

Daytime Arctic Cloud Detection Based on Multi-Angle Satellite Data With Case Studies

Tao SHI, Bin YU, Eugene E. CLOTHIAUX, and Amy J. BRAVERMAN

Global climate models predict that the strongest dependences of surface air temperatures on increasing atmospheric carbon dioxide levels will occur in the Arctic. A systematic study of these dependences requires accurate Arctic-wide measurements, especially of cloud coverage. Thus cloud detection in the Arctic is extremely important, but it is also challenging because of the similar remote sensing characteristics of clouds and ice- and snow-covered surfaces. This article proposes two new operational Arctic cloud detection algorithms using Multiangle Imaging SpectroRadiometer (MISR) imagery. The key idea is to identify cloud-free surface pixels in the imagery instead of cloudy pixels as in the existing MISR operational algorithms. Through extensive exploratory data analysis and using domain knowledge, three physically useful features to differentiate surface pixels from cloudy pixels have been identified. The first algorithm, enhanced linear correlation matching (ELCM), thresholds the features with either fixed or data-adaptive cutoff values. Probability labels are obtained by using ELCM labels as training data for Fisher's quadratic discriminant analysis (QDA), leading to the second (ELCM-QDA) algorithm. Both algorithms are automated and computationally efficient for operational processing of the massive MISR data set. Based on 5 million expert-labeled pixels, ELCM results are significantly in terms of both accuracy (92%) and coverage (100%) compared with two MISR operational algorithms, one with an accuracy of 80% and coverage of 27% and the other with an accuracy of 83% and a coverage of 70%. The ELCM-QDA probability prediction is also consistent with the expert labels and is more informative. In conclusion, ELCM and ELCM-QDA provide the best performance to date among all available operational algorithms using MISR data.

KEY WORDS: Classification; Clustering; Feature selection; Multiangle imaging spectroradiometer; Quadratic discriminate analysis; Remote sensing.



1. INTRODUCTION

Sensitivity of Earth's climate to increasing amounts of atmospheric carbon dioxide is a topic of general and scientific interest, and an important public policy issue as well. Today's global climate models generally predict that global surface air temperatures will increase by 1.5–3.5 K, with a doubling of atmospheric carbon dioxide levels throughout the twenty-first century. These same models also predict that the strongest dependences of surface air temperatures on increasing atmospheric carbon dioxide levels will occur in the Arctic (Giorgi and Bi 2005). As the Arctic warms, changes in the properties and distribution of ice- and snow-covered surfaces, atmospheric water vapor, and clouds can potentially lead to further warming and thus strong sensitivity to increasing amounts of atmospheric carbon dioxide. A systematic study of these dependences requires accurate Arctic-wide measurements (e.g., Francis, Hunter, Key, and Wang 2005), especially of cloud coverage, because clouds play an important role in modulating the

sensitivity of the Arctic to increasing surface air temperatures (e.g., Kato et al. 2006).

Ascertaining the properties of clouds in the Arctic is a challenging problem, because liquid- and ice-water cloud particles often have scattering properties similar to those of the particles that compose ice- and snow-covered surfaces. As a result, the amount of visible and infrared electromagnetic radiation emanating from clouds and snow- and ice-covered surfaces is often similar, which leads to problems in the detection of clouds over these surface types. Without accurate characterization of clouds over the Arctic, we cannot assess their impact on the flow of solar and terrestrial electromagnetic radiation through the Arctic atmosphere, and cannot ascertain whether they are changing in ways that enhance or ameliorate future warming in the Arctic.

With the launch of the Multiangle Imaging SpectroRadiometer (MISR) onboard the National Aeronautics and Space Administration (NASA) Terra satellite in 1999, novel electromagnetic radiation measurements made at nine view angles became available for scientific study. Unlike traditional multispectral sensors that take measurements in a single view, the MISR sensor comprises nine cameras (Fig. 1), with each camera viewing Earth scenes at a different angle in four spectral bands (blue, green, red, and near-infrared). The nine view zenith angles of the cameras are 70.5° (Df), 60.0° (Cf), 45.6° (Bf), and 26.1° (Af) in the forward direction; 0.0° (An) in the nadir direction and 26.1° (Aa), 45.6° (Ba), 60.0° (Ca), and 70.5° (Da) in the aft direction. (The “f” in the letter designation of the cameras represents the “forward” direction, and the “a” represents the “aft” direction.) The Da camera collects data from a location 7 minutes after the Df camera.

MISR produces a vast amount of data due to its global coverage at a high spatial resolution. The MISR cameras cover an approximate 360-km-wide swath on the Earth's surface that extends across the daylight side of the Earth from the Arctic down

Tao Shi is Assistant Professor, Department of Statistics, Ohio State University, Columbus, OH 43210 (E-mail: taoshi@stat.osu.edu). Bin Yu is Professor, Department of Statistics, University of California, Berkeley, CA 94720 (E-mail: binyu@stat.berkeley.edu). Eugene E. Clothiaux is Associate Professor, Department of Meteorology, Pennsylvania State University, University Park, PA 16802 (E-mail: cloth@essc.psu.edu). Amy J. Braverman is Statistician and Senior Member, Information Systems and Computer Science, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 (E-mail: Amy.Braverman@jpl.nasa.gov). This research was supported in part by National Science Foundation grants FD01-12731 (Yu), CCR-0106656 (Shi and Yu), DMS-03036508 (Shi and Yu), and DMS-0605165 (Yu) and ARO grant W911NF-05-1-0104 (Yu). It was also supported in part by a Miller Foundation Professorship to Yu in Spring 2004. Clothiaux was supported by National Aeronautics and Space Administration (NASA) grant NNG04GL93G and Jet Propulsion Laboratory contract 1259588. Braverman's work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the NASA. MISR data were obtained at the courtesy of the NASA Langley Research Center Atmospheric Sciences Data Center. The results given in Section 4.2 are courtesy of Guilherme Rocha, a doctoral degree student of Bin Yu at the University of California, Berkeley. The authors thank L. Di Girolamo, D. J. Diner, R. Davies, and R. Kahn for helpful discussions and suggestions, and especially 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 labeled data for this study were produced using his MISRLEARN software package. The full version of this article is available online at www.stat.berkeley.edu/~binyu/publications.html.

In the Public Domain
Journal of the American Statistical Association
June 2008, Vol. 103, No. 482, Applications and Case Studies
DOI 10.1198/016214507000001283

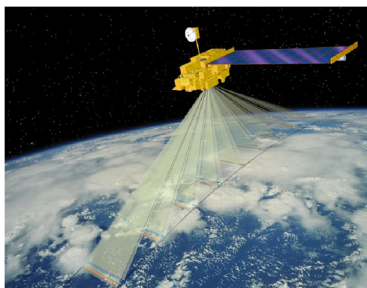


Figure 1. Cartoon illustration of the Terra satellite with the view directions of the nine MISR cameras. Image is courtesy of the MISR science team at the Jet Propulsion Laboratory.

to Antarctica in about 45 minutes. There are 233 geographically distinct but overlapping, MISR swaths, which are also called *paths*. MISR collects data from all paths on a repeat cycle of 16 days; that is, it covers the exact same path every 16 days. In the MISR data products, each path is subdivided into 180 *blocks*, with the block numbers increasing from the North Pole to the South Pole. Each complete trip of MISR around the Earth is given its own *orbit* number. Each MISR pixel covers a $275\text{ m} \times 275\text{ m}$ region on the ground, producing a huge amount of data as a result of MISR's global coverage. For example, one orbit of data is massive—MISR collects 3.3 megabits per second on average, with a peak rate of 9.0 megabits/second. Due to transmission channel constraints, these massive data sets cannot be downlinked; therefore, only the red radiances and all channels from the nadir camera are transmitted at full 275-m resolution, whereas the blue, green, and near-infrared radiances from the non-nadir cameras are aggregated onboard to a $1.1\text{ km} \times 1.1\text{ km}$ resolution before transmission.

