and a 91.82% non-classification rate for pixels in DATA-17JULY2002. For DATA-20JUNE2001 misclassification mostly occurs in regions of thin, low-altitude clouds, for which matching of objects in the different camera views is extremely difficult, if not impossible. The distinct differences in the MISR radiances from the nadir to the most forward looking MISR camera motivated our development of three physical features for detecting clouds in polar regions.

# 4 Feature Selection

In our current approach for detecting clouds in polar regions, we analyze the statistical properties of radiances scattered from clouds and snow- and ice-covered surfaces, rather than using object disparities in the different camera images to identify clouds. We focus on the scattering properties of snow- and ice-covered surfaces under clear skies. In the terrain projected MISR product (L1B2-Terrain) all of the MISR camera radiances are registered to the same location on the surface, allowing us to investigate easily the angular dependence of the radiances from the same reflecting surface. Since all four MISR bands have similar spectral signatures over ice, snow and clouds and only MISR red band radiances have 275 $m$ spatial resolution for all angles, we use the MISR red band data in this study. We develop three features to distinguish clouds from their underlying snow- and ice-covered surfaces. The features are the linear correlation in the radiances from the different MISR cameras, smoothness across the reflecting surfaces, and the magnitude of forward scattering from the reflecting surface. In section 4.1 we motivate our choice of these three features, while in section 4.2we provide the details on their computation.

## 4.1 Rational for the Three Features

The first feature is the linear correlation in the MISR radiances measured at different view angles. In previous work Shi et al. (2002) developed a simple algorithm, called Linear Correlation Matching Classification (LCMC), for detecting high altitude clouds over snow and ice surfaces. The LCMC

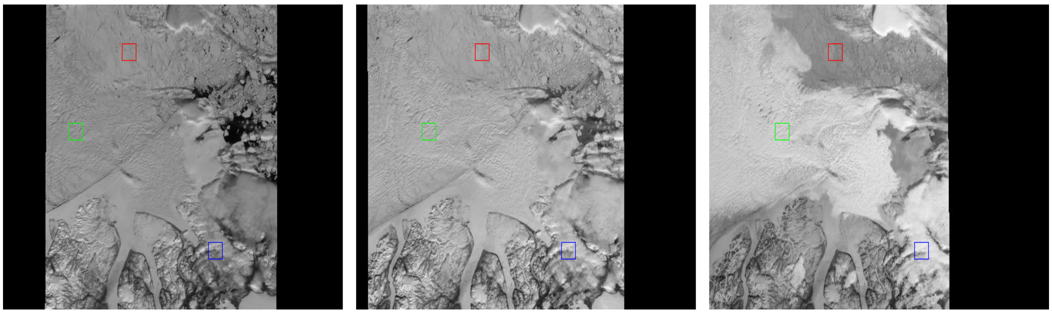

Figure 4: MISR red radiance images for blocks 19–21 of orbit 7898, i.e., images of Data-20June2001. Left: An-camera red radiance image. Center: Bf-camera red radiance image. Right: Df-camera red radiance image. The red, green and blue boxes delineate clear-sky, low-altitude cloud and high-altitude cloud regions, respectively.

algorithm capitalizes on the fact that spatial correlations between radiances observed at different angles in clear scenes are different than those in cloudy scenes. Specifically, for clear scenes the radiances measured by different MISR cameras are, in many cases, approximately proportional to the intensity of the light source, namely the sun. Therefore, they should be strongly correlated. However, for cloudy scenes this is not the case as the MISR radiances from the same cloud patch are now registered to different locations in both the terrain- and ellipsoid-projected MISR radiance products. As a result, the linear correlations in the radiances at the same projected locations are generally lower than for clear sky as they originate from different parts of clouds.

For example, consider the clear, high cloud and low cloud regions indicated by the red, blue and green boxes, respectively, in Figure 4. Producing scatter plots of the An- and Bf-camera radiances for an 8 by 8 array of pixels in each of these boxes leads to the results illustrated in Figure 5. Note the high degree of correlaton in the radiances from the surface and low-altitude cloud regions compared to the high-altitude cloud region. As this plot illustrates, LCMC produces low correlations for high clouds but is unable to distinguish reliably low clouds from the surface. Moreover, LCMC has difficulties over homogeneous scenes, such as smooth glacial ice flows, where the variations in the radiances across the scene are quite small, with subsequent poor linear correlations in the radiances from different MISR cameras.

To separate low-altitude clouds from rough surfaces, as well as smooth surfaces from clouds, we need additional features. To identify smooth surfaces we use the standard deviation of the An-camera radiances over a small window. This second feature will have small values for radiances emanating from "clear" homogeneous surfaces such as the glacial flows towards the bottom of Figure 4. For such smooth features the correlations between radiances from different angles may approach zero, being dominated by instrument noise. To demonstrate this point more clearly let $(X_{\mathrm{An}}, X_{\mathrm{Bf}})$ denote the measured An- and Bf-camera radiances for one pixel within a window of pixels. Suppose $X_{\mathrm{An}} = \mu_{\mathrm{An}} + \epsilon_{\mathrm{An}}$ and $X_{\mathrm{Bf}} = \mu_{\mathrm{Bf}} + \epsilon_{\mathrm{Bf}}$, where random variables $\mu_{\mathrm{An}}$ and $\mu_{\mathrm{Bf}}$ are the true An- and Bf-camera radiances and $\epsilon_{\mathrm{An}}$ and $\epsilon_{\mathrm{Bf}}$ are instrument noise. We assume $corr(\mu_{\mathrm{An}}, \mu_{\mathrm{Bf}}) = \rho$ and $var(\mu_{\mathrm{An}}) = var(\mu_{\mathrm{Bf}}) = \phi^2$. Furthermore, we assume $\epsilon_{\mathrm{An}}$ and $\epsilon_{\mathrm{Bf}}$ are independent of $\mu_{\mathrm{An}}$, $\mu_{Bf}$, and each other with mean 0 and variance $\sigma^2$. Then, the correlation between $X_{\mathrm{An}}$ and $X_{\mathrm{Bf}}$ is $CORR(X_{An}, X_{Bf}) = \rho/(1 + (\sigma/\phi)^2)$. Therefore, when the scene variability, $\phi^2$, is small relative to the instrument noise, $\sigma^2$, the observed correlation between the camera radiances is

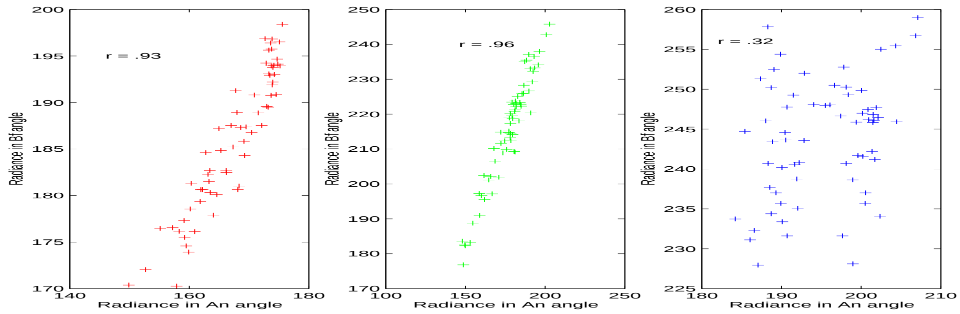

Figure 5: Scatter plots showing correlations between the red radiances of the MISR Bf and An cameras for clear scenes (left), scenes with low-altitude clouds (center), and scenes with high-altitude clouds (right). The correlations are computed from 8 by 8 arrays of radiances located within the red, green and blue boxes in Figure 4.

weaker than the true correlation $\rho$. In polar regions conditions with small scene variability appear to be indicative of cloud-free views of ice and snow, motivating our choice of the standard deviation, $SD_{An}$, of the An-camera radiances in a small window as an additional feature.

