Arctic Cloud Detection based on Multi-angle and Hyperspectral Satellite Images

Bin Yu

Statistics Department
University of California, Berkeley

http://www.stat.berkeley.edu/users/binyu

Joint work with Tao Shi, Eugene Clothiaux, Amy Braverman

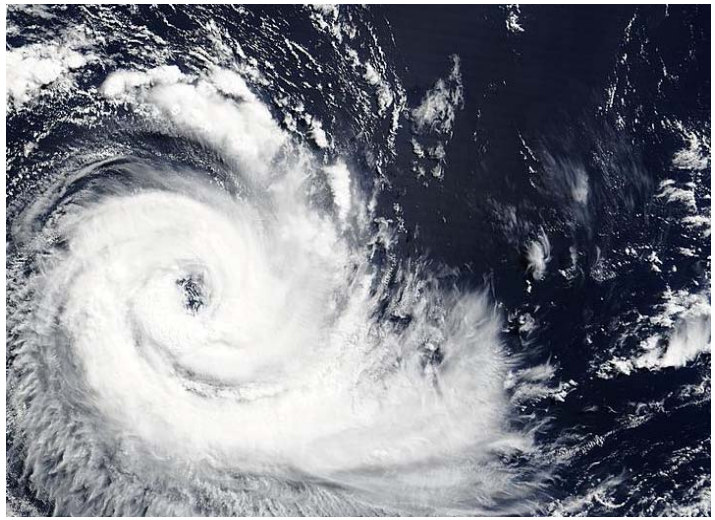
Today

- **Arctic Cloud Detection, Terra, MISR, MODIS**
- **Our Approach: Three Features, ELCMC, QDA**

  Comparing ELCMC with Expert Labels over 37 blocks of Data: 87% average accuracy
- **Fusing ELCMC with MODIS Labels: 94% average accuracy**
- **Conclusions and Future work**

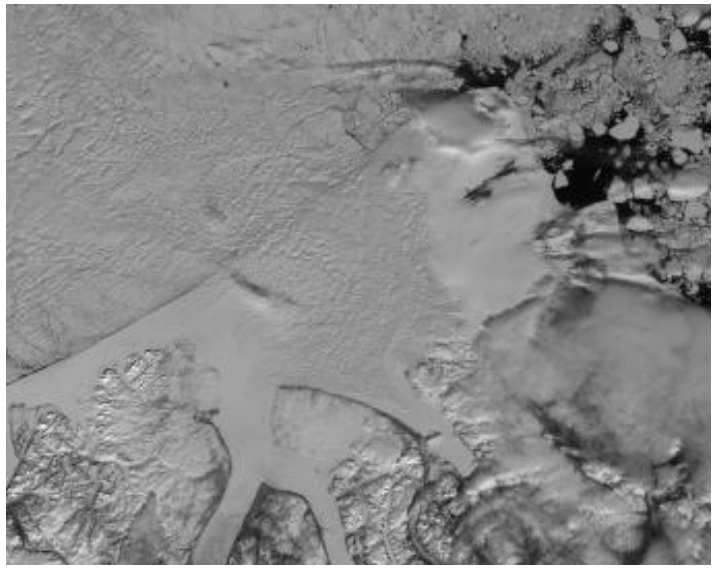
I.Cloud Detection over Arctic Regions

- **Uncertainties about cloud radiation feedback on the global climate are among the greatest obstacles in understanding and predicting earth's future climate.**

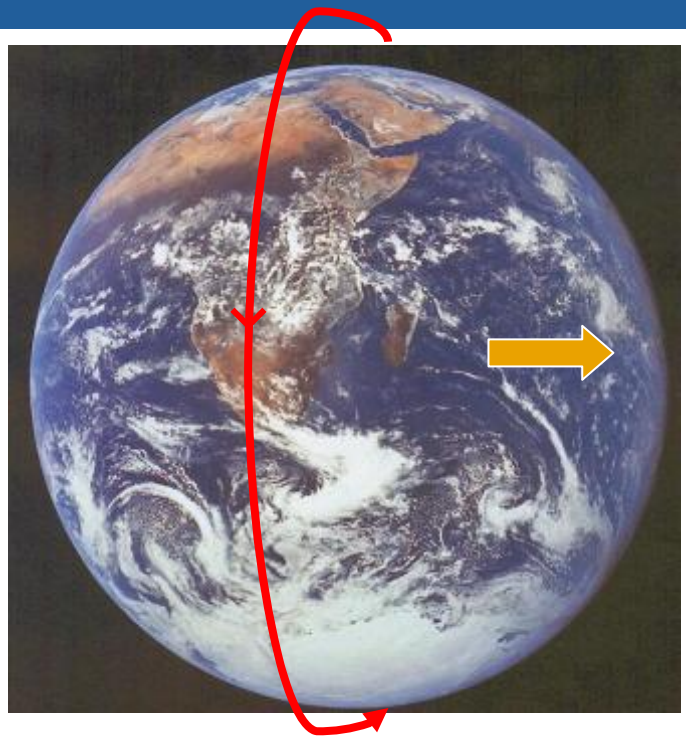


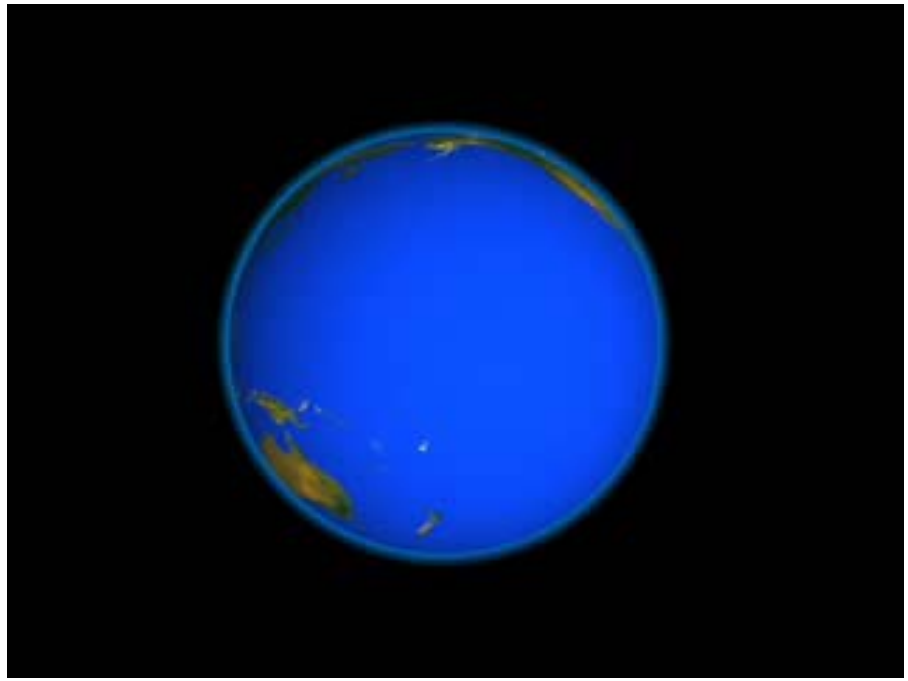
- **Clouds above snow- and ice-covered surfaces are especially difficult to detect because their temperature and reflectivity are similar to those of the surface.**

Human expert labels are used as “ground truth”, but expensive and not available on line.



EOS and its First Satellite: TERRA

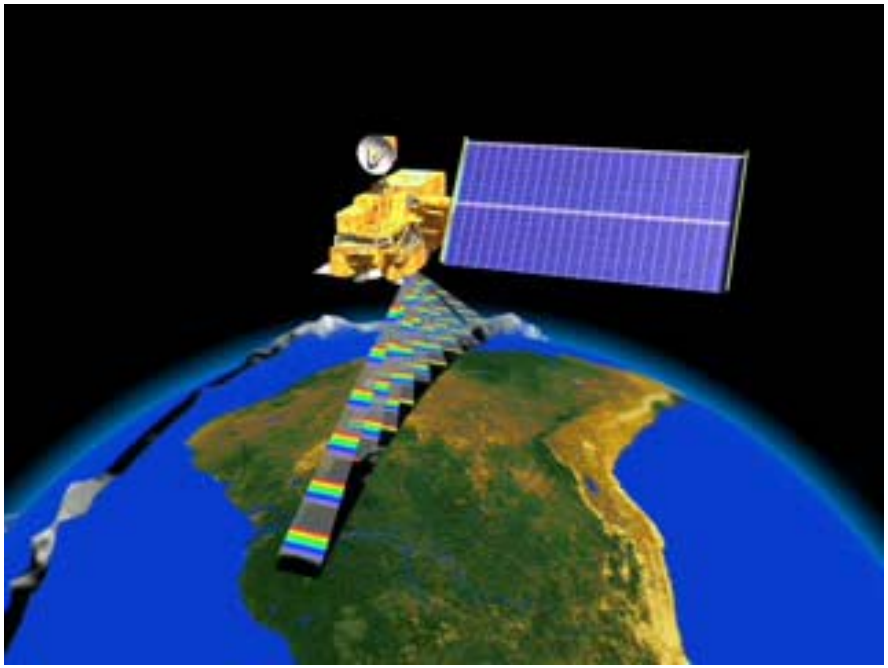




- **Earth Observing System (EOS) is designed by NASA to study earth from space with a multiple-instrument, multiple-satellite approach. Its goal is to improve the scientific understanding of global climate changes and provide the scientific basis for environmental policies.**