The MISR angular radiances represent a completely new type of remote sensing observation, and from the start they exhibited strong indications of possible separability of clouds from snow- and ice-covered surfaces (Diner et al. 1999a). However, the MISR operational cloud detection algorithms (see details in Sec. 2.1) were designed before MISR was launched, and they were not particularly targeted at detecting clouds over bright surfaces in polar regions. As a result, existing MISR operational algorithms do not work well over polar regions. Moreover, the massive data set size poses a hurdle for any algorithm and imposes somewhat severe computational constraints on candidate statistical solutions to the polar cloud detection problem. The sheer size of MISR data is a very important determinant of the statistical methodologies that are viable in this setting, and the approach of choice must incorporate nonstatistical, scientific, and operational considerations.

Cloud detection in the Arctic is not only scientifically important, but also statistically challenging. Standard classification frameworks are not readily applicable, because it is impossible to get expert labels operationally, considering the size of the data and resource time constraints. Even in an offline setting, a state-of-the-art classifier, such as Gaussian kernel support vector machines (Vapnik 1995), trained on offline expert labels does not deliver satisfactory accuracy when applied to MISR radiance measurements (cf. Sec. 4.1). A clustering framework is not suitable either, because each data unit (defined as each set of three consecutive MISR blocks in this article) could be

totally cloud-covered or fully cloud-free. When a data unit is fully cloud-covered, the clusters correspond to different types of cloudy pixels, but not to whether or not a pixel is cloudy or clear. Therefore, the statistical challenge is to combine classification and clustering schemes in a computationally efficient manner.

The goal of this work is to build operational cloud detection algorithms that can efficiently process the massive MISR data set one data unit at a time without requiring human intervention (expert labeling). We use the results of an extensive exploratory data analysis and interactions with the MISR science team to construct a labeling scheme, or detector. For each data unit, we combine clustering information on the current data unit with the detectors learned from previous data units. Thus the algorithm is sequential, and it adaptively combines classification and clustering methods.

The methodological novelty in our approach is to search for *cloud-free* instead of cloudy ice- and snow-covered surface image pixels. The current operational MISR algorithms look for cloudy pixels using a combination of methods, including parallax between two different MISR cameras and brightness differences. In contrast, our method exploits correlations in brightnesses among multiple views of the same scene inherent under unobstructed, cloud-free conditions. We show that this simple reversal of the detection target is indeed fruitful. It is also a prime example of how statistical insight can bring fresh new tools to bear on problems traditionally solved in nonstatistical ways.

Modeling the surface has an advantage over modeling clouds in our clustering and classification framework because the surface does not change much between different views, whereas clouds are always different from one view to the next. Our algorithms are based on three physically useful features: for characterizing the scattering properties of ice- and snow-covered surfaces the correlation (CORR) of MISR images of the same scene from different MISR viewing directions, the standard deviation (SD_{An}) of MISR nadir camera pixel values across a scene, and a normalized difference angular index (NDAI) that characterizes the changes in a scene with changes in the MISR view direction. We arrive at this set of features after substantial exploratory data analysis by both statisticians and atmospheric scientists.

We propose an enhanced linear correlation matching (ELCM) algorithm based on thresholding the three features with values either fixed or data-adaptive. We envision that the ELCM algorithm will be applied to MISR data units sequentially. The CORR and SD_{An} feature cutoff values, or thresholds, are set to fixed values during the operational processing, because they are stable over data collected from different times and locations. The NDAI threshold is initially set based on offline analysis of the first sets of data units. To classify subsequent data units, the NDAI threshold is either kept at its previous value or updated using a data-adaptive algorithm applied to the current data unit. Labels resulting from the ELCM algorithm are then used to train the Fisher's quadratic discriminant analysis (QDA) to produce probability labels, leading to our second (ELCM–QDA) detector. The QDA step produces the probability of “cloudiness” as a more informative probability prediction.

To assess the accuracies of the ELCM and ELCM–QDA algorithms, they are applied to extensive MISR data sets and

tested against expert labels. The expert labels are provided for 10 orbits of MISR data collected over the Arctic during the daylight season of 2002, and they represent the best available validation data in adequate quantities (i.e., 5.086 million 1.1-km pixels).

The rest of the article is organized as follows. Section 2 contains a brief review of MISR algorithms and introduces the MISR data investigated in this study. Section 3 first describes our proposed methodology, starting with our rationale for the study and computation of the features, then proposes our ELCM algorithm based on **adaptive thresholding of the features** (Sec. 3.2) and demonstrates that probability predictions can be obtained over the partly cloudy scenes by using ELCM labels to train Fisher's QDA (Sec. 3.3). Section 4 presents the testing results. ELCM and ELCM-QDA results are compared against the expert labels along with the MISR operational algorithms, strengths and weaknesses of our algorithms are explored, and finally, the expert labels are used to illustrate that our features are **better than the raw radiation measurements for separating clear from cloudy pixels**. Section 5 concludes the article with a summary and a discussion on potential scientific impacts of our algorithms.

2. MISR OPERATIONAL ALGORITHMS AND DATA DESCRIPTION

Viewing the atmosphere and surface from multiple angles, MISR has **stereo** capabilities that can be used to retrieve the elevation of objects, such as clouds, on or above the surface of the Earth. In addition to stereo information, MISR non-nadir camera angles provide radiation scattering patterns of different objects (e.g., clouds). These two sources of novel information motivated the MISR level 2 top-of-atmosphere cloud (L2TC) algorithm (Diner et al. 1999b), which produces two cloud masks: the stereo-derived cloud mask (SDCM) and the angular signature cloud mask (ASCM). In Section 2.1 we briefly review the MISR operational algorithms and their shortcomings in polar regions. In Section 2.2 we describe the data used in this study.