To detect low-altitude clouds we use the Normalized Differential Angular Index ($NDAI$), originally proposed by Nolin et al. (2002) to characterize the angular scattering properties of snow and ice surfaces. As we have shown, low-altitude clouds are not easily distinguishable from clear scenes using linear correlation because the radiances from low-level clouds may be registered to the same location on the reference surface, just as for clear-sky regions (Figure 5a,b). However, low-level cloudy regions tend to be much brighter than clear regions in the forward scattering direction Df angle images (Figure 4c). The difference in brightness between zenith and forward scattering results from clouds tending to scatter more photons in the forward direction, whereas snow- and ice-surface scenes scatter photons relatively equally in all directions.

For daylight Arctic regions forward scattering is recorded by the MISR forward pointing cameras, and especially the MISR Df-camera view. Therefore, for the Arctic regions we define $NDAI$ as $NDAI = (X_{\rm Df} - X_{\rm An})/(X_{\rm Df} + X_{\rm An})$. In the scenes that we have analyzed to date $NDAI$ is large for cloudy areas and small for clear snow and ice surfaces. In principle, $NDAI$ could be used to detect high-altitude homogeneous clouds with large spatial extents (Figure 3a). High clouds of small spatial extent and high clouds with significant inhomogeneities in them are problematic (Figure 3b), as now the relevant Df- and An-camera radiances associated with them are variable and are registered to distant locations on the terrain and ellipsoid reference surfaces. However, in conjunction with the linear correlation and surface smoothness tests high-altitude clouds should be separable from other cloud types as well as a variety of different surfaces. In fact, our working hypothesis for this study is that these three features are sufficient to identify and distinguish most surface and cloud types.

To illustrate separation of “cloudy” and “clear” scenes in the $CORR$, $SD$ and $NDAI$ feature space, we consider the features from the clear-sky and low- and high-altitude cloud regions illustrated in Figure 4. Using methods detailed in section 4.2, we computed values for the three features for these three regions and present them as two- and three-dimensional scatter plots in Figure 6. As Figure 6 clearly illustrates, the points occupy disjoint regions in the space of three features.

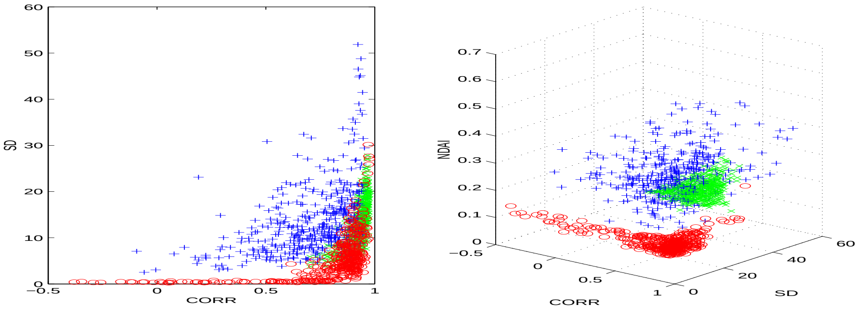

Figure 6: Scatter plots of $SD$, $CORR$ and $NDAI$ for the data in Figure 4. Note that the colors of the scatterplot symbols match the color of the boxes in Figure 4from which they come. Left: Two-dimensional scatterplot of $SD$ versus $CORR$. Right: Three-dimensional scatterplot of $SD$, $CORR$ and $NDAI$.

Moreover, synergy between the features is apparent, including the decrease in $CORR$ when $SD$ decreases.

### 4.2 Calculating Three Features

The data of interest in our study are blocks of MISR Level 1B2 terrain projected red band radiances from all nine cameras with a nominal spatial resolution of 275 m. The data are formatted in an $n$ by $m$ by 9 matrix $X$, where $n$ and $m$ are the number of rows and columns in a block and 9 corresponds to the nine MISR view angles. In this notation $X_{:,:,1}$ denotes the block of Df-camera radiances. Our goal is to compute features for each 1.1 km by 1.1 km region and to build classifiers at this same resolution based on the features. A schematic of the spatial representation of one MISR block of radiances, with one 1.1 km by 1.1 km region of 16 radiances to be classified indicated in the northwest part of the image, is illustrated in Figure 7a.

Since the Arctic region is the focus of this study, we use radiances from the MISR An, Af, Bf, and Df cameras to calculate the features. There are two reasons for using the MISR forward viewing cameras. First, for daylight Arctic scenes the MISR forward viewing cameras capture photons that are scattered in the forward direction. Second, radiances with large magnitudes are more effective in our linear correlation feature and radiances in the forward scattering direction are generally larger than those in the backward scattering direction. To compute the linear correlation of the radiances between camera views, we assign an 8 by 8 window around each target region composed of 4 by 4 pixels at 275 m resolution. We then reformat the 8 by 8 set of radiances in the An-, Af- and Bf-camera images for this target area into three 64 by 1 vectors. We use these vectors to compute the linear correlation coefficients between both the An- and Af-camera radiances, which is labelled as $Cor_{\mathrm{An,Af}}$, and the An- and Bf- camera radiances, which is labelled $Cor_{\mathrm{An,Bf}}$. For the target region we define $CORR = (Cor_{\mathrm{An,Af}} + Cor_{\mathrm{An,Bf}})/2$.

We use the average of a pair of linear correlation coefficients because the cloud can be registered to the same ground location in two view directions if the cloud movement is along the line of MISR flight. Having the MISR orbit, flight speed, and view angles just right to make the registration of a cloud in three images the same is extremely small. We use MISR Af- and Bf-camera radiances, instead of MISR Cf- and Df-camera radiances, because geo-registrations of the MISR Af and Bf

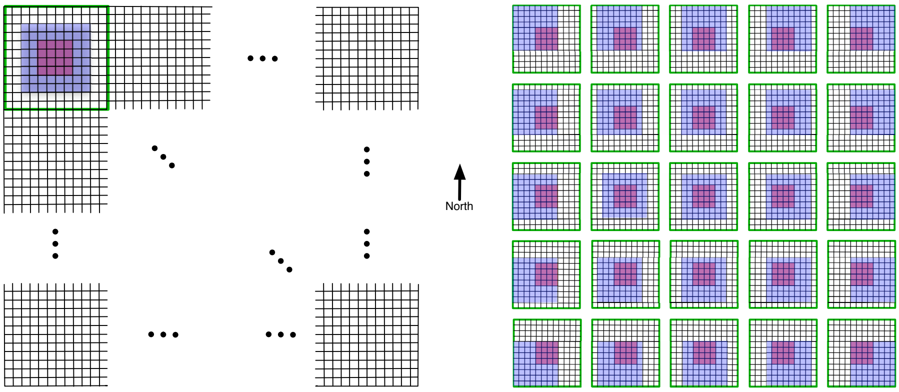

Figure 7: Left: Schematic representation of a block of MISR data. Each MISR block is a rectangular spatial array of 512 rows and 2048 columns of 275-m spatial resolution pixels. The red pixels in the figure define a 4 by 4 sub-array of pixels to be classified as “clear” or “cloudy”. The blue pixels define an 8 by 8 sub-array of pixels centered on the 4 by 4 target sub-array, while the green box defines a 12 by 12 sub-array of pixels centered on the target sub-array. Right: Locations (in blue) of a single 8 by 8 sub-array of MISR Df-camera radiances that are linearly correlated with the underlying 8 by 8 sub-arrays of MISR An-camera radiances in the computation of $NDAI$. The 4 by 4 sub-array of An-camera radiances to be classified is shown in red. Note that the single set of 8 by 8 MISR Df-camera radiances always covers the An-camera target region, but cycle over all possible 8 by 8 sets of MISR An-camera radiances contained within the 12 by 12 sub-array of An-camera search-window radiances.

cameras are more accurate than for the Cf and Df cameras. Misregistered pixels reduce linear correlations between camera images in the absence of clouds and should be avoided. If more than 16 of the 64 radiance values are missing in a target window of any camera view, we do not compute linear correlation coefficients for this window, assigning to it a missing value flag instead.