- **TERRA, the first EOS satellite, was launched on December 18, 1999.**

Cloud Detection Based on MISR Images



MISR has 9 angles

0º *(AN)*,
± 26.1º *(AF, AA)*,
± 45.6º *(BF, BA)*,
± 60º *(CF, CA)*,
± 70.5º *(DF, DA)*

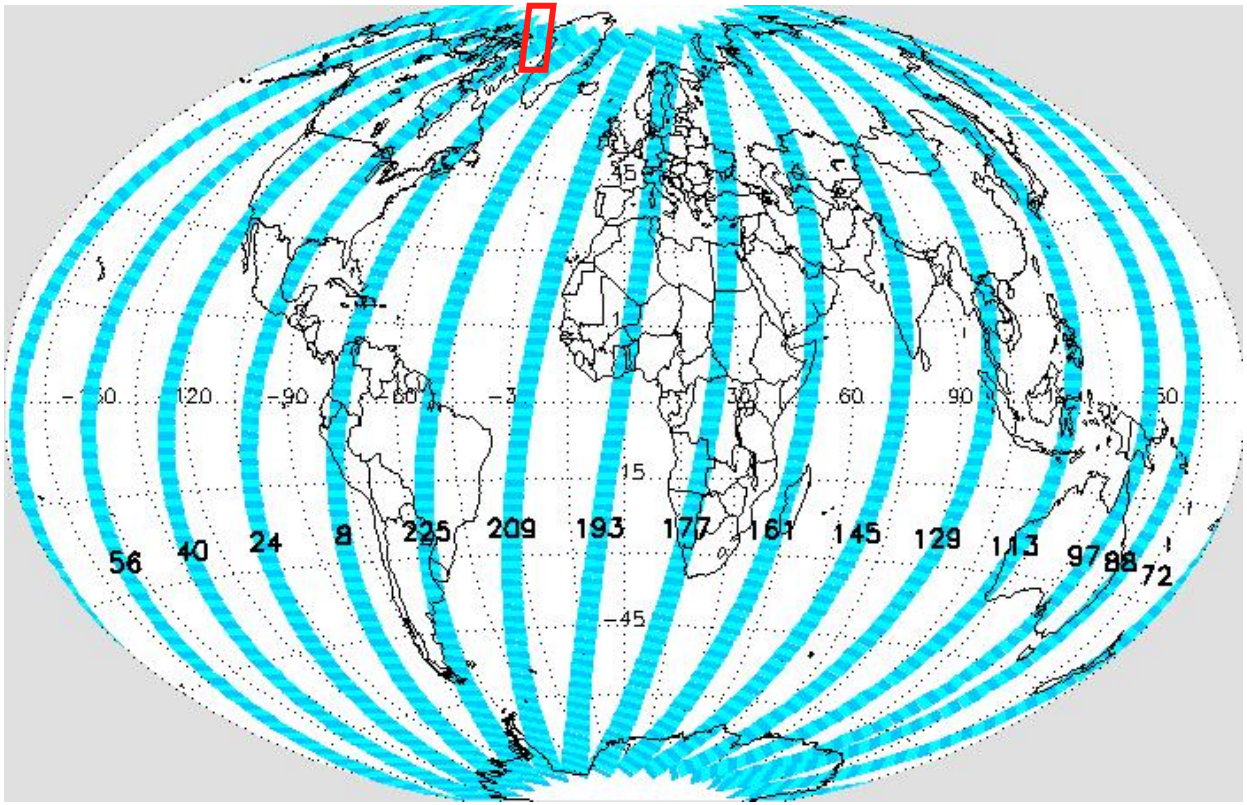
- **4 wavelengths in each angle. (443nm, 555nm, 670nm, and 865nm Near Infrared Red)**

- **Multi-angle Imaging Spectre Radiometer (MISR) was launched by NASA on December 18, 1999.**

- **Built and maintained for NASA by the Jet Propulsion Laboratory (*JPL*) in Pasadena, California.**

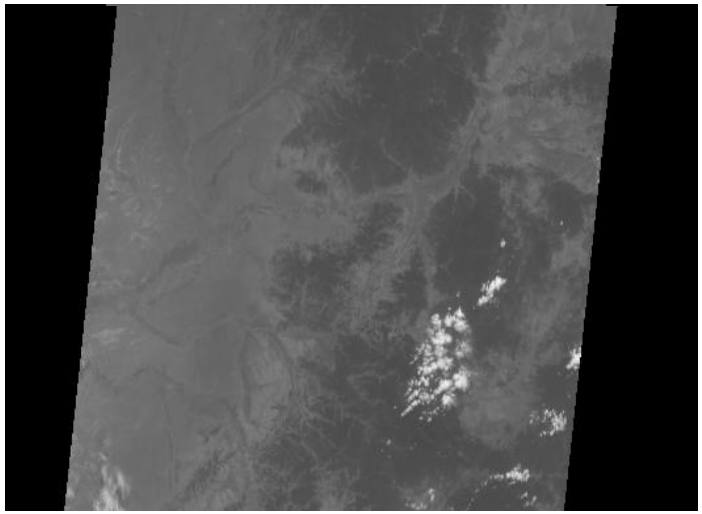
Challenges

- **Organization, transmission, and visualization of these massive data (MISR: 3.3 megabits/s on average and 9.0 megabits/s peak time)**
- **Streaming data or online processing**
- **Data fusion among EOS data sources**



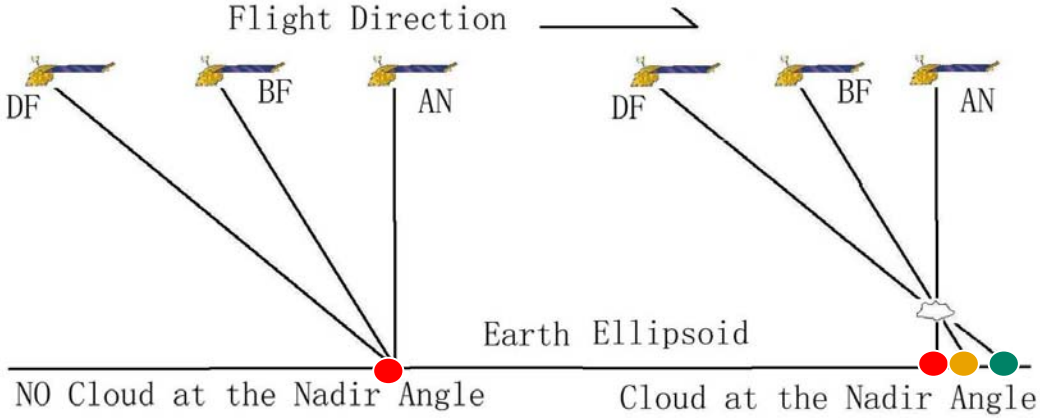
- **Orbit: a path from north pole to south pole (233 total)**
- **Block: each orbit is divided into 180 blocks, 1 degree for each.**
- **Block size: about 330 km in width, and 128 km in height.**