2.1 MISR Operational Algorithms

The L2TC algorithm uses MISR's stereo capability to detect clouds by comparing retrieved object heights with the underlying known terrain height. The rationale of the L2TC cloud height retrieval algorithm is based on the registration of MISR measurements to a known Earth-based reference ellipsoid. As shown in Figure 2, a cloud is registered to different reference

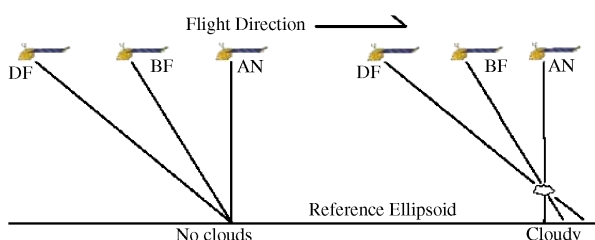


Figure 2. **Registration** of surface features and clouds to the reference ellipsoid. Note that only three of the nine MISR cameras are illustrated and that surface objects are registered to the same location, whereas clouds are registered to different locations.

ellipsoid locations for the different MISR cameras. The L2TC algorithm matches the same object in the different camera images, allowing **retrieval** of object heights and horizontal velocities through simple trigonometric relationships. The SDCM is derived by comparing retrieved object heights with the known terrain heights, and objects more than approximately 650 m above the terrain height are classified as clouds.

The ASCM is based on the **band-differenced angular signature** (BDAS), which characterizes the relationship between the difference of two solar spectral **reflectances** with view angle (Di Girolamo and Davies 1994; Diner et al. 1999b). Based on the relative contribution of air molecule (Rayleigh) scattering to total reflection, the longer the solar ray paths through Earth's atmosphere before reaching MISR, the larger the BDAS. Because radiation reflected from clouds has much shorter paths than radiation reflected from the Earth's surface, small BDAS indicates clouds. **Over ice- and snow-covered surfaces in the Arctic, the difference between the magnitude of blue radiation from the Df camera and near-infrared radiation from the Cf camera is thresholded to distinguish clouds from the surface.**

Both algorithms have difficulties in polar regions, especially for low clouds that are often present in daytime Arctic regions. The SDCM cannot detect low clouds because the distances between low clouds and the surface are often smaller than MISR height retrieval accuracy. The ASCM is good at detecting high and thin clouds, but has difficulty detecting low clouds over terrain because the Rayleigh scattering contribution to BDAS increases rapidly with decreasing in cloud-top height (Di Girolamo and Davies 1994). **Note that both the SDCM and ASCM use characteristics of clouds to find clouds. Our approach, which we propose in Section 3, attempts to overcome these shortcomings of the SDCM and ASCM by searching for surface pixels instead of cloudy ones.**

2.2 Data

The data used in this study were collected from 10 MISR orbits of path 26 over the Arctic, northern Greenland, and Baffin Bay. As described earlier, the repeat time between two consecutive orbits over the same path was 16 days, so the 10 orbits span approximately 144 days from April 28 through September 19, 2002 (a daylight season in the Arctic). We chose path 26 for our study because of the richness of its surface features, which include permanent sea ice in the Arctic Ocean, snow-covered and snow-free coastal mountains in Greenland, permanent glacial snow and ice, and sea ice that melted across Baffin Bay over the study period.

Six data units (MISR blocks 11–13, 14–16, 17–19, 20–22, 23–25, and 26–28) from each orbit are included in this study. Three of the 60 data units were excluded from this study because the surfaces were open water after the sea ice melted in the summer, and the MISR operational algorithm detects clouds over water well. **Therefore, the data investigated contained 57 data units with 7,114,248 1.1-km resolution pixels with 36 radiation measurements for each pixel.** We use the 275-m red radiation measurements to build some of our features, so the actual size of the data set is even larger. Our study concentrates on repeated visits over time so that the experts become familiar with the surface features, which dramatically aids the expert labeling process.

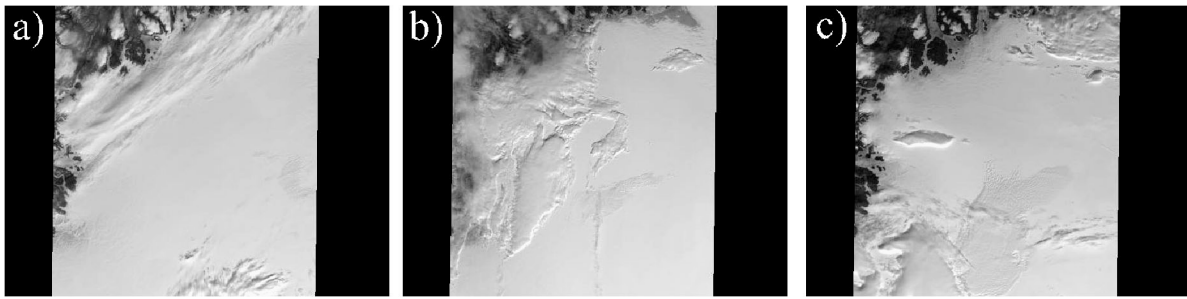


Figure 3. Data collected by the MISR An-camera for three consecutive orbits (i.e., 13257, 13490, and 13723) over blocks 20–22 of path 26.

Figure 3 shows images of three data units collected from MISR blocks 20–22 over three consecutive orbits (i.e., 13257, 13490, and 13723). The images show that the ice- and snow-covered surfaces are as bright as or even brighter than clouds, which contradicts the traditional assumption applicable at lower latitudes that clouds are brighter than surfaces. As a result, traditional cloud detection algorithms based on this assumption can not be applied in polar regions. Furthermore, the differences between different patches of images are not a necessary result of cloud versus surface, because an entire data unit might be fully cloud-covered or fully clear.

To evaluate the performance of our proposed methods and existing MISR operational algorithms, one of the authors hand-labeled the data. Because high-quality ground-based measurements are rare in the polar regions, expert labeling of clear and cloudy scenes is currently the best method for producing validation data for assessing automated polar cloud detection algorithms. After the expert determines the labels (i.e., clear or cloudy) for a patch of image pixels, tools developed by the MISR science team at Jet Propulsion Laboratory, called “misrdump” and “misrlearn” (Dominic Mazzoni, personal communication), are subsequently used to label the pixels in MISR nadir camera images as clear or cloudy. For example, labels for the data shown in Figure 3 are plotted in Figure 4, with white for cloudy and gray for clear. Expert labels are given only to those pixels of which the expert is highly confident based on his knowledge of the measurements, with more ambiguous regions left unlabeled (black pixels). As a result of this conservative labeling scheme, the expert labels cover about 71.5% (5,086,002) of the total valid pixels. We use these labels to evaluate the performance of different cloud detection algorithms.

3. METHODS

We propose two new cloud detection algorithms for the Arctic that are automated and fast for operational processing of the massive MISR data. Both algorithms are based on three physically useful features for identifying surface pixels and thereby obtain cloudy pixels by exclusion. Each pixel is treated as independent of the others, because the computational cost of most spatial models is too high to be implemented in operational data processing. Spatial smoothness of surface and cloudy pixels is taken into account implicitly in our algorithms, because two of our three features are based on local patches of pixels at the 275-m resolution. To be precise, our strategy for solving this arctic cloud detection problem is outlined in three steps:

- *Step 1.* Construct three features based on EDA and domain knowledge (Sec. 3.1).
- *Step 2.* Build an ELCM cloud detection algorithm by setting thresholds (fixed or data-adaptive) on each feature, and apply ELCM to each data unit to produce the first cloud detection product (Sec. 3.2).
- *Step 3.* Predict probability of cloudiness, the second cloud detection product, for the partly cloudy data units by training Fisher’s QDA on the labels produced by the ELCM algorithm (Sec. 3.3).