To compute standard deviations for the An-camera radiances we again use an 8 by 8 window of radiances centered on the 4 by 4 pixel region of interest (Figure 7a). We then compute the standard deviation as $SD_{\text{An}} = \sqrt{\sum_{i=1}^{k}(x_i - \bar{x})^2/(k-1)}$, where $x_{\text{i}}$, with $i \leq 64$, denotes the radiances and $\bar{x}$ is the sample mean. Similar to the computation of $CORR$, if the number of missing values in an 8 by 8 window exceeds 16, we assign a missing value to $SD_{\text{An}}$. We expect $SD_{\text{An}}$ to be small when there is no cloud over a very smooth land surface, e.g., a frozen river or a high-altitude plateau covered by packed snow.

The third feature is a measure of the forward scattering properties of the scene viewed by MISR. After Nolin et al. (2002), we compute the Normalized Differential Angular Index ($NDAI$) as $\mathbf{NDAI} = (\bar{X}_{\text{Df}} - \bar{X}_{\text{An}})/(\bar{X}_{\text{Df}} + \bar{X}_{\text{An}})$. Registration of the Df camera is generally less accurate than for the An camera. So, in the computation of $NDAI$ we we first set the location of the 8 by 8 window of radiances to the same target location in the An- and Df-camera images (e.g., blue pixels in Figure 7a). Since the misregistration of the Df-camera image relative to the An-camera image is no more than two pixels, and usually must less than this, we shift the Df-camera window of 8 by 8 pixels by up to ±2 pixels to find the best match of the Df-camera window to the An-camera window. This leads to 25 different locations in the Df-camera image that we match to each An-camera target region (Figure 7b). We use the minimum linear correlation coefficient across these 25 locations to identify the best Df-camera location and we compute $NDAI$ using it.

# 5 Methodology and Results

The three physical features that we use in this study are motivated by their different information content for clear- and cloudy-sky conditions. A natural method for separating clouds from the surface using these features is by setting threshold values on them. As illustrated in Figure 6, the three-dimensional scatter plot shows that "cloudy" pixels have larger $NDAI$ values than "clear" pixels. The two-dimensional scatter plot of $CORR$ and $SD_{\mathrm{AN}}$ shows that "clear" pixels are clustered close to two lines, represented as $CORR = 1$ and $SD_{\mathrm{AN}} = 0$, while "cloudy" pixels are located farther away from these two lines. So, properly chosen thresholds will, to some extent, separate "clear" and "cloudy" pixels in the three-dimensional space of our features.

After investigating the distributions of the three features using scenes from different orbits and visually inspecting results obtained from different thresholds, we conclude that thresholds for $CORR$ and $SD_{\mathrm{AN}}$ are stable and robust across different scenes and under different sunlight conditions. However, appropriate thresholds for $NDAI$ vary a great deal from scene to scene. These results motivate our setting the thresholds for $CORR$ and $SD_{AN}$ to empirically determined fixed values of $threshold_{\mathrm{CORR}} = 0.8$ and $threshold_{\mathrm{SD}} = 2$. To determine the threshold for $NDAI$ we automatically fit a one-dimensional mixture Gaussian model to $NDAI$ values computed from three adjacent blocks of data using the Expectation/Maximization (EM) algorithm (see details in section 5.1). We call the combination of three thresholds thus develop the Enhanced Linear Correlation Matching Classification (ELCMC) algorithm, which is designed to operate on every set of three MISR blocks of data.

Inspecting Figure 6, we see immediately that the class boundaries between "clear" and "cloudy" are not rectangles, as in the ELCMC algorithm. To go beyond the thresholding method in the ELCMC algorithm, we investigate Quadratic Discriminate Analysis (QDA) and Gaussian kernel Support Vector Machine (SVM) approaches to construct curvier and smoother decision boundaries, in the hope of improving upon the results obtained from the ELCMC algorithm. Both QDA and SVM methods require training data, which are unavailable when new scenes are encountered. Therefore, in the current study we use ELCMC classification results as training data for the QDA and SVM methods. In this paradigm we label our two approaches as the ELCMC-QDA and ELCMC-SVM methods.

Misclassified pixels in the ELCMC results are now problematic in that they now contaminate the training data for the QDA and SVM methods. To ameliorate this problem note that SVM methods are more robust than QDA when the training data contain errors because the hinge loss function used in SVM is an L1-type loss (see details in section 5.3). Consequently, as we will see in section 5.3.2, a SVM approach does lead to slightly better classification results against expert labelled data than QDA. Keep in mind, though, that the QDA method is computationally cheaper and thereby more suited for on-line processing of MISR data in the generation of cloud masks in polar regions.

## 5.1 The Three-feature Based ELCMC Algorithm

The underlying logic of the ELCMC algorithm is straightforward. To illustrate this point, again consider the two- and three-dimensional scatter plots in Figure 6 for the clear-sky surface (red), low cloud (green) and high cloud (blue) regions in Figure 4. For clear-sky regions with underlying smooth surface features $SD_{\mathrm{An}}$ is generally less than 2, while for rough surface features and clouds it usually greater than 2. To separate rough surface features from clouds we use the $CORR$ and $NDAI$ features. We note that $CORR$ is generally greater than 0.8 for rough surface features and low clouds and less than 0.8 for high clouds, while $NDAI$ is greater for low clouds than for rough surface features. In the three-dimensional scatterplot of Figure 6, a threshold of 0.2 is sufficient to separate rough surfaces from low clouds. For all 22 orbits of MISR path 26 that we have inspected to date, we find similar results.

These observations motivate the following sequence of tests for separating clear from cloudy regions:

**ELCMC Algorithm**:
A 1.1 km by 1.1 km target region is claimed "clear" when
· $SD_{\mathrm{AN}} < threshold_{\mathrm{SD}}$ **OR**
· $CORR > threshold_{\mathrm{CORR}}$ **AND** $NDAI < threshold_{\mathrm{NDAI}}$.
When the above tests fail, the region is labelled as "cloudy."

The algorithm claims "clear" for two types of situations: small $SD_{\mathrm{AN}}$ suggests smooth ice or snow surfaces, on which both linear correlation and forward scattering tests can fail; large $CORR$ and small $NDAI$ together suggest no clouds above the snow and ice surface. To determine the best thresholds, we investigated all 23 orbits of MISR data by visually comparing the results from different thresholds with images of the observed radiances. We find that the best thresholds for $SD_{\mathrm{An}}$ and $CORR$ are very stable across different locations. The threshold for $SD_{\mathrm{An}}$ is mainly related to the instrument noise level of the An camera, which changes little with the magnitude of the radiance and from one location to the next. Thus, $threshold_{\mathrm{SD}}$ is set to 2 for all scenes. The linear correlation between MISR camera views depends on both the ratio of the radiance to the level of the instrument noise (i.e., the signal to noise ratio) and the accuracy of MISR geo-registration. Both factors do not change much from one location to the next under nominal satellite platform operations, so we set $threshold_{\mathrm{CORR}}$ to a fixed value of 0.8. Although there are small scene and time changes in the optimal thresholds on $SD_{\mathrm{AN}}$ and CORR, these changes are small relative to values that we provide above.

Compared to the stability of $threshold_{\mathrm{SD}}$ and $threshold_{\mathrm{CORR}}$, the best $threshold_{\mathrm{NDAI}}$ varies across a large range of values, usually falling between 0.08 and 0.40. The magnitude of forward scattered radiation depends on at least two factors and both of these two factors vary across location and time. One factor is the exact angle between the line of sight of a MISR camera to a reflecting surface and the line from the sun to the same reflecting surface, which is mostly determined by the location of the NASA EOS Terra satellite and the height of the reflecting surface. The other factor is the size and shape of the particles that form clouds as the directionality of scattering by clouds is affected by these properties of cloud particles. Since both factors are unknown before data processing, $threshold_{\mathrm{NDAI}}$ should be chosen in a data driven manner.