MISR Operational Cloud Detection Algorithm



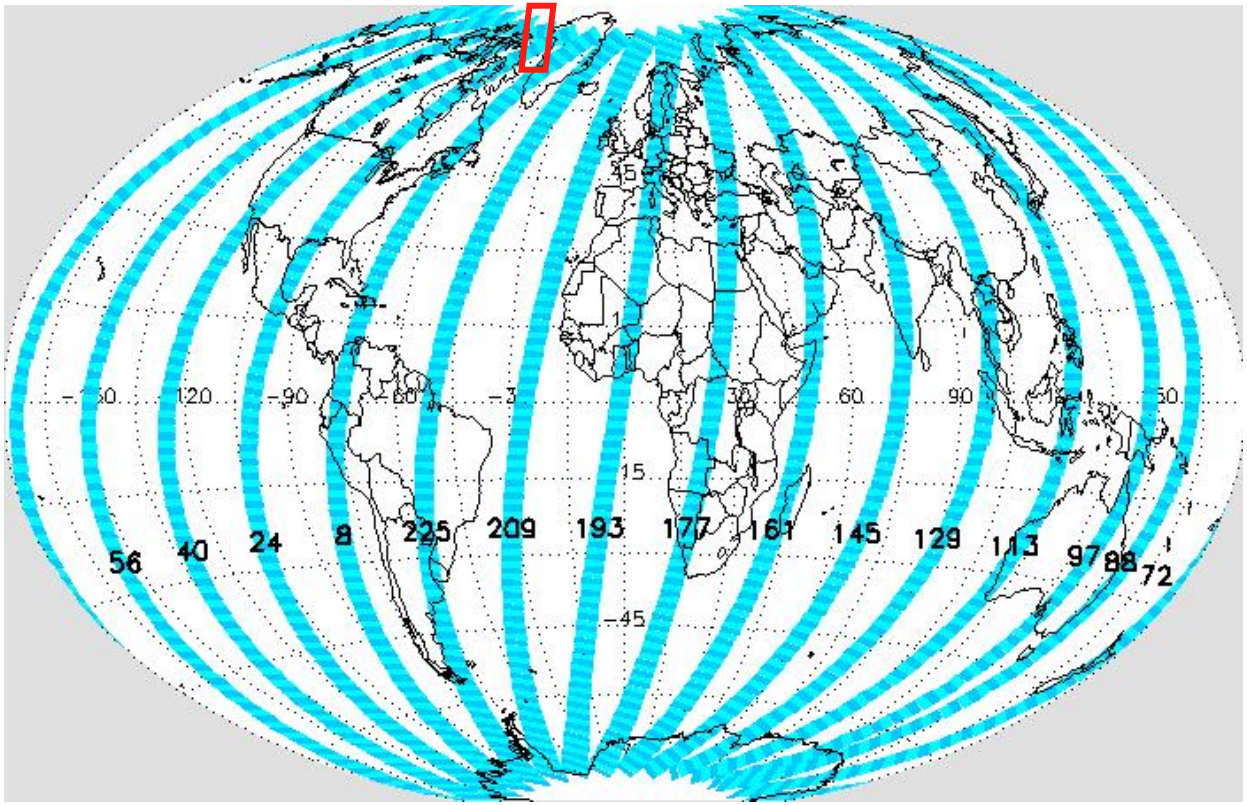
Clouds are registered at different ground locations in different angles.

The MISR operational SDCM retrieves the cloud height and cloud movement by matching the same clouds in three angles.



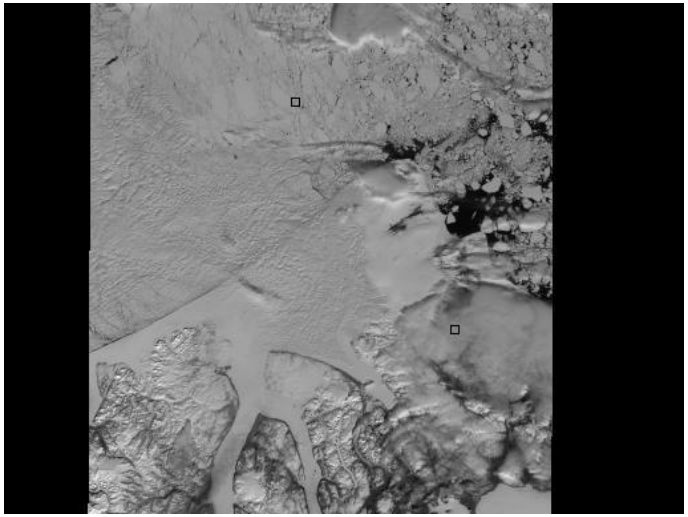
The algorithm works well over dark surfaces, such as deep ocean and vegetation covered land surface, but does not work well over snow and ice covered surfaces because good matching is very difficult.

Our study area

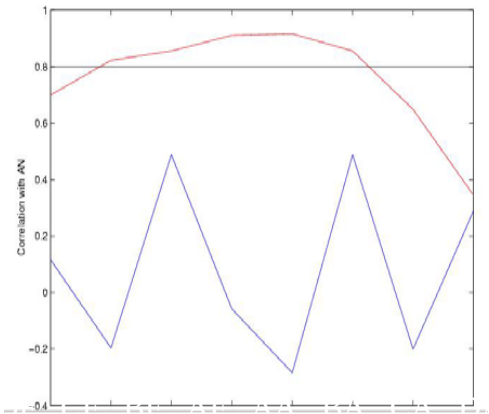


- **Data collected over Greenland and Baffin Bay on TERRA path 26 during 2002 summer season.**
- **Dr. Eugene Clothiaux hand-labeled 11 orbits using MISR and MODIS radiances for validation of different algorithms.**

Linear Correlation Matching Clustering (LCMC): Looking for Ice/Snow, not Clouds



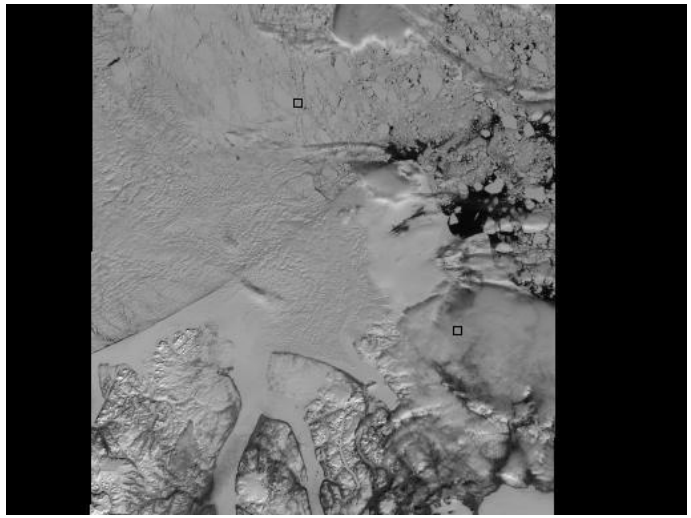
MISR red AN image collected over Greenland on June 21, 2001.

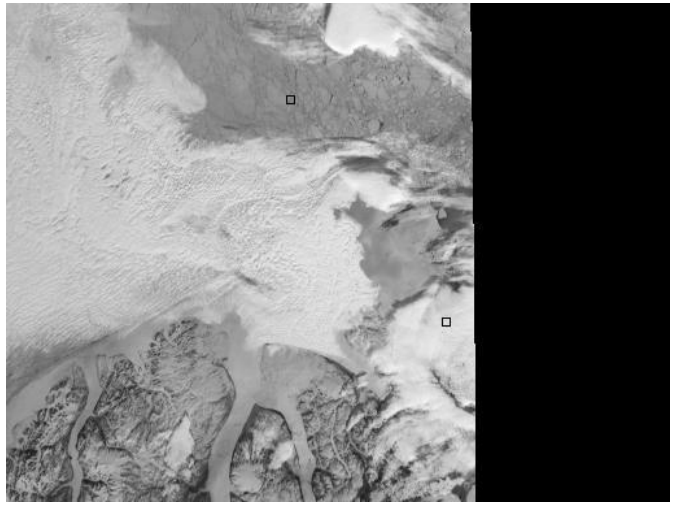


- **Correlations between angles are strong over snow- and ice-covered surfaces, and weak in areas covered by high cloud.**
- **Clouds are brighter than snow and ice covered surfaces in forward angles**.

Correlation of AN red radiance with red radiances at other angles.

Problems of LCMC





MISR red images collected over Greenland on June 21, 2001. Left: AN. Right: DF.

Correlation fails where

- **the surface is smooth (e.g. frozen rivers)**
- **there are thin clouds**

Remedies: SD to measure smoothness
Forward scattering via DF to deal with thin clouds

ELCMC: Enhanced Linear Correlation Matching Clustering

ELCMC thresholds 3 features based on 275m, terrain projected red radiances. It is a CLUSTERING ALGORITHM AND EXPERT LABELS ARE NOT USED.

1. ***correlation between angles*** ($CORR = (r_{AF\text{-}AN}+r_{BF\text{-}AN})/2$)

2. ***surface smoothness*** (SD_{AN})

3. ***the angular signature of the radiances from different angles***: ***Normalized Difference Angular Index (NDAI) see Nolin, Fetterer, and Scambos (2002)***

$$NDAI = \frac{Radiance_{DF} - Radiance_{AN}}{Radiance_{DF} + Radiance_{AN}}$$

ELCMC: on-line implementation details

1. Data unit = three blocks (to reduce all clear or all cloudy situation)

2. Thresholds fixed for SD (2.0) and CORR (0.75)

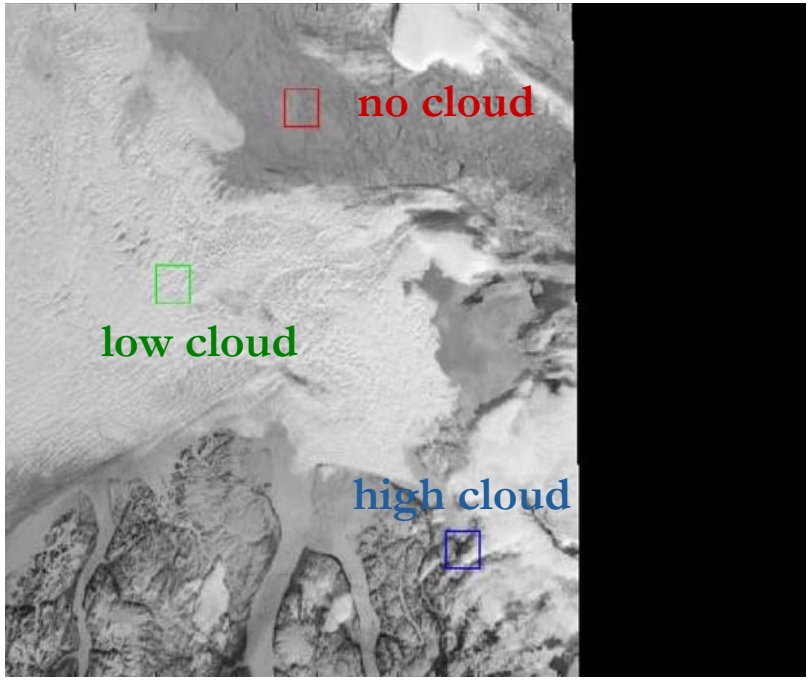
3. NDAI EM initialization is based on expert labels (only once)

   Run EM on current data unit on middle 95% data and use the "dip" between two modes as candidate threshold hold.