Recall that expert labels are not available for online data processing. In step 2, only for the data units collected on the first day do we set the best thresholds for the three features using the expert labels offline. For subsequent data units from the same block range, the thresholds are determined by combining clustering information from the data unit being processed with the thresholds learned from the previous visit.



Figure 4. Expert labels for blocks 20–22 of MISR orbits (a) 13257, (b) 13490, and (c) 13723. White represents high confidence cloudy; gray, high confidence clear; and black, unlabeled pixels.

3.1 Constructing Three Physical Features

MISR arctic cloud detection poses a massive data analysis problem. For each pixel, there are 36 dimensions (4 wavelengths at 9 angles). In each data unit, there are 196,608 ($= 384 \text{ rows} \times 512 \text{ columns}$) pixels at the 1.1-km resolution. In our study, only MISR red radiation data are used for constructing the features, for two reasons. First, based on exploratory data analysis and the scattering properties of liquid- and ice-water particles, all four bands have similar reflectance signatures over ice, snow, and clouds (Clothiaux, Barker, and Korolev 2005). Second, only the red radiances have 275-m spatial resolution, the highest spatial resolution of MISR.

In this study we use the radiation measurements that are projected to the underlying surface (i.e., surface-registered data). During MISR operational processing, MISR red radiation measurements are registered to the reference ellipsoid in two different ways. One way is the reference ellipsoid registration illustrated in Figure 2, whereby the radiation measurements are registered directly to the reference ellipsoid, ignoring terrain. In the second terrain-projection registration process (Fig. 5), the radiation measurements are projected (over land) to the terrain surface and then given the reference ellipsoid coordinates of the terrain surface. Surface-registered data refer to ellipsoid-projected measurements over the oceans and terrain-projected measurements over land surfaces.

Over all data units, we investigated the distributions of a large collection of features including linear combinations of angular radiances, correlations among different angles and wavelengths, nonlinear transformations of radiances, spatial patterns of clouds, and smoothness of reflecting surfaces. Combining extensive exploratory data analysis with specific domain knowledge, such as the fact that ice and snow surfaces scatter radiation more isotropically than clouds, which have more forward scatter, we found three physically useful features that can differentiate surface pixels from cloudy ones.

Let $I_{m,\ell}^k$ denote the (m, ℓ) pixel of the 275-m resolution red radiation measurements for the k th MISR camera angle, where k runs from 1 to 9 and represents MISR cameras Df to Da.

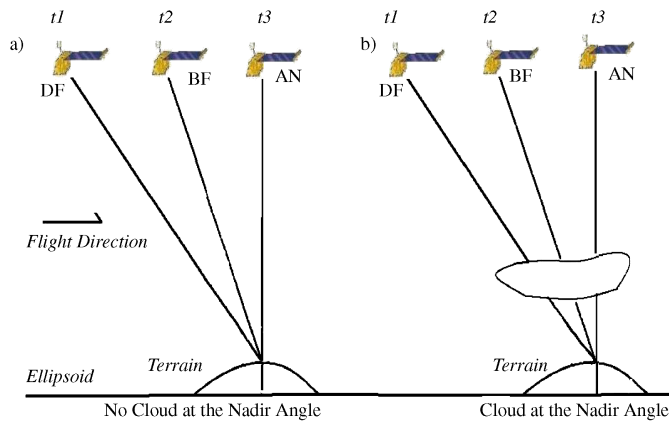


Figure 5. (a) Registration of MISR radiation measurements to the terrain when no clouds are present. (b) Registration of MISR radiation measurements to the terrain when a cloud is present. Note that measurements from different parts of a cloud are mapped to the same location on the terrain.

3.1.1 First Feature: CORR. The first feature, CORR, is revised from Shi, Yu, and Braverman (2002) and is an average linear correlation of radiation measurements at different view angles. We define CORR at a 2.2-km spatial resolution, but with a spacing of 1.1 km, using the 275-m data. Each 1.1-km spaced, 2.2-km resolution correlation LC_{ij} ($i = 1, \dots, 384$; $j = 1, \dots, 512$) in a data unit is constructed from an 8×8 group of pixels at 275-m resolution. Using these 8×8 275-m resolution pixels, we denote the linear correlation coefficient LC_{ij} between the k_1 and k_2 MISR view directions ($k_1 \neq k_2$ and $1 \leq k_1, k_2 \leq 9$) by

$$LC_{ij}^{k_1, k_2} = \frac{\sum_{m=4i-5}^{4i+2} \sum_{\ell=4j-5}^{4j+2} (I_{m,\ell}^{k_1} - \bar{I}_{i,j}^{k_1})(I_{m,\ell}^{k_2} - \bar{I}_{i,j}^{k_2})}{\sqrt{\sigma_{i,j}^{k_1} \sigma_{i,j}^{k_2}}}, \quad (1)$$

where $I_{m,\ell}^k$ is the 275-m resolution radiation measurements at location (m, ℓ) , and $\bar{I}_{i,j}^k$ and $\sigma_{i,j}^k$ are the mean and standard deviation of the 64 radiation measurements for the k th angle associated with location (i, j) .

The means are arithmetic averages, and the standard deviations are given by

$$\sigma_{i,j}^k = \sqrt{\left(\frac{1}{64-1}\right) \sum_{m=4i-5}^{4i+2} \sum_{\ell=4j-5}^{4j+2} (I_{m,\ell}^k - \bar{I}_{i,j}^k)^2}. \quad (2)$$

We expect high correlations over cloud-free areas or low clouds, because the radiation measured by each MISR camera is scattered from the same surface (Fig. 5a). Low correlations for radiation measurements are expected from clouds, especially high clouds (Fig. 5b). However, high correlations for clouds may occur under rare circumstances because of cloud motion. Therefore, we define the first feature as the average of the correlation coefficients obtained from the MISR Af/An and Bf/An camera pairs,

$$CORR_{ij} = (LC_{ij}^{4,5} + LC_{ij}^{3,5})/2,$$

because in practice, this average is always relatively low for high altitude clouds.

High values of CORR suggest either clear (cloud-free) conditions or the presence of low altitude cloud that is registered to the same location on the underlying surface. But if we declare clear for high CORR pixels and cloudy for low CORR ones, then we can produce two types of errors: Smooth cloud-free terrain surfaces can be classified as cloudy (low signal-to-noise ratios lead to low correlations), and low altitude clouds can be classified as clear. To avoid these errors, we need two more features.

3.1.2 Second Feature: SD. To identify smooth surfaces, the standard deviation [eq. (2)] within groups of MISR An ($k = 5$) camera red radiation measurements is useful:

$$SD_{ij} = \sigma_{i,j}^5.$$

It has small values for radiation emanating from smooth surfaces over which the correlations between different MISR view directions are dominated by instrument noise.

3.1.3 Third Feature: NDAI. The third feature is the NDAI developed by Nolin, Fetterer, and Scambos (2002),

$$NDAI_{ij} = \frac{\overline{I_{ij}^1} - \overline{I_{i,j}^5}}{\overline{I_{ij}^1} + \overline{I_{i,j}^5}}, \quad (3)$$

where the average radiation measurements are over a 4×4 group of 275-m resolution red radiation measurements (and thus have a spatial resolution of 1.1 km).