We employ a method to select $threshold_{\mathrm{NDAI}}$ based on modelling the distribution of $NDAI$ values across three MISR blocks as a mixture of two Gaussian distributions, one for the "cloudy" area and the other for the "clear" area. The choice of two Gaussian distributions to model the $NDAI$ distribution is made for two reasons: two Gaussians fit the distribution of $NDAI$ well and the computation of mixed Gaussian fits by the EM algorithm is easier than fitting a mixture of other distributions, such as Laplacian distributions. Three blocks of MISR data are modelled together, representing a compromise between ensuring both "cloudy" and "clear" areas within the analysis region and separability of $NDAI$ values for "cloudy" and "clear" areas. Although modelling more

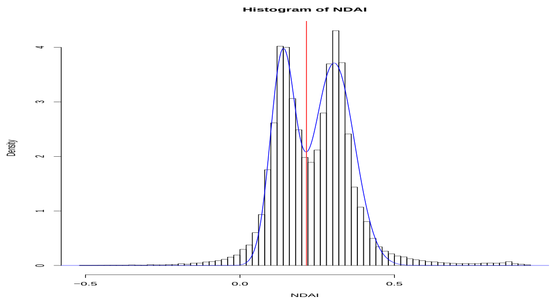

Figure 8: Histogram (black lines) of $NDAI$ from Data-20June2001 together with the fit (blue line) of the two one-dimensional Gaussian functions mixture model to it. The vertical red line indicates the $NDAI$ threshold derived from this histogram.

than three MISR blocks of data together provides a greater probability of having both ”clear” and ”cloudy” scene types in the block range, the distributions of $NDAI$ from both “cloudy” and “clear” areas in extended block ranges may change as a result of changes in the angles between the illumination and view directions to the reflecting surface.

The one-dimensional mixture of two Gaussian distributions is fitted by an EM algorithm and the dip in the fitted density is taken as $threshold_{\mathrm{NDAI}}$, assuming, of course, that there is a dip in the distribution between 0.08 and 0.40. Before applying the EM algorithm on $NDAI$ distributions, the upper and lower 2.5% of data are trimmed off to avoid outliers and extreme values in both tails, thereby improving robustness of the fitting scheme. To find the dip in the fitted Gaussian distribution functions we set the derivative of the distribution function to zero and solved for a solution between the means of the two Gaussian distributions in the mixture. When no dip is found within the expected range of $NDAI$ values from 0.08 and 0.40, the threshold from either the previous orbit or the next orbit is selected if one of the two is available. If neither of these two thresholds exists, we use the average of all available thresholds for this path and range of blocks. Finally, we use the dip in the fit of the two Gaussians to the distribution, instead of Bayes decision rule, because the dip is more robust than Bayes decision rule when the empirical distribution has a heavier tail than a mixture Gaussian model. Support for this choice is evident in 22 orbits that we inspected.

An example of a $NDAI$ distribution and our model fit to it is illustrated in Figure 8. These results are generated using MISR radiances from DATA-20JUNE2001. For this case the dip in the fit is quite clear, occurring at a value of 0.215. Using $threshold_{\mathrm{SD}} = 2.0$, $threshold_{\mathrm{CORR}} = 0.80$ and $threshold_{\mathrm{NDAI}} = 0.215$, we obtain a misclassification rate of 6.05% against the expert labelled data. To test the validity of these thresholds, we search for the best thresholds for this particular data set using a small proportion, approximately 10%, of randomly selected expert labels. The best thresholds we find are $threshold_{\mathrm{SD}} = 2.2$, $threshold_{\mathrm{CORR}} = 0.76$, and $threshold_{\mathrm{NDAI}} = 0.210$ with a misclassification rate of 5.82%. Most of the change is a result of the lower value for $threshold_{\mathrm{CORR}}$.

To illustrate the ELCMC algorithm performance for DATA-20JUNE2001 we present the Df-camera image, the expert labels and ELCMC results in Figure 9. The black areas on the left and right margins of Figure 9c are not classified because only the central area in this image has data from the MISR Af, An, Bf and Df cameras. Relative to the ”clear” and ”cloudy” labelled

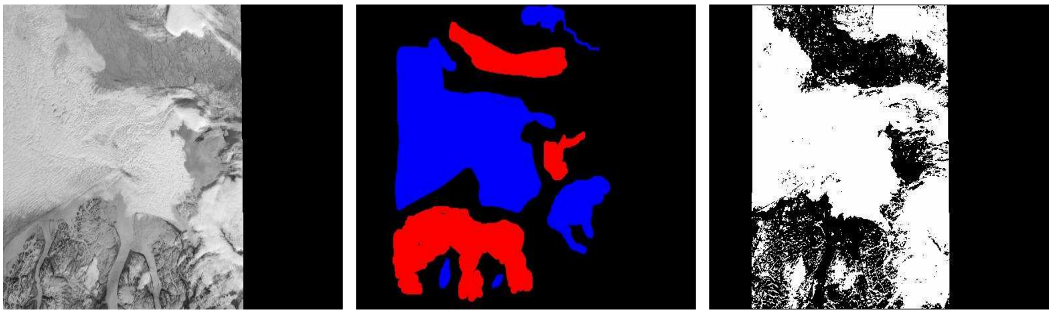

Figure 9: ELCMC algorithm cloud-detection results for Data-20June2001. Left: Df-camera red radiance image. Center: Image of expert labels, where red labels indicate clear pixels, blue labels indicate cloudy pixels, and black labels indicate pixels ignored in the labelling process. Right: ELCMC algorithm classification results, where white labels indicate cloudy pixels.

pixels, the algorithm has a misclassification rate of 6.05%. For expert labelled “clear” areas the misclassification rate is 15.09%, while for ”cloudy” areas the misclassification rate is only 1.14%. That is, the algorithm readily identifies clouds while classifying many pixels in clear regions of rough terrain as cloudy. For example, towards the bottom left of Figure 9, the expert labels show that there are no clouds over this mountainous area. However, in this region the ELCMC algorithm claims some pixels as “cloudy”. An analysis of these false positive ”cloudy” pixels indicates that they are labelled ”cloudy” as a result of the linear correlation test. We subsequently trace this result to slight misregistration problems between the different MISR camera images. Unfortunately, in polar regions terrain height information is sometimes inaccurate, leading to MISR registration errors when the MISR data are projected to the terrain. These errors reduce linear correlations between different MISR camera images for these scenes resulting in the misclassifications that we have found.

In summary, the ELCMC algorithm in this exploratory data analysis is accurate to approximately 94% in separating cloud from surface features. Computations of the $NDAI$ thresholds by fitting a two Gaussian distribution model to the data is simple and fast. But any misregistration errors in the MISR camera images used in the linear correlation test are going to cause some ”clear” pixels of rough terrain to be misclassified as ”cloudy” pixels. To circumvent these problems we explore two nonlinear classifiers, i.e., QDA and SVM, in an attempt to improve upon the ELCMC algorithm. We also investigate $\log(SD_{\text{AN}})$, instead of $SD_{\text{AN}}$, as a feature because the distribution of $SD_{\text{AN}}$ is heavy tailed and has a much larger range than both the $CORR$ and $NDAI$ features. Moreover, QDA models data as Gaussian distributions that are symmetric, and distributions of $log(SD_{\text{AN}})$ tend to be more symmetric than those for $SD_{\text{AN}}$. For training of SVM algorithms the elements of the input vectors should have comparable magnitudes, so $log(SD_{\text{AN}})$, which has a range of values closer to those of $CORR$ and $NDAI$, is more appropriate for this application as well. Finally, we apply QDA and SVM methods directly to the red radiances from all nine MISR cameras to test the value of our three-feature set.

## 5.2 Quadratic Discriminate Analysis Classifiers

The ELCMC algorithm divides the three-dimensional feature space into rectangular regions. However, Figure 6 shows that the boundary between the "clear" and "cloudy" points is nonlinear and does not particularly follow any one of the three coordinate axes. Fisher's Quadratic Discriminate Analysis (QDA) supplies a quadratic classification boundary that is expected to be more accurate than simply thresholding on each feature. Implementation of QDA requires training data to construct the boundary (cf. Mardia et. al. 1979; Ripley 1996). We use three blocks of MISR data to train and test QDA for the same reasons in the development of the ELCMC algorithm, namely as a trade-off between ensuring "clear" and "cloudy" pixels in each scene and enhancing separability amongst the different features.