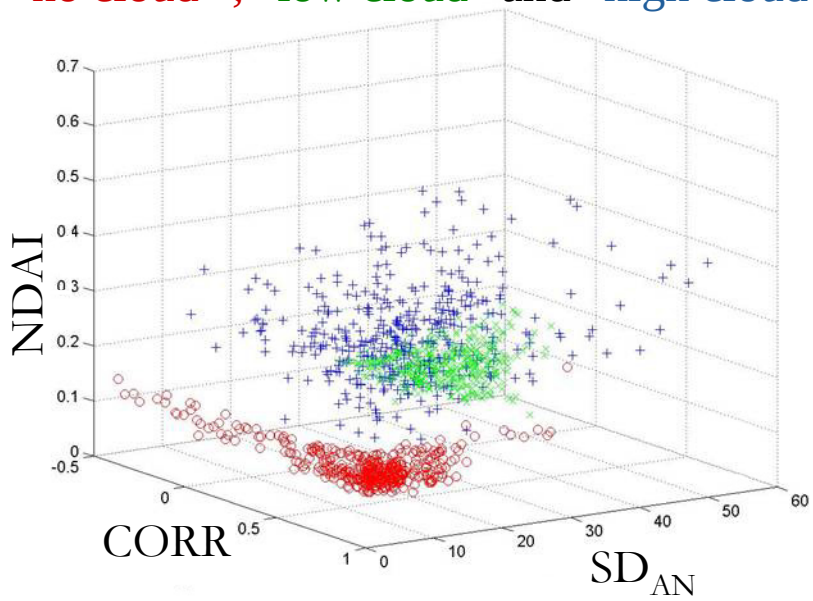
   If candidate threshold falls within (0.08, 0.4), use it; otherwise use the threshold from previous visit.

Processing time for one data unit is 1 or 2 seconds on a laptop.

ELCMC (continued)



"no cloud" , "low cloud" and "high cloud"

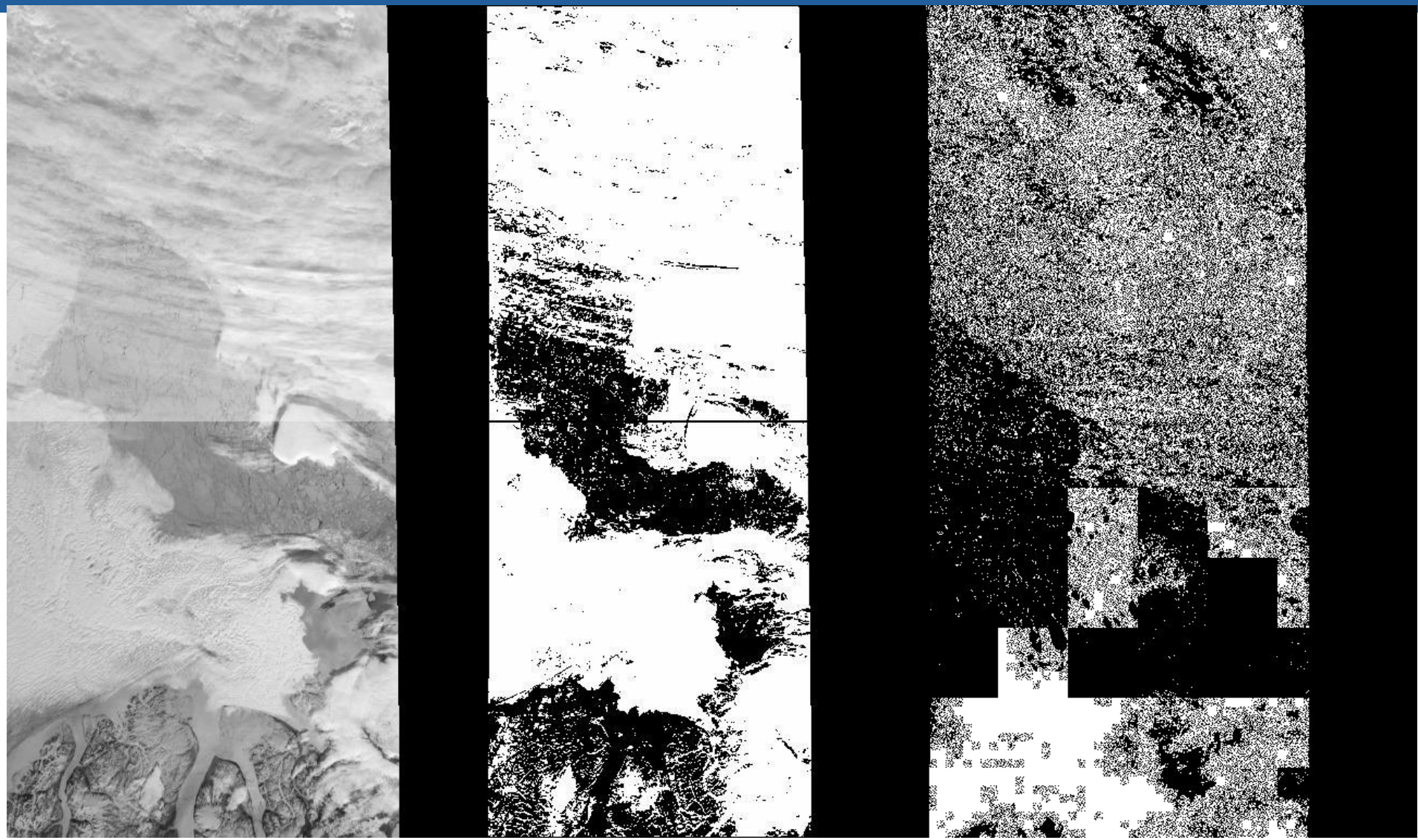


Declare "clear (snow and ice)" when

1. $SD_{AN} < threshold_{sd}$ *or*
2. $CORR > threshold_{corr}$ *and* $NDAI < threshold_{ndai}$

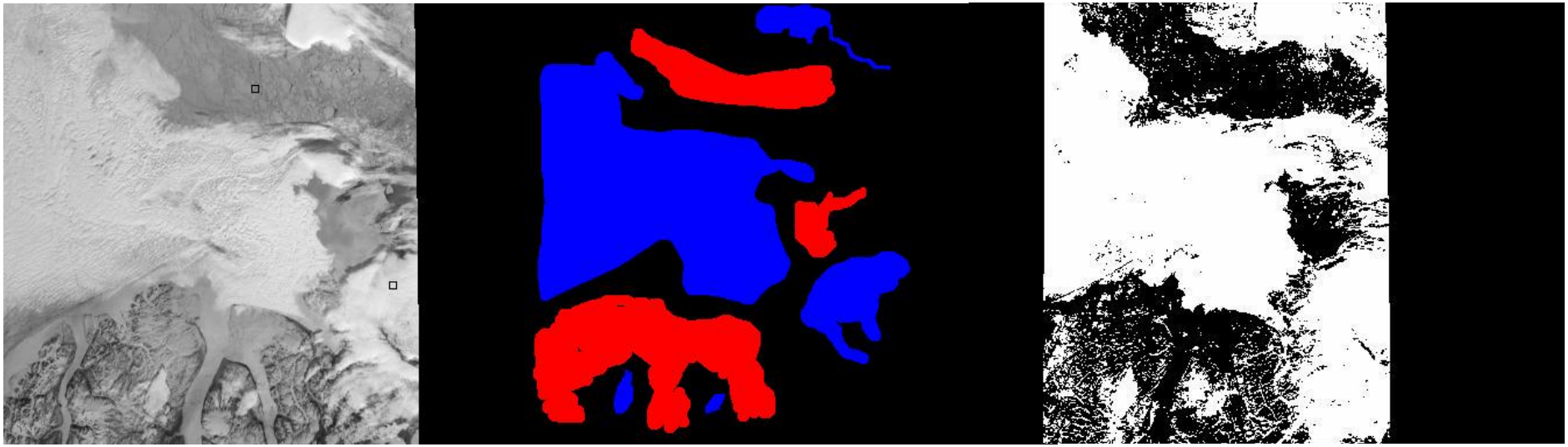
Otherwise "cloudy". And thresholds are either fixed or adaptively chosen.