This feature is motivated by the fact that surface-leaving radiation at visible wavelengths is more isotropic from ice- and snow-covered surfaces than from low-altitude clouds (Stephens, Cambell, and Vonder Haar 1981). In our current implementation, as described in (3), MISR Df-camera measurements are compared with those from the An camera. Over clouds in the arctic region, radiation collected in the Df camera is much higher than that collected in the An camera; thus larger NDAI values suggest the presence of clouds.

Our investigation suggests that clear and cloudy pixels separate well in the space spanned by CORR, SD_{An} , and NDAI. In addition, the features have another desirable property for building an efficient cloud detection algorithm: The distributions of features from clear surfaces are about the same for data collected from different geo-locations or at different times. We illustrate the importance of this stability property of the features in our combined clustering and classification algorithms in Section 3.2. Unfortunately, the MISR radiation measurements themselves do not have these “separability” and “stability” properties (see details in the long version of this article, available www.stat.berkeley.edu/~binyu/publications.html).

3.2 Enhanced Linear Correlation Matching Algorithm

As discussed in Section 1, for operational processing of MISR data, expert labels are not available to train a detector for each data unit. Thus detectors must be trained either on data units without expert labels or on data units collected previously. Moreover, the results of clustering algorithms do not always reflect a separation between clear and cloudy conditions, especially when data units are fully cloud-covered or totally clear.

In this section we propose an operational algorithm that combines clustering information on the current data unit with detectors learned from previous data units. The quality of the clustering results on the current data unit depends on the “separability” property of the feature space. To borrow a detector developed from previous data units requires that the distributions of previously collected clear and cloudy pixels in the space spanned by the features be similar to those for the current data unit. The “stability” and the “separability” described earlier are two essential properties of our physical features and are critical to the success of our method. Unfortunately, in a space of radiation measurements, clear and cloudy pixels are neither readily separable nor stable across time and space due to changes in solar illumination directions of uneven terrain and the magnitude of the solar radiation. Therefore, our feature development is both necessary and vital to the success of our methods.

Given the aforementioned two desirable properties of the three features, a natural and computationally efficient detector is to label pixels by thresholding the features. The thresholds are either fixed values or values chosen by a data-adaptive

method. We call this approach ELCM because two additional features are added to the linear correlation matching classification (LCMC) of Shi et al. (2002). We now describe this procedure in detail.

3.2.1 ELCM Algorithm.

A $1.1 \text{ km} \times 1.1 \text{ km}$ pixel is labeled clear when

- $SD_{An} < \text{threshold}_{SD}$ or
- $CORR > \text{threshold}_{CORR}$ and $NDAI < \text{threshold}_{NDAI}$.

When the foregoing tests fail, the pixel is labeled cloudy.

The rationale of the ELCM algorithm follows directly from our choice of three features. Ice- and snow-covered surfaces under clear skies are either very smooth, with extremely small SD_{An} , or have both high correlations between different view directions and relatively low forward scattering. Clouds are rarely if ever both extremely smooth (producing low correlations between views) and relatively weakly forward scattering. After investigating the distributions of expert-labeled clear and cloudy pixels from different orbits and comparing results obtained from different thresholds to the expert labels, we concluded that thresholds for CORR and SD_{An} are stable and robust across all data units, and we set them to the empirically determined and fixed values of

$$\text{threshold}_{CORR} = .75 \quad \text{and} \quad \text{threshold}_{SD} = 2.0.$$

Unfortunately, appropriate thresholds for NDAI vary from one data unit to another, so threshold_{NDAI} must be learned adaptively.

3.2.2 The Choice of threshold_{NDAI} . We used a clustering algorithm on the current data unit and the threshold learned from previous data units at the same location to set threshold_{NDAI} . For each data unit in our first orbit, we performed a grid search of threshold_{NDAI} from $[0, 1]$ in steps of 1×10^{-5} to identify the value leading to the smallest classification error relative to the expert labels. For each data unit in later orbits, we fitted a one-dimensional mixture model of two Gaussian distributions to observed NDAIs by the EM algorithm initialized by the k -means algorithm (Dempster, Laird, and Rubin 1977). The dip between the two components is found by a numerical grid search with the same step size of 1×10^{-5} between the two fitted means in the two-component Gaussian mixture distribution. The dip, if it exists, is used as the threshold, and it usually falls in the empirically determined range of .08–.40. Otherwise, the threshold from the previous (16 days earlier) data unit is used. When fitting NDAI distributions, we trimmed off the upper and lower 2.5% tails to avoid outliers and extreme values in both tails, thereby improving the robustness of the fitting scheme. For the same robustness consideration, we used the dip in the fitted distribution instead of the Bayes rule cutoff.

An example of an NDAI distribution and the mixture model fitted to it is illustrated in Figure 6. These results are based on MISR blocks 20–22 of orbit 13490. For this data unit, the dip in the fitted distribution is quite clear, occurring at a value of .215 and falling in the expected range from .08 to .40, and the threshold from the previous visit is not needed. Using $\text{threshold}_{CORR} = .75$, $\text{threshold}_{SD} = 2.0$, and $\text{threshold}_{NDAI} =$

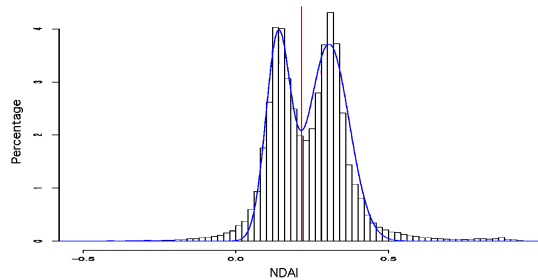


Figure 6. Histogram (vertical bars) of NDAI from blocks 20–22 of orbit 13490 together with the fit (blue line) of the two one-dimensional Gaussian functions mixture model to it. The vertical red line indicates the NDAI threshold derived from this histogram.

.215 to classify the pixels in the data unit, we obtained a 96.16% agreement rate against the expert labels. The accuracies for MISR blocks 20–22 of orbits 13257 and 13723 (Fig. 7) were 87.53% and 94.36%.

Further improvement of $threshold_{NDAI}$ selection may be achieved when the algorithm is implemented online. Because of the computational constraints in this case study, we investigated only the data collected from path 26, which led to a 16-day time lag between two visits. For global data, the time lag between two consecutive MISR visits was actually much shorter than 16 days in the polar region, because MISR paths significantly overlap spatially at high latitudes. Therefore, the threshold learned from the previous visit should perform much better than that learned 16 days earlier, and we would expect even better performance of the ELCM algorithm.

3.3 Probability Prediction by Training QDA on ELCM

The ELCM algorithm is designed to provide either clear or cloudy detection labels for each pixel with valid radiation measurements. Because the thresholds produced by the algorithm are not perfect, neither are its labels. Moreover, it is not always possible to accurately classify a pixel as clear or cloudy if it is partly cloudy, as always occurs to varying degrees near cloud boundaries. Therefore, reporting a probability of cloudiness is desirable and more informative than providing only a binary clear versus cloudy label. We used Fisher's QDA trained from ELCM labels to provide an estimate of probability, or confidence, of cloudiness.