In a two class classification problem, QDA models each class density as a multivariate Gaussian distribution:

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_{\mathrm{k}}|^{1/2}}\ e^{-\frac{1}{2}(x-\mu_{\mathrm{k}})^T\Sigma_{\mathrm{k}}^{-1}(x-\mu_{\mathrm{k}})},$$

where $k = 1, 2$ denotes the class label. Let $\pi_{\mathrm{k}}$ be the prior probability of class $k$. The a posterior distribution for $x$ belonging to class $k$ is then given by

$$P(x \in \text{Class } k|X = x) = \frac{f_{\mathrm{k}}(x)\pi_{\mathrm{k}}}{f_1(x)\pi_1 + f_2(x)\pi_2}.$$

The classification rule of QDA is to place $x$ in the class that has the largest a posterior probability at $x$. Using training data, parameters $\pi_{\mathrm{k}}$, $\mu_{\mathrm{k}}$, and $\Sigma_{\mathrm{k}}$ are estimated by the empirical class proportions, means and variances, and subsequently substituted into the above two equations to form the classifier.

To test the value of our three-feature set relative to the MISR radiances, we apply the QDA method to two sets of features, one consisting of our three-feature set $(log(SD_{\mathrm{AN}}), CORR, NDAI)$ and the other comprised of the red radiances from the nine MISR cameras. For both sets of input to QDA, we performed two experiments that used different training data from DATA-20JUNE2001. In the first experiment the training data are the expert labelled data, as we want to quantify the accuracy of the QDA classifiers when accurate labels are available. We use two-thirds of the 60,111 labelled pixels, selected at random, in order to estimate the QDA parameters for the two sets of inputs. We subsequently apply the two QDA classifiers to the remaining one-third of pixels and compute their misclassification rates. In the second experiment the training data are the labels generated by the ELCMC algorithm when applied to DATA-20JUNE2001. Now, we are testing if QDA can improve upon the ELCMC algorithm results by generalizing the decision boundaries of the ELCMC algorithm to nonlinear ones. As before, the results of the QDA method are compared with the expert labels to obtain error rates. The results of these two experiments when applied to the two sets of input features are listed in Table 1 of summary section 5.5.

As Table 1 illustrates, the QDA approach when developed with expert labels produces much lower misclassification rates, i.e., 0.96% with radiances as input and 2.94% with the three-feature set of inputs, than the ELCMC algorithm, whose misclassification rates are 6.05% for automatically chosen thresholds and 5.82% for the optimal set of thresholds developed from the expert labels. That is, the QDA approach with either feature set performs better than the ELCMC algorithm when accurate labels are available. Unfortunately, accurate labels are usually unavailable. Using ELCMC algorithm results as training data, the QDA method based on radiances has a misclassification rate

Figure 10: QDA cloud-detection results for Data-20June2001 when the QDA algorithm is trained using ELCMC algorithm classification results. Left: Df-camera red radiance image. Center: ELCMC-QDA algorithm results when applied to our set of three features. Right: ELCMC-QDA algorithm results when applied to the red radiances from the nine MISR cameras. White labels indicate cloudy pixels.

of 3.59%, while the QDA method based on the three features has a misclassification rate of 4.09%. In both cases QDA provides a nontrivial improvement over the ELCMC algorithm misclassification rate of 6.05%. We illustrate the results from the ELCMC-QDA approach for DATA-20JUNE2001 in Figure 10b with the three features as input and in Figure 10c for the nine MISR camera radiances. The two sets of results are self-consistent, as we would expect from the misclassification rates, and both sets of inputs lead to fewer mistakes, especially over the "clear" mountain area towards the bottom and left of the image, than the ELCMC algorithm (Figure 9).

The differences in the misclassification rates when training with expert labels and ELCMC algorithm-produced labels are caused by mistakes in the ELCMC algorithm labels. Overall, when training labels are accurate, the QDA method is accurate; otherwise, errors in the training produce errors in the classifications. We tested on both sets of inputs a robust version of QDA based on Minimum Covariance Determinant estimators (c.f. Rousseeuw, 1985; Croux and Haesbroeck, 1999), but it does not improve the results. This result motivated our search for other types of robust classifiers that we could train with ELCMC algorithm results. The Support Vector Machine (SVM) approach is one candidate for providing robust classification results when trained with errors in the data.

### 5.3 Support Vector Machine Classifiers

Research on machine learning algorithms led to the development of the Support Vector Machine (SVM) approach (Vapnik, 1996; Burges, 1998; Cristianini and Shawe-Taylor, 2000; Schölkopf and Smola, 2002; Hastie et al., 2001), which is able to construct nonlinear classifiers as does QDA. Originally motivated by Vapnik (1996) using the concept of margins, SVMs turn out to have equivalent regularization formulations in Reproducing Kernel Hilbert Spaces (RKHS), as shown by Wahba et al. (1999). The SVM approach has proven to be successful in many classification problems with high dimensional inputs and is robust to mistakes in the training data due to its hinge loss function, which is related to $L1$ type of loss function. Therefore, a SVM approach might be more effective than QDA when using ELCMC labels as training data.

In a two-class classification problem we have training data, i.e., predictors $(x_1, x_2, \cdots, x_N)$, $(x_i \in R^p)$ and corresponding labels $(y_1, y_2, \cdots, y_N) \in \{-1, 1\}$. A SVM maps $x$ to a feature space

$z = \phi(x)$ $(z \in R^{\mathrm{m}})$ and searches for the linear hyperplane in $R^{\mathrm{m}}$ with the largest margin across the transformed features $z$. Letting $z_{\mathrm{i}} = \phi(x_{\mathrm{i}})$, we can write the original optimization problem (Vapnik, 1996) solved by a SVM as

$$\min_{\omega,b} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\xi_{\mathrm{i}}, \tag{3}$$

$$\text{subject to: } y_i(\omega z_i - b) \geq 1 - \xi_i;\ \xi_i \geq 0\,\forall i.$$

Equation (3) is a quadratic optimization problem with linear constraints, where $C$ is a penalization parameter for constraint violation. The computation of a SVM involves an inner product in the feature space, $\mathbf{K}(x_{\mathrm{i}}, x_{\mathrm{j}}) =< \phi(x_{\mathrm{i}}), \phi(x_{\mathrm{j}}) >$, which is called the kernel function. The classifier learned from the optimization problem is in the form of $f(x) = sign(\hat{\omega}z - \hat{b})$, which can be written as $f(x) = sign(\sum_{i=1}^{N}\alpha_{\mathrm{i}}\mathbf{K}(x_{\mathrm{i}}, x) + c)$ with the nonzero $\alpha_{\mathrm{i}}$ called support vectors.

Wahba et al. (1999) proved that a SVM with kernel $\mathbf{K}(\cdot,\cdot)$ in the predictor space is the solution of the following regularization problem:

$$\text{minimize } L(f, \text{training data}) + \lambda J(f), \tag{4}$$

where $L$ is the empirical loss function and $J(f)$ is a penalty function. Here, $f$ is a member of a Reproducing Kernel Hilbert Space (RKHS) with kernel $\mathbf{K}$ on the predictor space and a hinge loss function $L(f, \text{training data}) = \sum_{n=1}^{N}[1 - y_{\mathrm{n}}f(x_{\mathrm{n}})]_{+}$, where $[s]_{+} = s$ if $s > 0$ and zero otherwise. The function $J(f)$ is the norm in the RKHS of the reproducing kernel $\mathbf{K}$. Since the hinge loss function shares one piece with the $L1$ loss function, it provides a SVM with some robustness to mislabelled data.

We choose Gaussian kernels in this cloud detection problem for two reasons. First, the Gaussian kernel is the most commonly used in machine learning and has been applied successfully to many classification problems. Second, smooth boundaries between "cloudy" and "clear" areas (c.f. Figure 6) need to be found in this problem. Lin and Brown (2002) demonstrated in the case of regression that the SVM Gaussian kernel tends to approximate smooth boundaries well and conjectured that it would work well in classification problems. The Gaussian kernel is defined as

$$\mathbf{K}(x_{\mathrm{i}}, x_{\mathrm{j}}) = exp\left(-\|x_{\mathrm{i}} - x_{\mathrm{j}}\|^2/2\sigma^2\right), \tag{5}$$

for $x_{\mathrm{i}}$ and $x_{\mathrm{i}}$ two points in the predictor space. In our cloud detection problem, the predictors are either $x = (log(SD), CORR, NDAI)$ or 1.1-km resolution red radiances from the nine MISR cameras and we attempt to classify them as either resulting from a "clear" or "cloudy" 1.1-km resolution region.