An Example of ELCMC Result



DF red image **ELCMC** **MISR result**

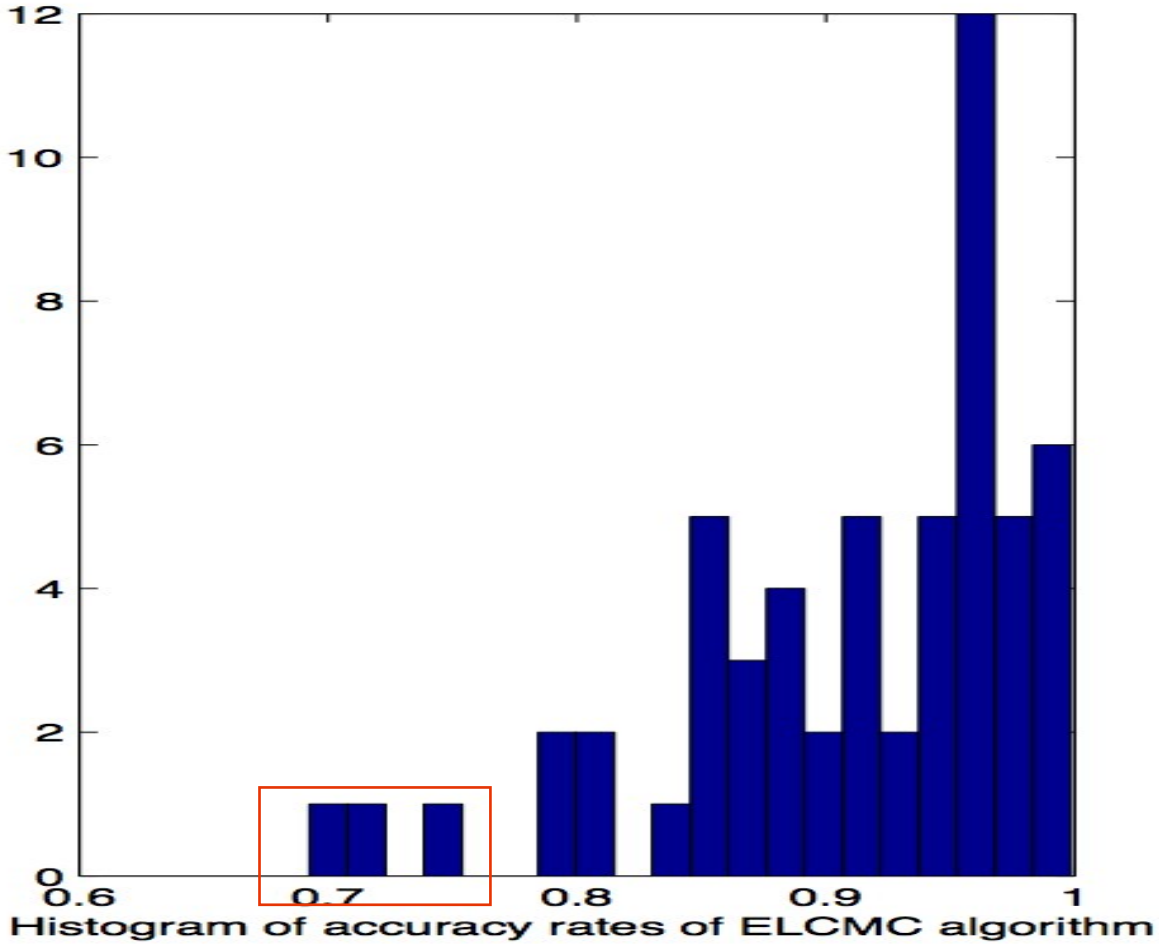
Accuracy against Expert Labels



DF red image | Expert labels | ELCMC cloud mask

On this data set, ELCMC algorithm provides 93.95% accuracy on expert labels, comparing with a 46.75% accuracy rate of the MISR operational algorithm.

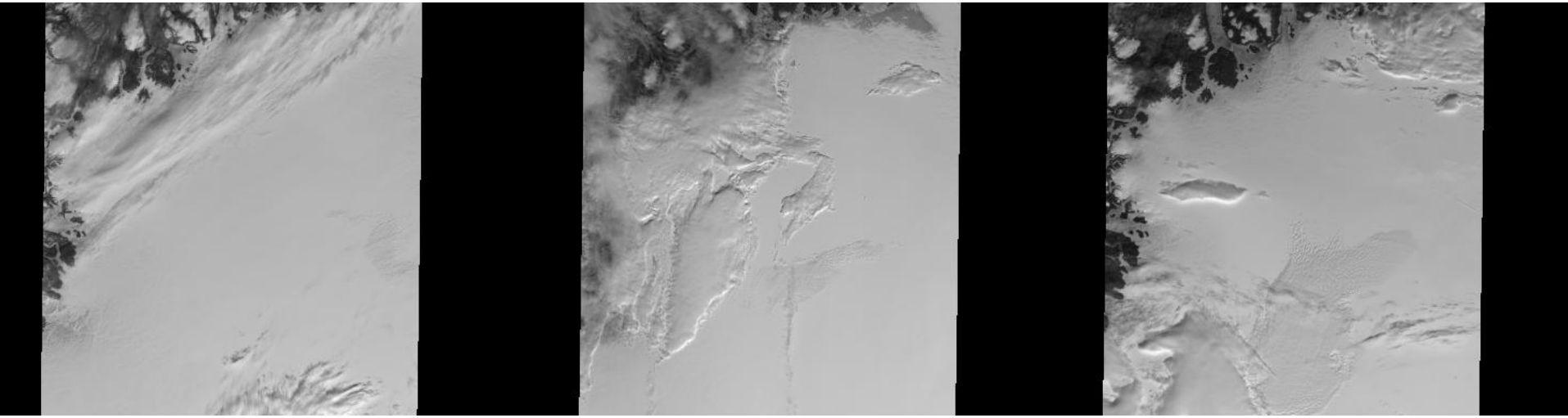
Overall Accuracy against Expert Labels



Over all 60 blocks we tested, ELCMC algorithm provides a 92% of accuracy, but it has very low accuracy (~70%) over a few blocks.

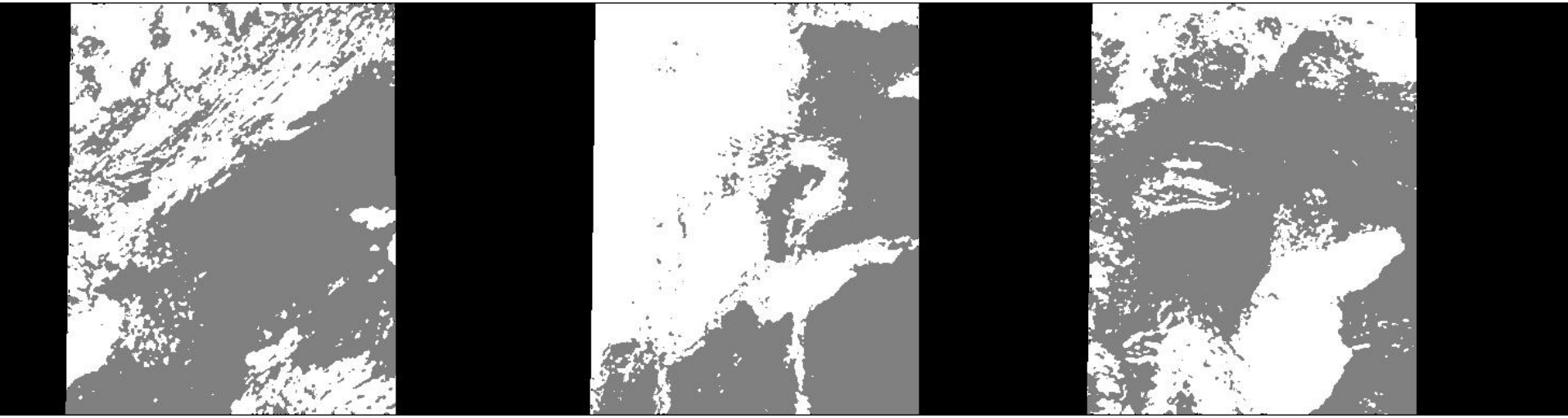
More Data and Expert Labels

Data collected over same location but in different days (9-day gap)