If ELCM results show that an entire data unit is almost totally clear or cloudy with 98% of the labels in the same class, then we report just the labels. Otherwise, the ELCM labels are taken

as input data to train QDA on the three features. Of the 57 data units in this study, 32 were partly cloudy and 25 were either totally clear or cloudy.

In a two-class classification problem, QDA models each class density as a multivariate Gaussian distribution $N(\mu_s, \Sigma_s)$, where $s = 1, 2$ denotes the class label. Let π_s be the prior probability of class s . The posterior probability $P(x \in \text{class } s | x)$ is given by Bayes's rule. The parameters π_s , μ_s , and Σ_s are estimated by the empirical (i.e., training data) class proportions, means, and covariance matrices and subsequently substituted into the foregoing two equations. The estimated posterior probability $P(x \in \text{class } s | x)$ serves as the predicted probability of cloudiness for a pixel.

Compared with other methods that model only the conditional probability $P(s|x)$, such as logistic regression, QDA is more computationally efficient and robust to errors in the training labels, because it models the joint distribution $P(s, x)$. The computation of QDA involves only estimating the vector of means and the covariance matrix, which is much more efficient than iterative weighted refitting of logistic regression. In our setup, means and covariances of X are estimated from labels produced by the ELCM thresholding algorithm. These labels contain errors due to the threshold selection. However, mistakes in labels do not greatly affect the estimates of means and covariances if the mistakes are only a very small proportion of the total training sample.

Logistic regression models only the conditional distribution of $P(s|X)$. When it is fitted from labels produced from the ELCM algorithm, the logistic regression keeps faith to labels and thus is more strongly affected by mislabels in the ELCM results. Simply put, QDA works better for this problem than logistic regression. This phenomenon is also behind the recent surge of interest in semisupervised learning (Zhu, Ghahramani, and Lafferty 2003; Zhou, Schölkopf, and Hofmann 2005), which aims to take into account the distribution of the predictor, or feature, distribution. A robust version of QDA based on minimum covariance determinant estimators (cf. Rousseeuw 1985; Croux and Haesbroeck 1999) also was tested, but it did not improve the estimates or the labeling results. Thus we used ELCM-QDA to provide probability labels.

4. RESULTS

To assess the performance of the ELCM and ELCM-QDA algorithms, as well as the MISR operational SDCM and ASCM algorithms, we now compare their results against the expert labels. In addition to our proposed operational algorithms that

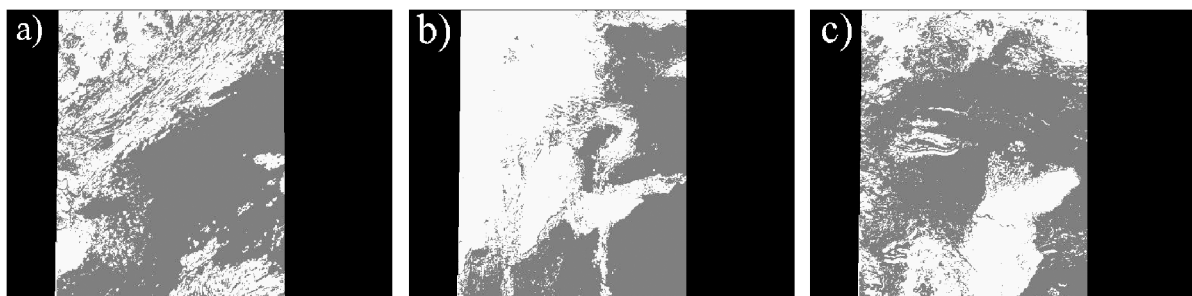


Figure 7. ELCM algorithm results for blocks 20–22 of MISR orbits (a) 13257, (b) 13490, and (c) 13723; these data units are identical to those presented in Figures 3 and 4. White represents pixels classified as cloudy; gray represents those classified as clear.

do not use expert labels (except the first orbit), we also test a simple-minded offline SVM (Vapnik 1995) classifier trained on randomly sampled expert labels to illustrate the value of our feature selection and clustering/classification framework. We discuss results from the data unit with the lowest agreement rate between the ELCM algorithm and expert labels to provide insight into the reasons why ELCM fails at certain geo-locations. Besides comparing the results from all available operational algorithms, we illustrate the separability and stability properties of the three MISR features and explain why our feature selection is necessary and works well.

4.1 Comparisons of ELCM, ELCM-QDA, ASCM, SDCM, and SVM

Among the pixels of the 57 data units under study, 5,086,002 pixels are labeled by the expert; these compose 71.5% of all pixels. Moreover, as discussed earlier, an offline Gaussian kernel SVM using MISR red radiances is also trained on 20,000 randomly selected expert labels from the first 20 data units (1,000 from each data unit) and tested on the remaining 37 data units. The accuracy of this SVM classifier demonstrates how well a state-of-the-art classifier performs when applied to this problem in a simple-minded way. The training and testing procedures of SVM follow the standard 10-fold cross-validation methods. Besides accuracies, we also report the algorithms' coverage rates, defined as the percentage of pixels for which a label is provided. By design, ELCM, ELCM-QDA, and SVM provide 100% coverage, whereas ASCM and SDCM have coverage rates of 70.12% and 26.64% over the 57 data units.

The results, given in Table 1, show that the ELCM algorithm agreement rate of 91.80% over the 5 million testing pixels is 8.57% higher than the MISR ASCM (83.23%) algorithm and 11.90% higher than the SDCM (80.00%) algorithm. This represents a significant improvement from both scientific and statistical standpoints. The offline SVM has an 80.99% agreement rate when using the expert labels for training, much lower than the ELCM algorithm (but comparable to the SDCM or ASCM algorithms). The low accuracy of offline SVM can be explained by the fact that the distributions of both the clear and cloudy pixels depend on when and where the data were collected, but a single classifier cannot be adaptive to either time or location. The ELCM algorithm is built in a data-dependent way and adapts to the changes in the distribution.

The ELCM-QDA algorithm does not improve overall agreement rates with expert labels relative to the ELCM algorithm.

Table 1. Agreement (accuracy) rates relative to expert labels and coverages of the ELCM, MISR operational SDCM and ASCM, and offline SVM algorithms

	ELCM	SDCM	ASCM	Offline SVM
Agreement with expert	91.80%	80.00%	83.23%	80.99%
Coverage	100%	26.64%	70.12%	100%

However, the ELCM-QDA algorithm goes beyond ELCM's binary labels of cloudy versus clear by providing probability labels. ELCM-QDA results from the three data units discussed across Figures 3, 4, and 7 are illustrated in Figure 8. One indication that the ELCM-QDA algorithm is performing reasonably is the frequent occurrence of probabilities of around .5 (green area in Fig. 8) at cloud boundaries (gray area in Fig. 4). We expect this result because boundary pixels are at the transition between clear and cloudy regions and this is consistent with our motivation to use probability labels.