The software that we used to train the SVM is the Ohio State University SVM Classifier Matlab Toolbox (Junshui Ma et al. – `http://www.eleceng.ohio-state.edu/~maj/osu_smv/`). The OSU SVM toolbox implements SVM classifiers in C++ using the LIBSVM algorithm of Chih-Chung Chang and Chih-Jen Lin. The LIBSVM algorithm is a simplification of both SMO by Platt (1999) and SVMLight by Joachims (1999). Both the SMO and SVMLight algorithms break the large SVM Quadratic Programming (QP) optimization problem into a series of small QP problems. The computational complexity of training SVM using SMO and SVMLight are both empirically around $O(N_1^2)$ in the examples shown in Joachims (1999), where $N_1$ is the training sample size. The complexity of testing is $O(m\,N_2)$ where $N_2$ is the test size and $m$ is the number of support vectors, which usually increases linearly with the size of the training data set. The LIBSVM algorithm uses $C$ in equation (3) as a parameter for the cost of constraint violations and $\sigma$ as the bandwidth parameter of the Gaussian kernel in equation (5). Both $C$ and $\sigma$ must be provided by the user and, as will be discussed below, we compute values for them using expert-labelled data from some scenes and then we fix them for other scenes.

In the analysis to follow we describe and compare the dependencies of SVM performance on the source of the SVM training data. First, using expert labelled scenes, we compare the performance of our SVM classifiers when trained with the set of three features and the red radiances from all nine MISR cameras. Second, we assess SVM performance degradation when the labelled scenes are derived from the ELCMC algorithm, again using the set of three features and the nine MISR camera red radiances as input features. The motivation for these experiments is the same as for the QDA method analysis. The SVM results are compared with the QDA results using the same training and testing data in order to assess the relative performance of the two methods.

### 5.3.1 SVM Results with Expert Labelled Training Data

Using the expert labelled data, we first train our first SVM with the three-feature set of $log(SD)$, $CORR$, and $NDAI$. We then train a second SVM using the red radiances from the nine MISR cameras. We normalize all features and radiances to a mean of 0 and a standard deviation of 1, thereby ensuring that all input features have approximately the same magnitudes. These experiments are identical to those with the QDA, as we want to compare the results of the two methods and what are the optimal parameters.

In training a SVM computational demands increase rapidly with the number of samples in the training data set. Therefore, we use only 1%, 2%, 5%, and 10% of the 60,111 expert labels for DATA-20JUNE2001 to train the SVM, as opposed to the 66% that we use to train the QDA algorithm. As in the QDA algorithm, all of the training samples are drawn at random from the pool of expert labels. As mentioned earlier, we train two SVM classifiers, one on our set of three features and the other on the red radiances from the nine MISR cameras. We apply the classifiers to the expert labels not used in the training data set in order to assess SVM performance. To compare SVM and QDA algorithm performance we also train the QDA algorithm on exactly the same training data that we use for the SVM algorithm.

To ascertain optimal values for $C$ and $\sigma$ we first set $C$ and $\sigma$ to a pair of values within a large range of values for them. We would then run the SVM training procedure 20 times for each percentage of expert labels listed above. We then compute the mean and standard deviation of the misclassification rates for each set of 20 training runs. We repeat this procedure for different values of $C$ and $\sigma$, covering the large range of values for them. We then take as the optimal pair of values for $C$ and $\sigma$ the set that produced the lowest average misclassification rate. We obtain $C = 200$ and $\sigma = 0.01$ for the optimal SVM parameters when using our three features as input and $C = 180$ and $\sigma = 0.03$ for the optimal SVM parameters when using the nine MISR radiances as input. We obtain both of these sets of values using 10% of the expert labels for the training data set. We then use these values for $C$ and $\sigma$ when training the SVM using ELCMC algorithm results as training data. The misclassification rates that we obtained with the SVM and QDA approaches just described are reported in Table 2.

As Table 2 illustrates, the average misclassification rates of the SVM are slightly lower than for

Figure 11: SVM cloud-detection results for Data-20June2001 when the SVM is trained using ELCMC algorithm classification results. Left: Df-camera red radiance image. Center: ELCMC-SVM algorithm results when applied to our set of three features. Right: ELCMC-SVM algorithm results when applied to the red radiances from the nine MISR cameras. White labels indicate cloudy pixels.

the QDA algorithm when using the same training data, whether it uses the set of three features or the red radiances from the nine MISR cameras. The most accurate results were obtained using SVM on radiances, which produced a 0.72% misclassification rate using 10%, i.e., around 6000 values, of the expert labels. Note that the misclassification rates of the QDA and SVM algorithms decrease as the training sample size increases. Overall, we conclude that the SVM with expert labels for training produces significantly better results than the ELCMC algorithm, while its results are only slightly better than QDA when using expert labels. As expert labels are usually unavailable with new data, we now test the SVM when training it with ELCMC algorithm results.

### 5.3.2 SVM Results with ELCMC Labelled Training Data

In this study we use the output of the ELCMC algorithm (section 5.1) as training data for an SVM classifier. We perform the same experiments as in section 5.3.1, except we use randomly selected ELCMC classification results as training data in conjunction with the optimal values chosen in section 5.3.1 for $C$ and $\sigma$. The SVM classifiers are applied to the entire data set, while the misclassification rates are relative to the expert labels. In Table 1 we show the average and standard deviation of the misclassification rates for 20 runs using 10% of ELCMC algorithm results as training data. The average and standard deviation of the misclassification rate for the ELCMC-SVM approach using the set of three features as input is $3.93 \pm 0.13\%$, while for the nine MISR radiances as input the mean and standard deviation are $3.13 \pm 0.18\%$. Overall, the average misclassification rates for the SVM are lower than for the ELCMC algorithm and they are slightly lower than those from the ELCMC-QDA approach. The SVM classification results for one of the 20 runs are shown in Figure 11b and Figure 11c. The SVM appears generally robust to errors in the training data set, as expected.

In summary, the SVM performs slightly better than QDA as a second stage classifier that utilizes ELCMC algorithm results. One potential reason for this improvement is that the SVM more readily adapts to smooth classification boundaries compared to QDA methods. However, visual inspection of the results from the ELCMC-QDA and ELCMC-SVM methods shows no important differences between the two approaches. This finding is important as the computational cost of the SVM is much higher than for QDA. The computational costs of training a SVM with the LIBSVM algorithm

