

# Lab 4 - Cloud detection - Stat 215A, Fall 2017

Olivia Angiuli, Miyabi Ishihara, Yizhou Zhao

November 2, 2017

## 1 Introduction

In 1999, NASA's Multiangle Imaging SpectroRadiometer (MISR) data generated a massive data set of Arctic images that help provide insight into recent global changes of ice coverage due to increasing surface air temperatures. The dataset is rich, generated by a satellite that collected images along 233 different geographical paths, with 180 blocks in each path, providing images of any given region every 16 days.

One of the biggest statistical challenges of this dataset is to conclude, for each image, which regions are ice-covered, as opposed to snow-covered or simply unknown. Using only a small subset (three images) of this data, each of which produced radiances from 5 different angles, we use expert labels in order to train and evaluate classification models that can help differentiate between icy versus cloudy surfaces.

## 2 Exploratory Data Analysis

### 2.1 Data background

We are provided with three images, each of which provide the following information:

- **Expert labelling.** Experts visually inspected each image, classifying each pixel as cloudy (+1), not cloudy (-1), or unlabelled (0). These are the labels with which we will train our supervised algorithms, and with which we will test our results.
- **Radiances from five angles.** We were given satellite images for three different 275m x 275m regions, each from five different angles as shown below (where "F" indicates the forward direction of flight and "A" represents the rear direction):

As will be described in the next section, differences in radiance measurements of the same image from different angles can tell us about the presence or absence of clouds and/or ice.

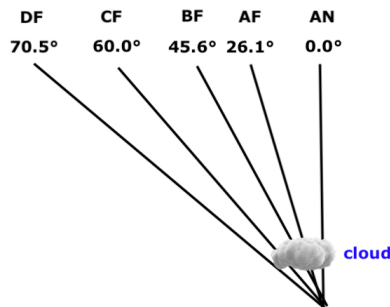


Figure 1: The five angles from which we have radiance measures, for each of the three images.

- **Feature values.** Yu et. al. developed three features that capture spatial information that helps distinguish between ice and clouds:

- **CORR** - *an average linear correlation of radiation measurements at different view angles.*

A cloud may obstruct the view of underlying ice from some angles but not others. Figure 1, above, illustrates a scenario in which DF provides an unobstructed view of the underlying image that the other angles cannot provide.

This suggests that low CORR images often suggest clouds: however, low-altitude clouds or smooth cloud-free areas may break this trend, requiring the following two features.

- **SD** - *standard deviation within groups of MISR.*

SD helps identify smooth surfaces (areas with low standard deviations), as well as to identify a baseline for background measurement error that can be filtered out in the model.

- **NDAI** - *a proxy for surface roughness, measured by the normalized difference between measurements in the forward versus backward pointing cameras.*

The intuition is that a low-altitude cloud has more roughness than an ice- or snow-covered surface. This feature combines with CORR to help differentiate between low CORR images that have low-altitude clouds versus are cloud-free.

## 2.2 Data visualization

The following plots visualize the three provided images. The first set of images shows the raw data from the perspective of the AN-camera. The second set of images shows the same three images, classified by whether an expert determined that a given area was ice (pink), unknown (green), or cloudy (blue).

Raw satellite data of snow/ice, from AN-camera

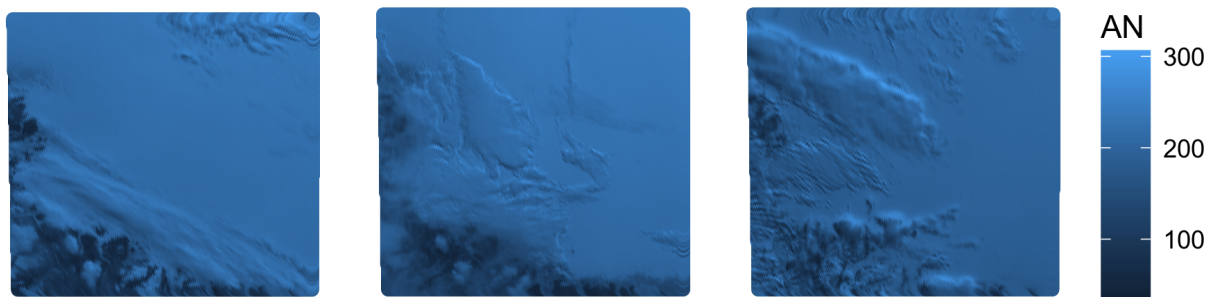


Figure 2: Raw (from AN-camera) image data for the three provided images.