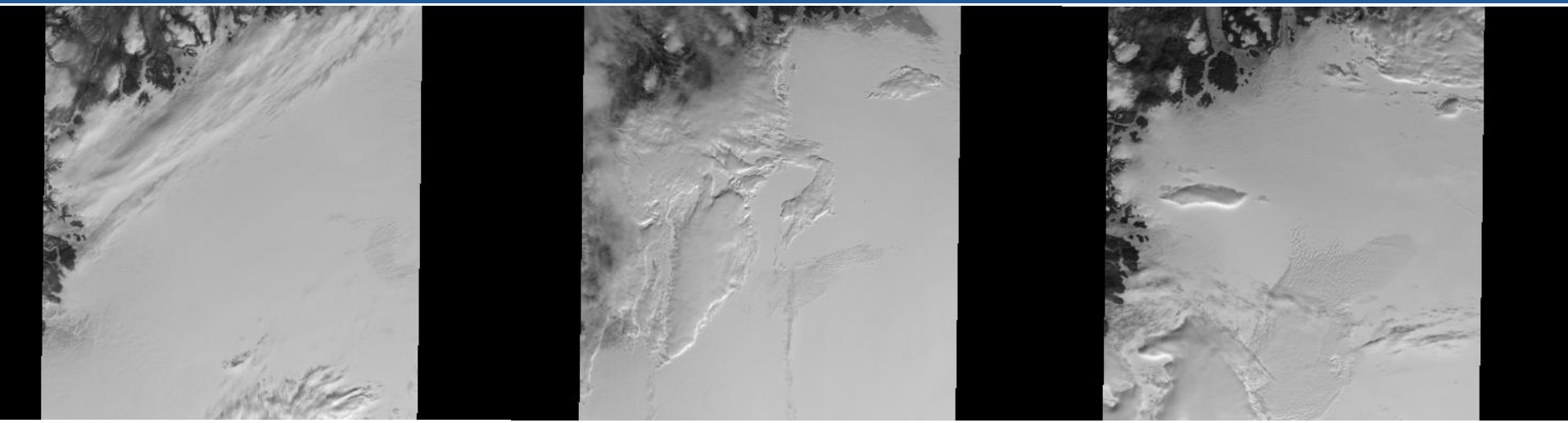
Orbit 13257 Block 20-22 | Orbit 13490 Block 20-22 | Orbit 13723 Block 20-22

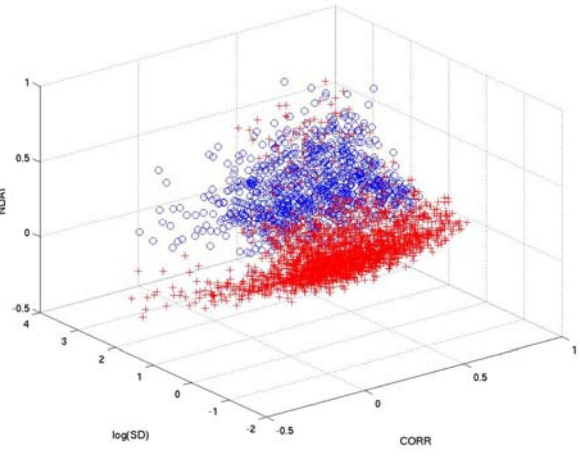
Comparing ELCMC maps with expert maps

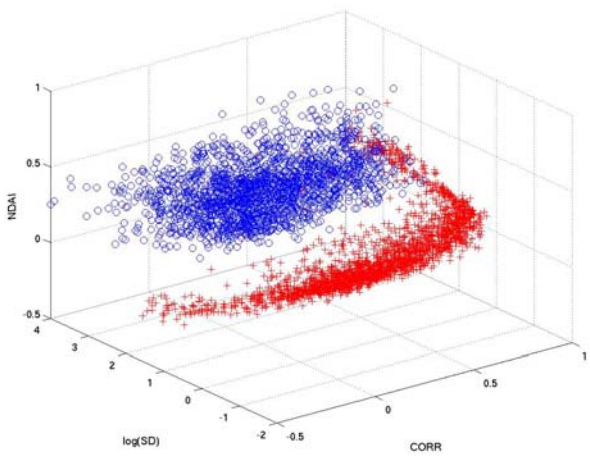


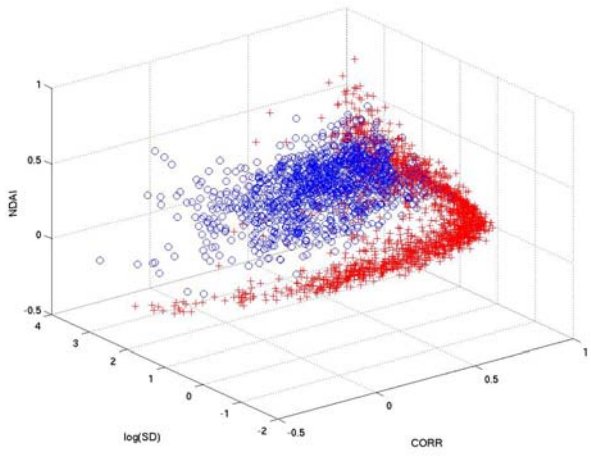
Orbit 13257 Block 20-22 Orbit 13490 Block 20-22 Orbit 13723 Block 20-22