A histogram of ELCM algorithm and expert label agreement rates for the 57 individual data units is displayed in Figure 9(a). Across all data units, the ELCM algorithm has agreement rates with expert labels at or well above 90% for most cases. Several data units have agreement rates as low as 70%, and we now analyze one of them to determine whether there are any systematic biases in the ELCM algorithm or the expert labels.

The data unit comprising blocks 17–19 of orbit 13956 has the lowest (71%) agreement rate between the ELCM algorithm results and the expert labels. We find that disagreements between the two occur mainly over regions labeled clear by the expert that were confirmed to be accurate by further investigation. The major reason for errors in the ELCM algorithm results is that CORR is low even though there is no cloud. The low correlation in this case is a result of poor terrain data registration, which usually occurs at sharp elevation changes (e.g., the coastline of Greenland). This leads to systematic errors of ELCM algorithm results over regions with rough topography. Currently, we do not have a solution to this problem, because it requires updating the MISR terrain height database, which is not planned at the present time.

4.2 Need for Feature Selection and Adaptive Thresholding

Randomly sampled expert labels are used as training data to demonstrate the separability and stability of our three features,

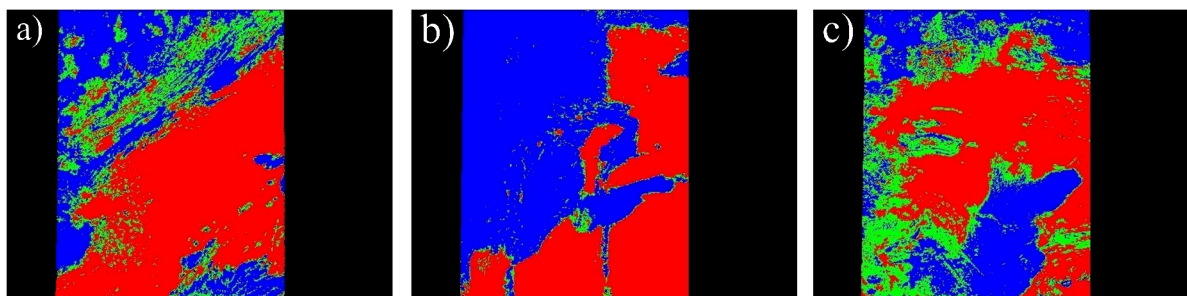


Figure 8. ELCM-QDA results for blocks 20–22 of MISR orbits (a) 13257, (b) 13490, and (c) 13723, which are the same data units illustrated in Figures 3, 4, and 7. Red is for pixels for which $P(\text{cloudy}|x) < .2$; green, for $.2 \leq P(\text{cloudy}|x) \leq .8$; and blue, for $P(\text{cloudy}|x) > .8$.

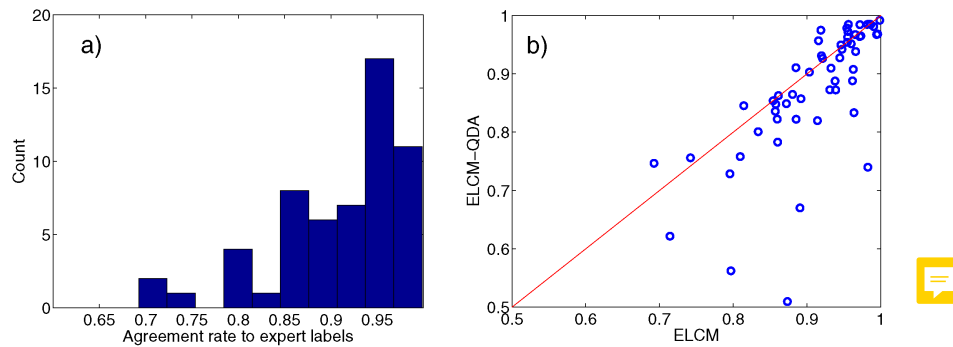


Figure 9. (a) Histogram of ELCM algorithm and expert label agreement rates for the 57 individual data units in the study. (b) Agreement rate of the ELCM-QDA algorithm with expert labels versus the agreement rate of the ELCM algorithm with expert labels for the 32 partly cloud data units in the study.

relative to the radiation measurements. To compare separability, QDA, logistic regression, and logistic regression with $L1$ penalization are trained on half of the expert labels of a data unit and tested on the other half from the same data unit. The results suggest that detectors based on features always outperform those based only on the radiation measurements. To compare stability, we apply QDA and logit trained with expert labels from one orbit (i.e., 13257) to a subsequent one (i.e., 13490) and find that the detectors based on features outperform the ones based on radiation (see details in the long version of this article at www.stat.berkeley.edu/~binyu/publications.html).

The choice of adaptively finding a threshold for NDAI is supported by the following experiment. With our features as predictors, the QDA trained on one previous visit (orbit 13257) classifies the current (orbit 13490) data unit worse, with 92.6% accuracy, than that trained on the current data unit itself, with 95.9% accuracy. As illustrated in Figure 10, degradation of detector performance is due mainly to the difference in the NDAI distributions between different visits.

5. SUMMARY, CONCLUSIONS, AND IMPACTS

This article has presented a statistical study that is but the first of many that must be pursued to accurately characterize the properties of clouds over the daylight Arctic. We have demonstrated that three physical features—the linear correlation of radiation measurements from different MISR view directions (CORR), the standard deviation of MISR nadir red radiation measurements within a small region (SD_{An}), and a normalized

difference angular index (NDAI)—contain sufficient information to separate clouds from ice- and snow-covered surfaces. The ELCM algorithm based on the three features is more accurate and provides better spatial coverage than the existing MISR operational algorithms for cloud detection in the Arctic. The ELCM algorithm combines classification and clustering frameworks in a way that makes it suitable for real-time, operational MISR data processing. Computations are sufficiently fast to handle the massive MISR data stream in an online fashion. Results from the ELCM algorithm can be used to train QDA to provide probability labels for partly cloudy scenes.

This work is significant for statistics in two ways that go beyond technical development and implementation of statistical methods. The past decade has experienced an explosion of Earth science data that support weather and climate studies. Applications of this data to current scientific problems, such as hurricane prediction and climate change, requires application of appropriate statistical methods to the data on a case-by-case basis. As such, statisticians have an important role to play in the analyses of these data if they choose to do so (Yu 2006). For this study of cloud detection in daylight polar regions, we found that three features provided sufficient separability and stability to separate clear (cloud-free) regions from clouds that a classifier no more sophisticated than QDA provided a performance comparable to that of much more sophisticated classifiers, such as SVMs. In the past, statisticians have often worked in a mode whereby they come in after the fact to work with science colleagues and develop *ex post facto* analysis method-

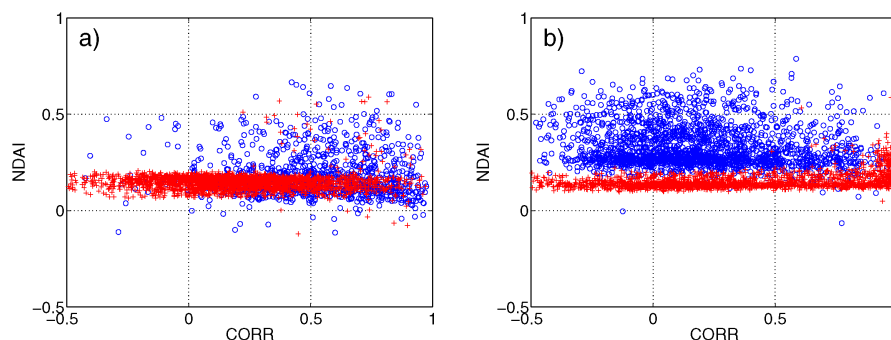


Figure 10. Scatterplots of NDAI versus CORR for MISR blocks 20–22 of orbits (a) 13257 and (b) 13490. Blue circles represent expert-labeled cloudy pixels; and red circles represent expert-labeled clear pixels.

ologies. Here, however, statisticians are directly involved in the nuts and bolts of data processing. Success was achieved only after sometimes challenging collaborations between atmospheric scientists, statisticians, and the MISR science and instrument teams at the Jet Propulsion Laboratory.