Expert-labelled satellite data of snow/ice

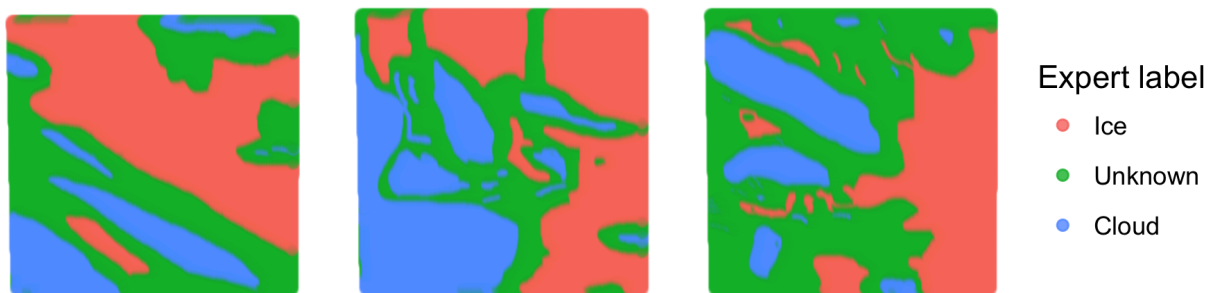


Figure 3: Expert-labelled image data for the three provided images.

### 2.3 Do different angles provide different information?

Yes! Below, we show image 1 from all 5 different angles that we are provided in the data set: from left to right,  $70.5^\circ$  (DF),  $60.0^\circ$  (CF),  $45.6^\circ$  (BF),  $26.1^\circ$  (AF) in the forward directions, and  $0^\circ$  (AN).

One can spot higher resolution in the bottom left corner (which corresponds to mostly cloudy and/or unknown regions) as the angles becoming increasingly acute. This supports the reasoning for Yu et. al.’s construction of the “CORR” and “NDAI” features: cloudy regions might have more variation in radiances between different angles.

Image 1, taken from increasingly acute angles

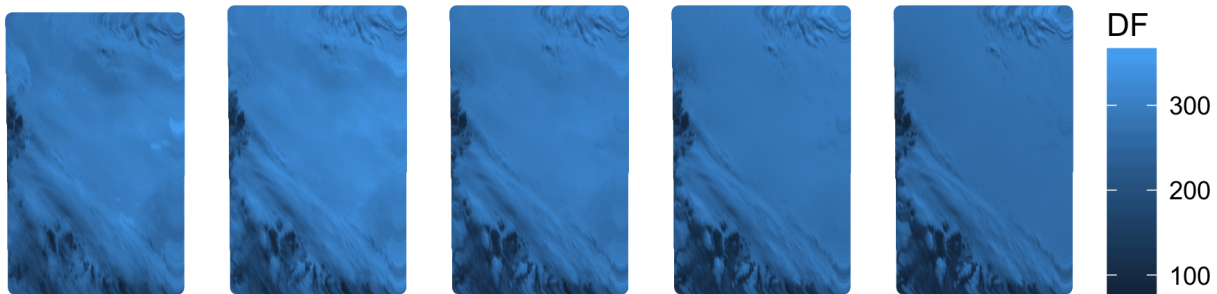


Figure 4: Image 1, by increasingly acute angles.

Plotting the conditional densities of cloud- versus ice-covered regions confirms that different angles do indeed help differentiate between cloud and ice. We notice two trends in particular. First, icy regions tend to occur at a resolution of around 275. This can effectively set a high “prior” for iciness for pixels with a similar resolution. Second, we notice that cloudy regions have a much wider range of resolutions that change with the angle of measurement. Again, this confirms Yu et. al.’s intuition that differences between measurements from different angles often correspond to cloudiness.