Three Dimensional Plots in Three Features



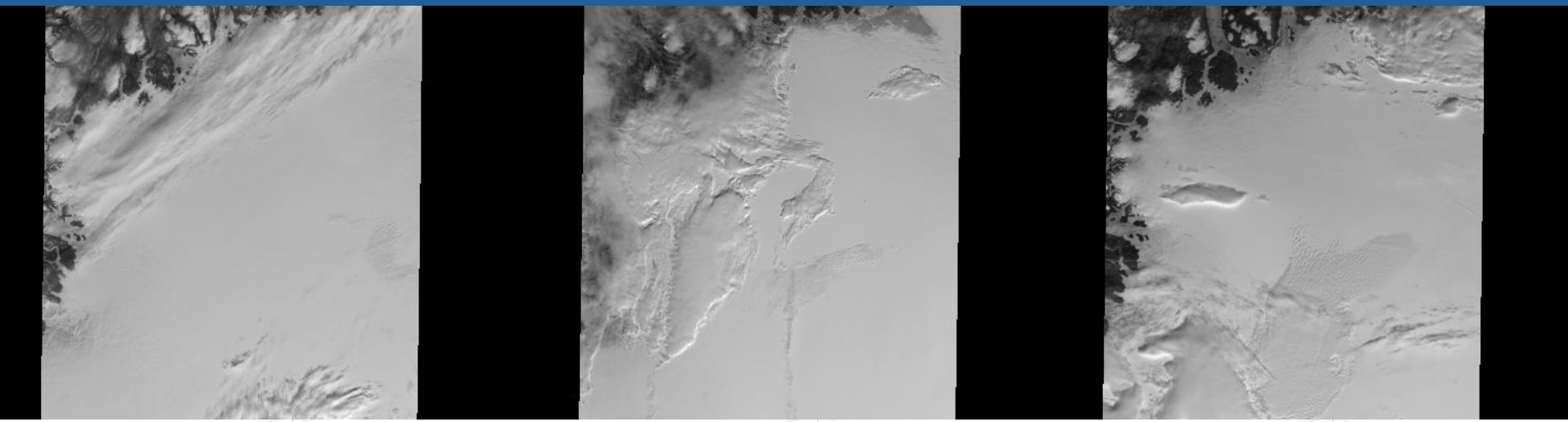


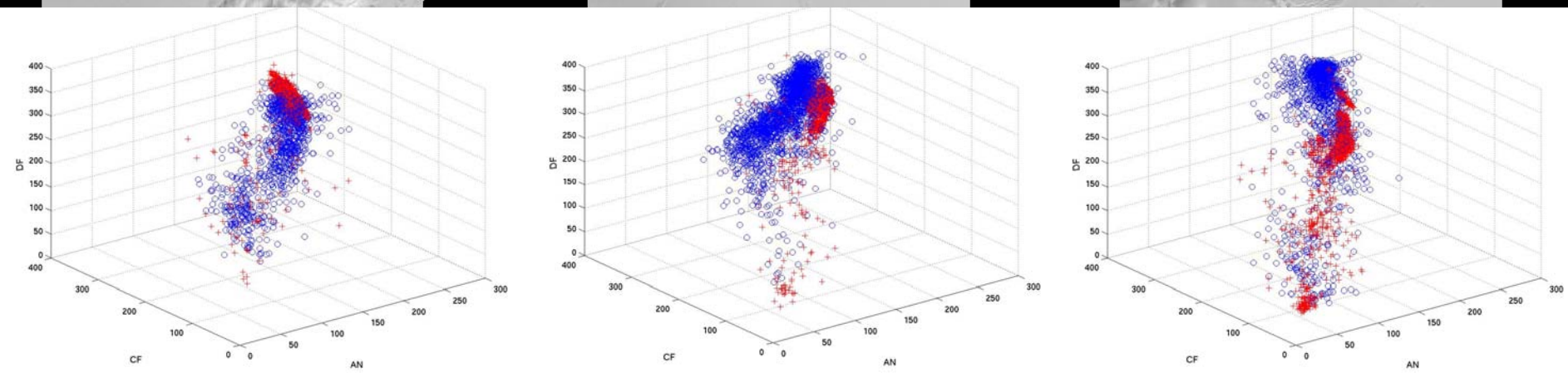




The distributions of “clear” and “cloudy” are “stable” across different orbits

MISR Radiances in Different Angles



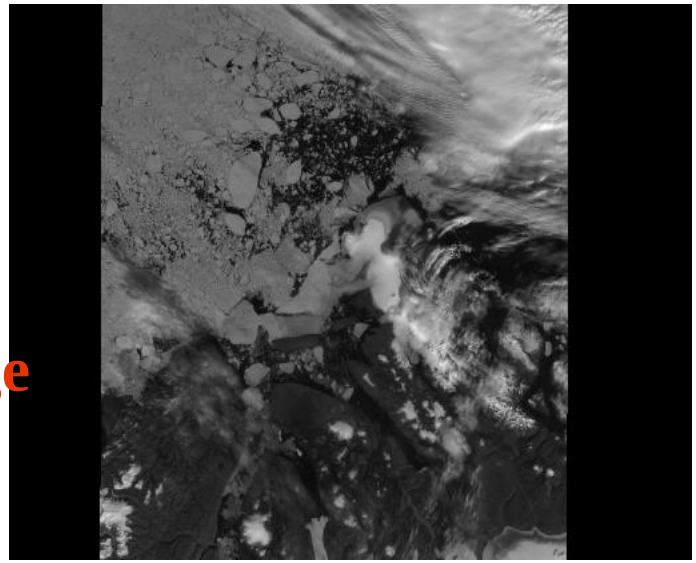


(1). “Cloudy” and “Clear” not separated very well

(2). The distributions of “clear” pixels and “cloudy” ones change across different orbits, so classifier learned from one dataset does not classify another well.

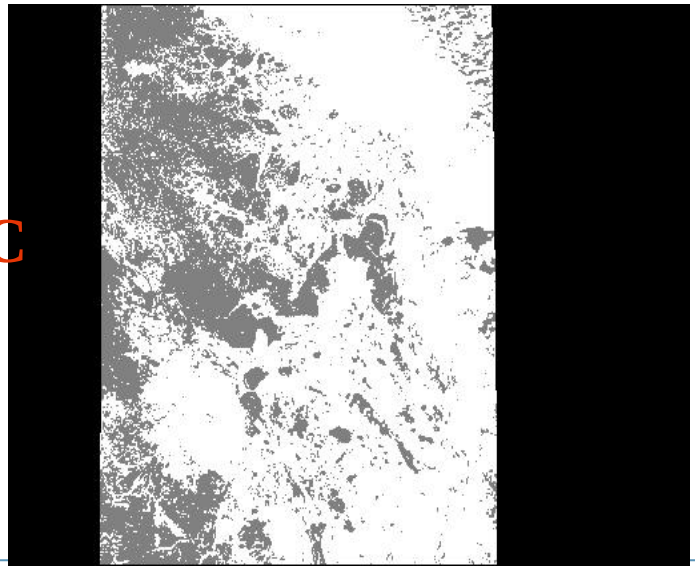
A 70% case…

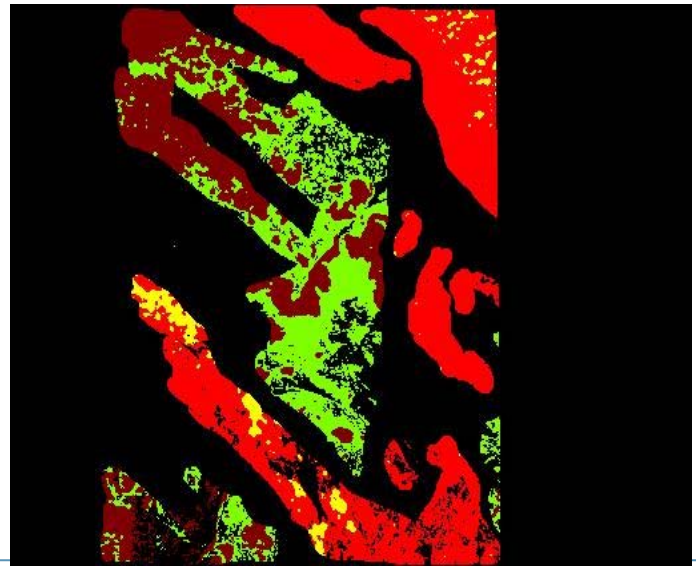
AN image



Expert Label

ELCMC Result (70%)

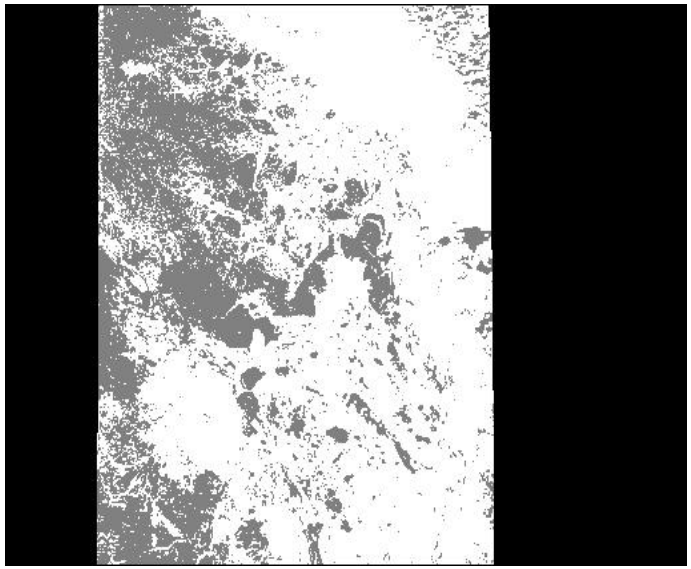


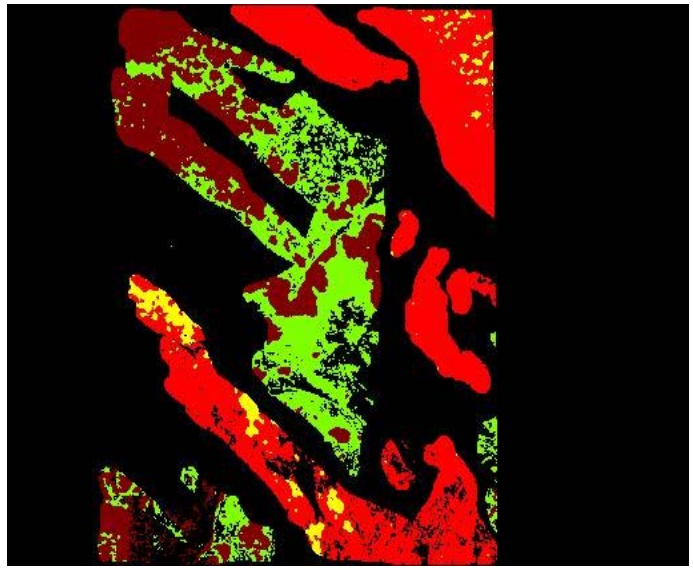


Differ-ence

Problems of ELCMC Identified in Case Studies

ELCMC result





Differ-ence

Two types of mistakes:

(1). ELCMC algorithm classifies clear surface as "cloudy" when the data registration is not good enough (Correlation between angles is reduced.)

(2). It classifies very smooth cloud as "clear" because the SD is too small. This case has not shown much in this example.

Better results with QDA?