A second significant aspect of this research is that it demonstrates the power of statistical thinking, and also the ability of statistics to contribute solutions to modern scientific problems. Of course, this is not in dispute in the statistics community, but our discipline's contributions are less clear when viewed from the standpoint of others (e.g., NASA managers) not trained as we are. Many do not understand what statistics has to offer beyond that offered by statistically savvy computer and atmospheric scientists. This work shows that by careful combination and application-specific modification of proven statistical methods, one can craft effective, innovative solutions to apparently complex scientific problems. It would not be possible to develop such a strategy without a fundamental understanding of statistical principles, nor would it be possible without a willingness to become immersed in the scientific fundamentals and issues of a specific field.

From a scientific standpoint, there is no question that the ability to detect clouds in polar scenes is a topic of great importance and interest. The year 2007 is ushering in a major international initiative to study and better understand Earth's processes at the poles—the International Polar Year (IPY) of 2007. The previous IPY, in 1957–58, involved some 80,000 scientists from 67 countries. Fifty years later, the current IPY will include information from satellite sensors, including MISR. Improved cloud masks from MISR, combined with those from the MODIS also on the Terra satellite, will enable a range of improved cloud and radiation studies (Shi, Clothiaux, Yu, Braverman, and Groff 2007). The ultimate aim of this cloud research is improved understanding of the flow of visible and infrared radiation through the atmosphere, so we can begin to tease apart the response of clouds to changes in arctic climate and their feedback on it. More reliable polar cloud properties also will translate into more accurate global climate model simulations through improved model cloud physics. These studies will eventually enable the scientific community to study how changing cloud properties may enhance or ameliorate any initial changes in the Arctic brought about by increasing concentrations of atmospheric carbon dioxide.

[Received August 2006. Revised April 2007.]

REFERENCES

- Clothiaux, E. E., Barker, H. W., and Korolev, A. V. (2005), "Observing Clouds and Their Optical Properties," in *3D Radiative Transfer in Cloudy Atmospheres*, eds. A. Marshak and A. B. Davis, Berlin: Springer, pp. 93–150.
- Croux, C., and Haesboeck, G. (1999), "Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator," *Journal of Multivariate Analysis*, 71, 161–190.
- Dempster, A., Laird, N., and Rubin, D. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 39, 1–38.
- Di Girolamo, L., and Davies, R. (1994), "A Band-Differenced Angular Signature Technique for Cirrus Cloud Detection," *IEEE Transactions on Geoscience and Remote Sensing*, 32, 890–896.
- Diner, D. J., Asner, G. P., Davies, R., Knyazikhin, Y., Muller, J.-P., Nolin, A. W., Pinty, B., Schaaf, C. B., and Stroeve, J. (1999a), "New Directions in Earth Observing Scientific Applications of Multangle Remote Sensing," *Bulletin of the American Meteorological Society*, 80, 2209–2228.
- Diner, D. J., Braswell, B. H., Davies, R., Gobron, N., Hu, J. N., Jin, Y. F., Kahn, R. A., Knyazikhin, Y., Loeb, N., Muller, J.-P., Nolin, A. W., Pinty, B., Schaaf, C. B., Seiz, G., and Stroeve, J. (2005), "The Value of Multiangle Measurements for Retrieving Structurally and Radiatively Consistent Properties of Clouds, Aerosols, and Surfaces," *Remote Sensing of the Environment*, 97, 495–518.
- 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," JPL Technical Document D-11399, rev. D, Jet Propulsion Laboratory, Pasadena, CA.
- Fisher, A. R. (1925), *Statistical Methods for Statistical Workers*, Edinburgh: Oliver and Boyd.
- Francis, J. A., Hunter, E., Key, J. R., and Wang, X. (2005), "Clues to Variability in Arctic Minimum Sea Ice Extent," *Geophysical Research Letters*, 32, L21501, doi:10.1029/2005GL024376.
- Giorgi, F., and Bi, X. (2005), "Updated Regional Precipitation and Temperature Changes for the 21st Century From Ensembles of Recent AOGCM Simulations," *Geophysical Research Letters*, 32, L21715, doi:10.1029/2005GL024288.
- Kato, T., Tang, Y., Gu, S., Hirota, M., Du, M., Li, Y., and Zhao, X. (2006), "Temperature and Biomass Influences on Interannual Changes in CO₂ Exchange in an Alpine Meadow on the Qinghai-Tibetan Plateau," *Global Change Biology*, 12, 1285–1298.
- Mardia, K. V., Kent, J. T., and Bibby, S. M. (1979), *Multivariate Analysis*, New York: Academic Press.
- 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 Transactions on Geoscience and Remote Sensing*, 40, 1605–1615.
- Rousseeuw, J. P. (1985), "Multivariate Estimation With High Breakdown Point," in *Mathematical Statistics and Applications*, eds. W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Dordrecht: B. Reidel, pp. 283–297.
- Shi, T., Yu, B., and Braverman, A. J. (2002), "MISR Cloud Detection Over Ice/Snow Using Linear Correlation Matching," Technical Report 630, University of California Berkeley, Dept. of Statistics.
- Shi, T., Clothiaux, E. E., Yu, B., Braverman, A. J., and Groff, D. N. (2007), "Detection of Daytime Arctic Clouds Using MISR and MODIS Data," *Remote Sensing of Environment*, 107, 172–184.
- Stephens, G. L., Campbell, G. G., and Vonder Haar, H. T. (1981), "Earth Radiation Budgets," *Journal of Geophysical Research*, 86, 9739–9760.
- Vapnik, V. (1995), *The Nature of Statistical Learning Theory*, New York: Springer.
- Yu, B. (2006), "Embracing Statistical Challenges in the Information Technology Age," Technical Report 706, University of California Berkeley, Dept. of Statistics.
- Zhou, D., Schölkopf, B., and Hofmann, T. (2005), "Semi-Supervised Learning on Directed Graphs," in *Advances in Neural Information Processing Systems*, eds. L. K. Saul, Y. Weiss, and L. Bottou, Cambridge, MA: MIT Press, p. 17.
- Zhu, X., Ghahramani, Z., and Lafferty, J. (2003), "Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions," *ICML*, 20.