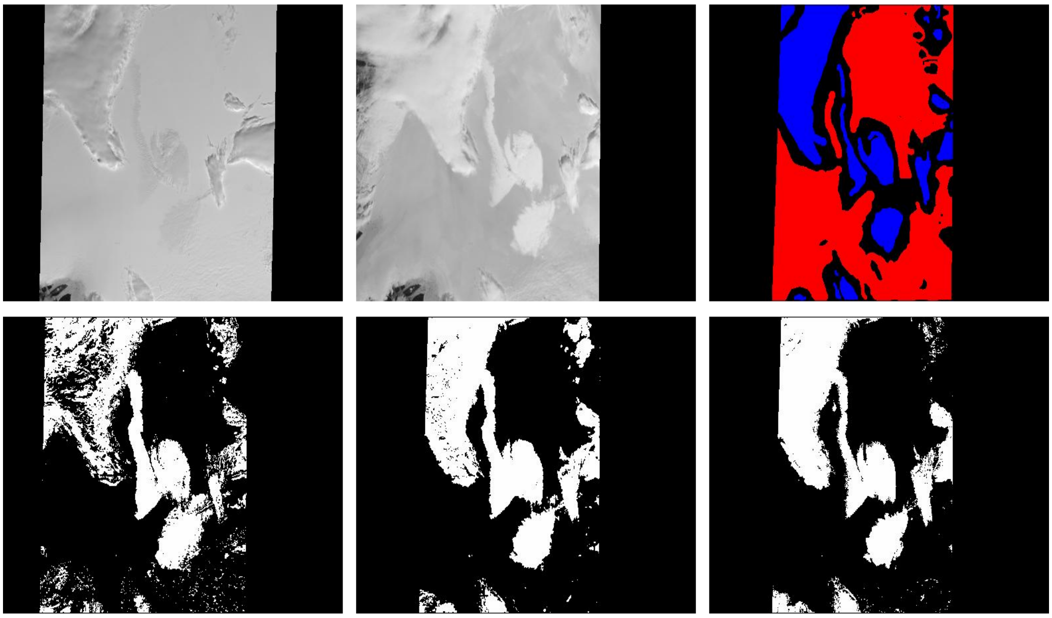

Figure 12: ELCMC-QDA algorithm cloud-detection results for Data-17July2002. Upper left: An-camera red radiance image. Upper center: Df-camera red radiance image. Upper right: Image of expert labels, where red labels indicate clear pixels, blue labels indicate cloudy pixels, and black labels indicate pixels ignored in the labelling process. Lower left: ELCMC algorithm classification results. Lower center: ELCMC-QDA algorithm results when applied to our set of three features. Lower right: ELCMC-QDA algorithm results when applied to the red radiances from the nine MISR cameras.

is approximately $O(N_1^2)$, whereas the costs for testing go as $O(mN_2)$, where $N_1$ and $N_2$ are the number of training and testing data samples, respectively, and $m$ is the number of support vectors, which increases linearly with $N_1$. The computational cost of QDA is $O(N_1)$ for training and $O(N_2)$ for testing. Therefore, the ELCMC-QDA approach is not nearly as computationally demanding as the ELCMC-SVM approach.

### 5.4 Classifier Results for DATA-17JULY2002

To test the three proposed classifier methods on a different orbit, as well as assess the capability of classifiers trained on one region to perform accurately in other regions, we analyze DATA-17JULY2002. Red radiances from the MISR An and Df cameras, expert labels and classification results for this scene are presented in Figure 12. The scene consists primarily of clear-sky snow-covered regions of the Greenland plateau and several cloud decks scattered across it. For this scene the MISR L2TC stereo-derived cloud product contains no retrievals for 91.82% of the pixels and misclassifications for 0.28% of the pixels. The L2TC algorithm has difficulties with this scene because of the lack of significant texture in the reflecting surfaces, which makes matching the same object in the different MISR camera views difficult.

We obtain the classification results presented in Figure 12 and Table 1 section 5.5 of in a manner identical to the earlier studies in sections 5.1, 5.2 and 5.3. In particular, we train the classifiers on expert labels, as well as ELCMC algorithms results, for this scene and then test on subsets of the training data withheld for this purpose. For the ELCMC algorithm we determine the threshold on $NDAI$ in the two Gaussian mixture model to be 0.177. As Table 1 illustrates, the ELCMC algorithm obtains a misclassification rate of 6.28%, which improved to 2.02% when supplemented by the ELCMC-QDA method applied to radiances and to 1.94% when supplemented by the same approach applied to features. The misclassification rates obtained by supplementing the ELCMC algorithm with an ELCMC-SVM approach are 2.08% with radiances as input features and 1.82% when using the set of three features as the input. Visual inspection of the results in Figure 12 once again demonstrates no significant performance differences between the ELCMC-QDA and ELCMC-SVM approaches.

To test the performance of classifiers trained on one region when applied to a different region, we apply the ELCMC-QDA and ELCMC-SVM algorithms trained on DATA-17JULY2002 to DATA-20JUNE2001. We obtain misclassification rates of 17.05% for ELCMC-QDA and 19.21% for ELCMC-SVM using the radiances as features and 20.25% for ELCMC-QDA and 25.70% for ELCMC-QDA with the set of three features as inputs. These results suggest that classifiers trained for one area are not appropriate without modification for other areas, mainly because of the requirement of different thresholds for different locations and times.

To demonstrate necessariness of feature selection, we consider the means of the expert labelled "clear" and "cloudy" sky radiances listed in Table 3. Clearly, the Df-camera radiances are greater in magnitude in cloudy regions as compared to clear-sky regions, confirming enhanced forward scattering from clouds relative to the underlying surface. However, the mean of Df angle radiances (355.36) of clear regions in DATA-17JULY2002 is even higher than that (321.87) of cloud regions in DATA-20JUNE2001. It suggests that the feature selection based ELCMC algorithm is a necessary step to provide good training data for QDA and SVM, although a quadratic boundary in the raw radiances space can separate "clear" and "cloud" pixels well in each location.

To determine the best thresholds for the data collected at different locations and times is not simple. One approach for reducing the significance of this problem is to develop features that are scene independent, while a second approach, albeit a time consuming one, is to develop specific time-dependent sets of thresholds for each location. Our research results presented here indicate a third possibility: use the ELCMC algorithm with two fixed thresholds and a third that can be computed automatically to provide classification results for training the ELCMC-QDA and ELCMC-SVM algorithms. Although the thresholds chosen in the ELCMC step are not always the best ones for a particular scene, the post-process QDAs or SVMs are robust enough to provide good approximations of the classification boundary based on the ElCMC labels. The results from the current study indicate that misclassification rates of approximately 2-4% can be obtained with this approach.

We wish to emphasize that the expert "cloudy" and "clear" labels for pixels in DATA-20JUNE2001 and DATA-17JULY2002 were provided with high confidence. As such, pixels with a nebulous classification, such as near cloud boundaries, were generally not part of the expert label data sets. Radiances originating near cloud boundaries will be registered to different locations on the reference ellipsoid from one MISR camera image projection to the next, a situation similar to the cloud with small spatial extent presented in Figure 3). For these types of scenes the use of features, rather than radiances, might produce more accurate results. We plan to investigate the performances of ELCMC algorithm post-processors for these types of difficult scenes using both radiances and specific feature sets as their input.

Table 1: Misclassification rates of tested classifiers

| Classifiers | Error Rates (SD) DATA-20JUNE2001 | Error Rate (SD) DATA-17JULY2002 |
|---|---|---|
| Operational algorithm [no retrieval percentage] | 53.25% [29.92%] | 0.28% [91.82%] |
| ELCMC on $SD$, $CORR$, and $NDAI$ | 6.05% | 6.28% |
| QDA on features using expert training labels | 2.94% | 1.55% |
| QDA on radiances using expert training labels | 0.96% | 1.38% |
| SVM on features using expert training labels | 1.82% (0.06%) | 1.20% (0.05%) |
| SVM on radiances using expert training labels | 0.72% (0.08%) | 0.60% (0.09%) |
| QDA on features using ELCMC training labels | 4.09% | 1.94% |
| QDA on radiances using ELCMC training labels | 3.59% | 2.02% |
| SVM on features using ELCMC training labels | 3.84% (0.07%) | 1.82% (0.16%) |
| SVM on radiances using ELCMC training labels | 3.19% (0.09%) | 2.08% (0.36%) |

NOTE: Misclassification rates are produced by the three classifiers when applied both to the set of three features and the red radiances from the nine MISR cameras obtained from Data-20June2001 and Data-17July2002.

Table 2: Mean misclassification rates of the QDA and SVM on expert labels

| Training size | QDA on features | QDA on radiance | SVM on features | SVM on radiances |
|---|---|---|---|---|
| 600(1%) | 3.20% (0.44%) | 1.17% (0.15%) | 2.22% (0.25%) | 1.26% (0.30%) |
| 1200(2%) | 3.13% (0.29%) | 1.04% (0.23%) | 2.02% (0.22%) | 1.05% (0.19%) |
| 3000(5%) | 3.10% (0.28%) | 1.00% (0.12%) | 1.90% (0.08%) | 0.90% (0.14%) |
| 6000(10%) | 3.06% (0.23%) | 0.99% (0.12%) | 1.82% (0.06%) | 0.72% (0.08%) |

NOTE: Mean misclassification rates of the QDA and SVM are trained and tested on 20 identical sets of random samples drawn from the expert labels of Data-20June2001. Standard deviations about the mean misclassification rates over the 20 runs are in parentheses.