Image 1 conditional densities, by measurement angles

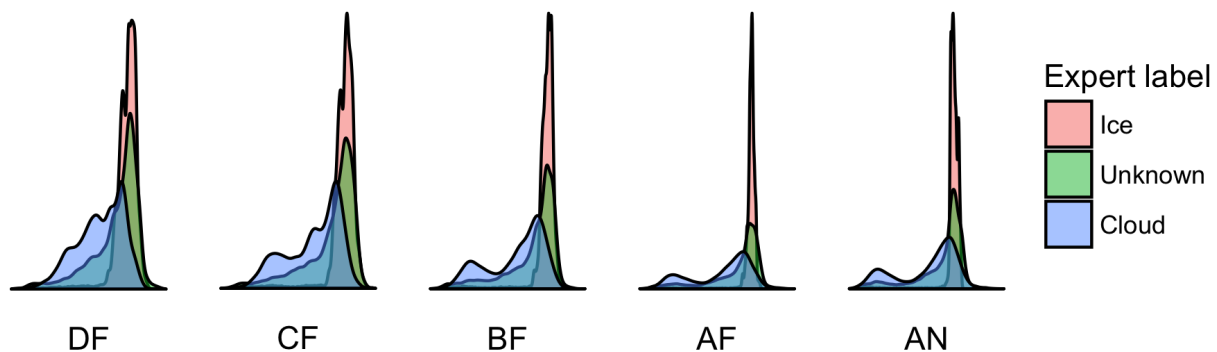


Figure 5: Conditional densities of ice vs. cloud vs. unknown by different angles. Cloud resolutions become increasingly bimodal as measurement angle decreases.

### 2.4 How do features help distinguish between cloud and ice?

Recall that the three features, as described in detail above, are:

- CORR (correlation of images from different view angles)
- SD (the standard deviation within groups of MISR)

- NDAI (a proxy for surface roughness, measured by the normalized difference between measurements in the forward vs. backward pointing cameras)

By itself, NDAI has the most discriminatory power between the presence of ice and the absence of ice (either cloudy or unknown). This can be seen by the distinct peaks in the leftmost graph below.

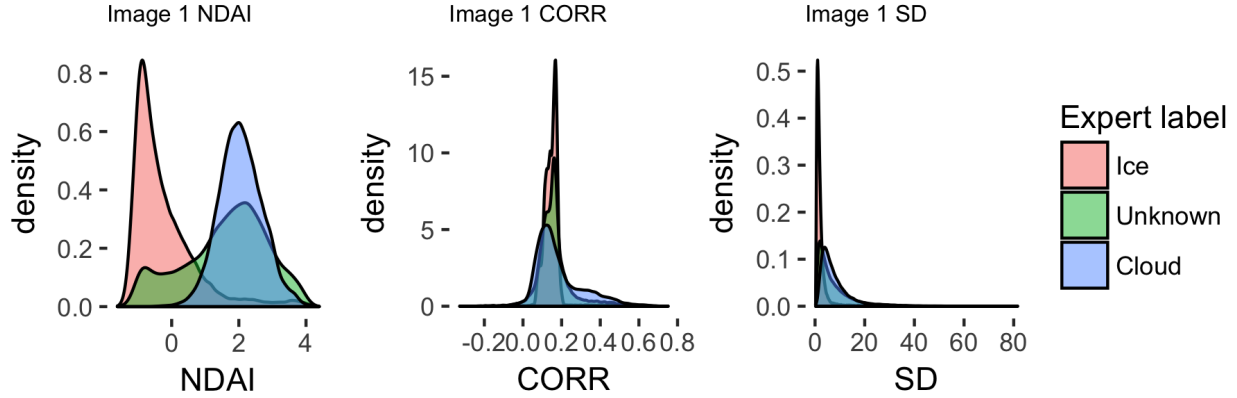


Figure 6: Conditional densities of ice vs. cloud vs. unknown by different angles. Cloud resolutions become increasingly bimodal as measurement angle decreases.