### 5.5 Summary Tables

As a summary of our study, we present the misclassification rates from all classifiers tested on DATA-20JUNE2001 and DATA-17JULY2002 in Table 1. To illustrate the dependence of the misclassification rates on the size of the training data sets we present in Table 2 results from QDA and SVM algorithms applied to DATA-20JUNE2001 that were trained with different percentages of the expert labelled data sets. Finally, in Table 3 we demonstrate the scene-dependence of the radiances for expert labelled “cloudy” and “clear” radiances, indicating the problems associated with training them on one regions and then applying them to other regions.

## 6 Conclusions and Discussions

In this paper we developed three features, i.e., $(SD_{\mathrm{An}}, CORR, NDAI)$, using MISR data for detection of clouds over polar regions. Three classification methods, i.e., ELCMC, ELCMC-QDA and ELCMC-SVM, were developed and applied to the set of three features for cloud detection in

Table 3: Means of the red radiances from the nine MISR cameras

| Mean Radiances | Df | Cf | Bf | Af | An | Aa | Ba | Ca | Da |
|---|---|---|---|---|---|---|---|---|---|
| Cloudy(20JUNE2001) | 321.87 | 257.60 | 206.87 | 174.34 | 163.89 | 190.79 | 192.95 | 197.70 | 197.58 |
| Clear(20JUNE2001) | 222.09 | 207.02 | 193.02 | 176.35 | 167.80 | 176.81 | 182.66 | 188.92 | 192.92 |
| Cloudy(17JULY2002) | 408.32 | 354.22 | 310.17 | 277.21 | 262.26 | 284.82 | 286.18 | 281.94 | 272.19 |
| Clear(17JULY2002) | 355.36 | 341.10 | 324.96 | 302.58 | 285.26 | 284.43 | 282.97 | 278.27 | 272.05 |

NOTE: Means of the red radiances are calculated from the nine MISR cameras for Data-20June2001 and Data-17July2002, partitioned according to expertly labelled “clear” and “cloudy” pixels.

polar regions. All three methods produced significantly more accurate results compared to the operational MISR stereo cloud detection algorithm, both in terms of classification accuracy when compared with expert labelled data and in terms of computational speed. In an operational setting the most fundamental method among the three is the ELCMC algorithm since we propose using it to create the training data for the QDA and SVM methods.

In this study the ELCMC algorithm produced misclassification rates of 6.05% and 6.28% against expert labels for data sets coming from MISR orbits 7898 and 13723, which are much lower than the rates for the MISR operational algorithm. Using the results of the ELCMC algorithm as training data for the QDA and SVM algorithms, we find that both the QDA and SVM algorithms improve upon the accuracy of the ELCMC results, with the SVM producing slightly better results than QDA. But the computational expense of the SVM is significantly greater than for QDA and we currently favor the ELCMC-QDA approach for processing massive amounts of MISR data. We obtained comparable results from the ELCMC-QDA approach when we replaced the set of three features with the red radiances from the nine MISR cameras. So, either radiances or the set of three features are appropriate as inputs to the ELCMC-QDA algorithm, although in practice greater efficiency can be gained by using the set of three features as they are always generated by the ELCMC algorithm as a pre-processor.

The methods that we tested in this study can facilitate, as a pre-processor, the MISR L2TC operational cloud-height retrieval algorithm . Using either ELCMC-QDA or ELCMC-SVM for ”clear” and ”cloudy” sky detection, one could then run the L2TC matching algorithm only on the “cloudy” pixels. Moreover, in the L2TC cloud-motion retrieval algorithm, we would be more confident that we are matching “cloudy” pixels and snow- and ice-covered surface features. This approach would improve the accuracy and speed of the L2TC cloud algorithms as pattern matching of “clear-sky” pixels would be eliminated.

As we have learned through this project, feature selection based on the physics knowledge and statistical exploratory analysis is very important for massive data set problems of the information technology age. Understanding the physics behind the problem is not only the final goal of the scientific research, but also provides us insights for solving the statistical detection problem. The traditional statistical approach as well as the cutting-edge method SVM from machine learning led to the huge improvements of our methods over the operational algorithm.

## Acknowledgements

Tao Shi and Bin Yu are partially supported by NSF grant CCR-0106656. Eugene Clothiaux is supported by Jet Propulsion Laboratory contracts 1216622 and 1259588. Amy Braverman's work is performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. MISR data were obtained at the courtesy of the NASA Langley Research Center Atmospheric Sciences Data Center. The authors would like to thank L. Di Girolamo, D.J. Diner, R. Davies, and R. Kahn for helpful discussions and suggestions. We would especially like to thank Dominic Mazzoni of the Jet Propulsion Laboratory for sharing, adapting and expertly supporting his software package MISRLEARN for our use in this study. All of the expert labelled data for this study were produced using his MISRLEARN software package.

## References

[1] Burges, C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Knowledge Discovery and Data Mining*, 2(2).

[2] Chang, C. and Lin, C. (2001), LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm/

[3] Croux, C., and Haesboeck, G.(1999) Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator, *Journal of Multivariate Analysis*, **71**, 161C190.

[4] Diner, D.J., Asner, G.P., Davies, R., Knyazikhin, Y., Muller, J. Nolin, A.W., Pinty, B., Schaaf, C.B., and Stroeve, J. (1999a). New directions in earth observing scientific applications of mulitangle remote sensing. *Bulletin of American Meteorological Society.* **80**, 2209–2228.

[5] Diner, D.J., Davies, R., Di Girolamo, L., Horvath, A. Moroney, C., Muller, J.-P., Paradise, S.R., Wenkert, D, and Zong, J. (1999b). MISR level 2 cloud detection and classification algorithm theoretical basis. Jet Propulsion Lab., Pasadena, CA, JPL Tech. Doc. D-11399, Rev. D, 1999.

[6] Hastie, T., Tibshirani, R.,and Friedman, J.(2001). *The Elements of Statistical Learning, Data Mining, Inference, and Prediction* Springer.

[7] Joachims, T. (1999). Making Large-Scale SVM learning Practical. In: *Advances in Kernel Methods - Support Vector Learning* Schölkopf, B., Berkgers, C. and Smola, A., MIT Press 1999.

[8] Lin, Y. and Brown, L. D. (2002). Statistical Properties of the Method of Regularization with Periodic Gaussian Reproducing Kernel. *Technical Report 1062r, Department of Satatistics, University of Wisconsin.* Accepted, Annals of Statistics.

[9] Mardia, K. V., Kent, J.T., and Bibby, S. M. (1979) *Multivariate Analysis.* Academic Press.

[10] Cristianini, N. and Shawe-Taylor, J. *An Introduction to Support Vector Machines.* Cambridge University Press, Cambridge, UK.

[11] Nolin, W. A., Fetterer, M.F., and Scambos, A.T. (2002) Surface Roughness Characterization of Sea Ice and Ice Sheets: Case Studies With MISR Data. *IEEE Trans. Geosci. Remote 1Sensing.* **40**. **July** ,1605–1615.

[12] Platt, J. (1999). Fast Training of Support Vector Machines using Sequential Minimal Optimization. In: *Advances in Kernel Methods - Support Vector Learning* Schölkopf, B., Berkgers, C. and Smola, A., MIT Press 1999.

[13] Ripley, B. D. (1996) *Pattern Classification and Neural Networks.* Cambridge, 1996.

[14] Rousseeuw, J.P. (1985), Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications edited by Grossmann, W. et al.*, B. Reidel, Dordrecht (1985), 283C297.

[15] Schölkopf, B.and Smola, A. (2002). *Learning with Kernels*, MIT Press, Cambridge, MA, 2002

[16] Shi, T., Yu, B., and Braverman, A. (2002). MISR cloud detection over Ice/Snow using Linear Correlation Matching. *Technical Report 630, Department of Statistics, University of California.*

[17] Vapnik, V., (1996). *The Nature of Statistical Learning Theory.* Springer, NY.

[18] Wahba, G., Lin, Y., and Zhang, H. (1999). GACV for Support Vector Machines, or Another Way to Look at Margin-Like Quantities, *Advanced in Large Margin Classifiers.*