The discriminatory power of NDAI can also be seen from the following side-by-side plot of NDAI resolution values versus the true classes of image 1. Note that higher values of NDAI tend to correspond to “no ice” regions.

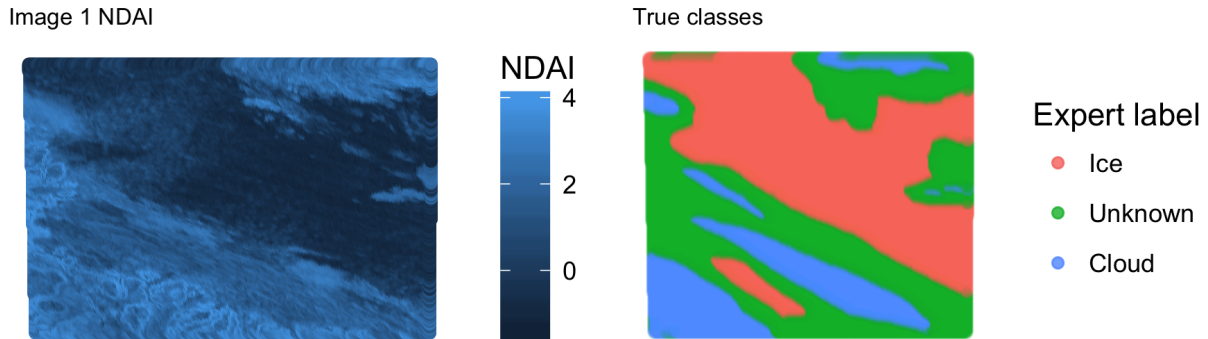


Figure 7: High NDAI values tend to correspond to unknown or cloudy regions; low NDAI values tend to correspond to icy regions.

Yu et. al. also discuss how NDAI and CORR, together, provide even more information than the two features alone: they hypothesize that high CORR + low NDAI images would more strongly suggest ice than high CORR alone. This relationship, though, is not readily seen in the following pairwise feature plots, largely because there are few high CORR pixels that correspond to ice-covered regions.

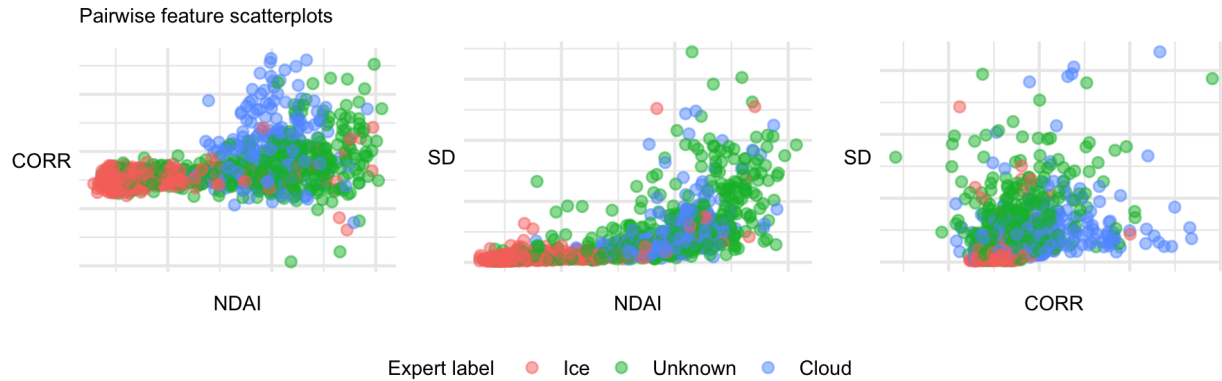


Figure 8: High NDAI values tend to correspond to unknown or cloudy regions; low NDAI values tend to correspond to icy regions.

### 3 Classifying ice- versus snow-covered surfaces

#### 3.1 Baseline approach: random forests

#### 3.2 Approach 2

#### 3.3 Approach 3

### 4 References

- (1) Shi, Tao, et al. "Daytime arctic cloud detection based on multi-angle satellite data with case studies." *Journal of the American Statistical Association* 103.482 (2008): 584-593.