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

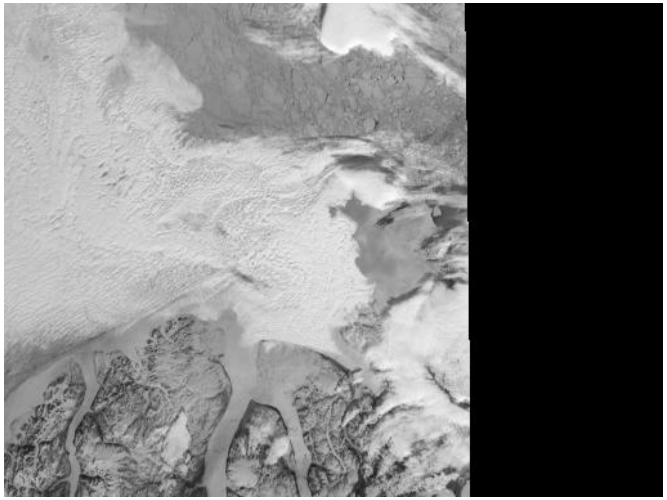
$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}$$

Let π_k be the prior probability of class k (k=1, 2). The posterior distribution for x belonging to class k is then given by

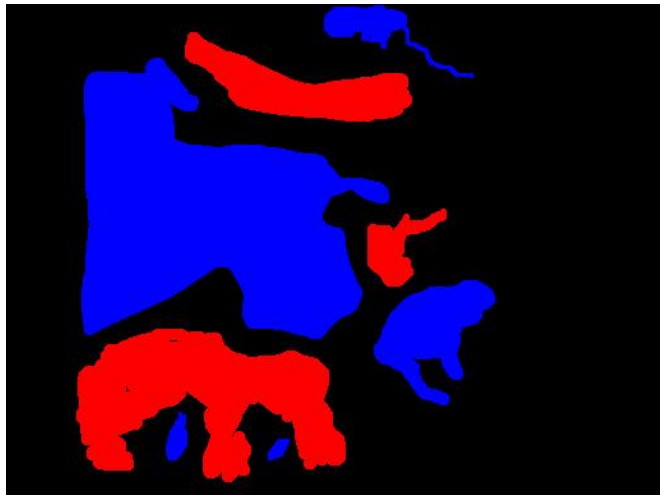
$$P(x \in \text{Class } k|X = x) = \frac{f_k(x)\pi_k}{f_1(x)\pi_1 + f_2(x)\pi_2}$$

QDA classifies x in the class that has the largest posterior probability at x. Using training data, parameters π, μ_k, and Σ_k are estimated by the empirical values.

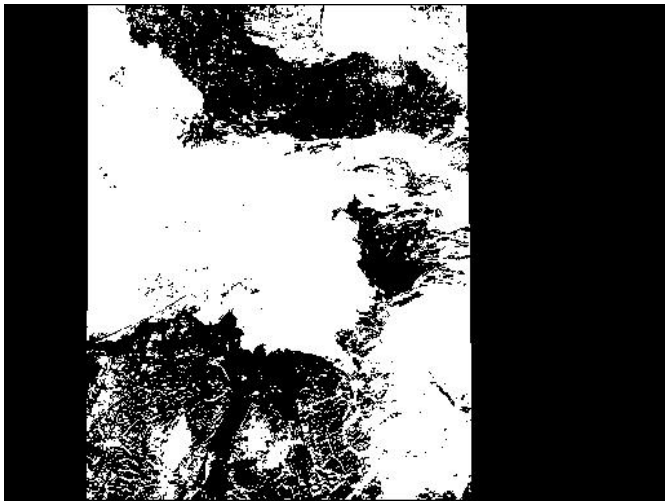
2% improvement in this case by QDA



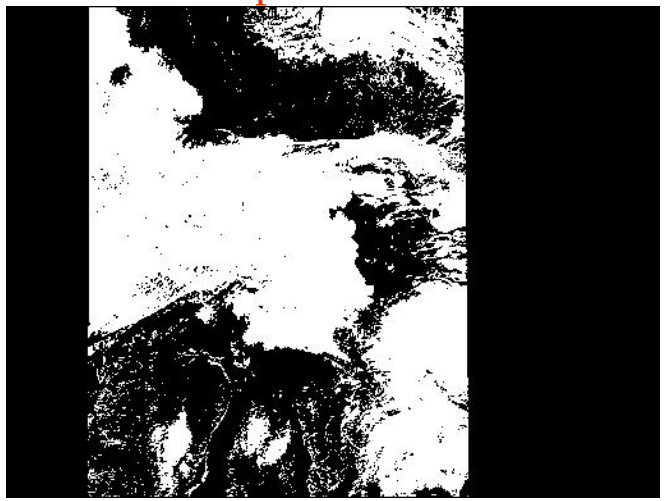
DF red image



Expert Labels



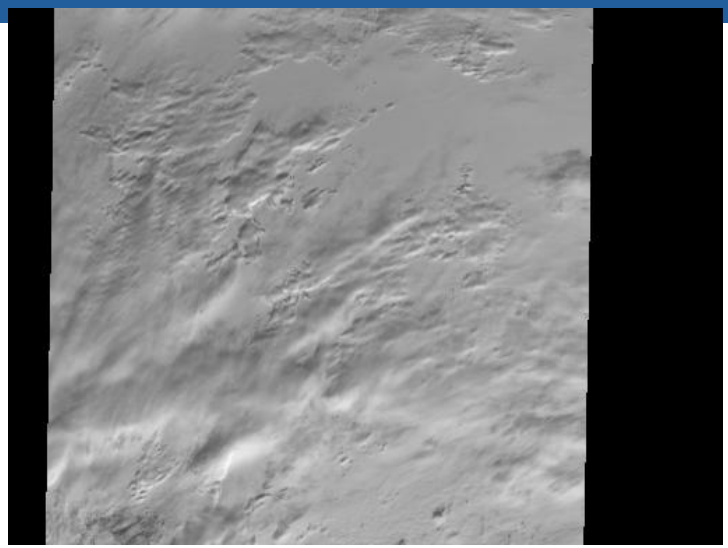
ELCMC Results



QDA Results with ELCMC Labels as Training

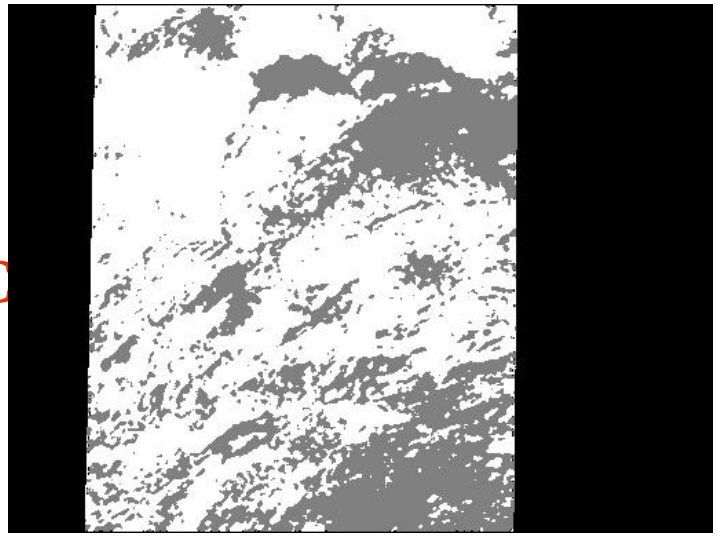
ELCMC-QDA gives soft labels

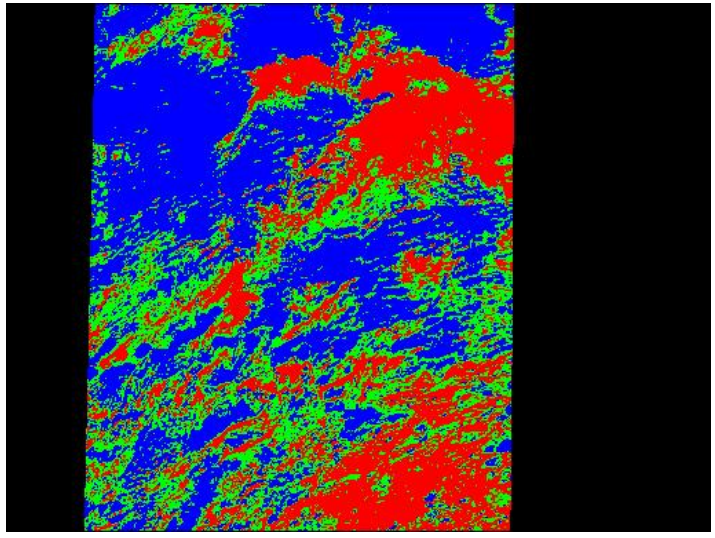
AN image



Expert Label

ELCMC results

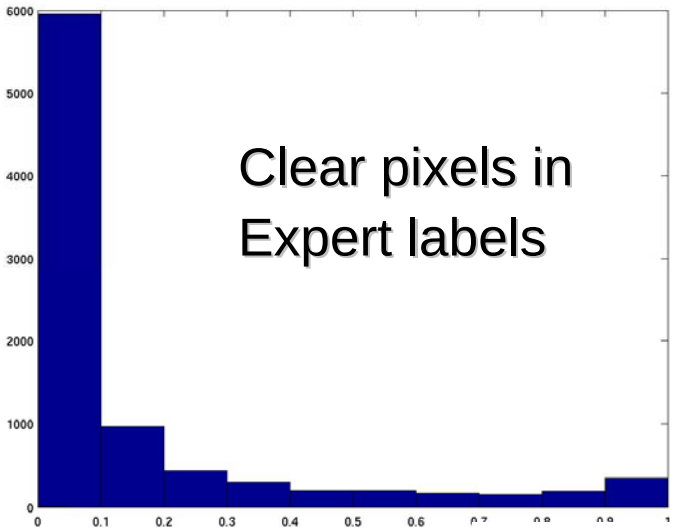


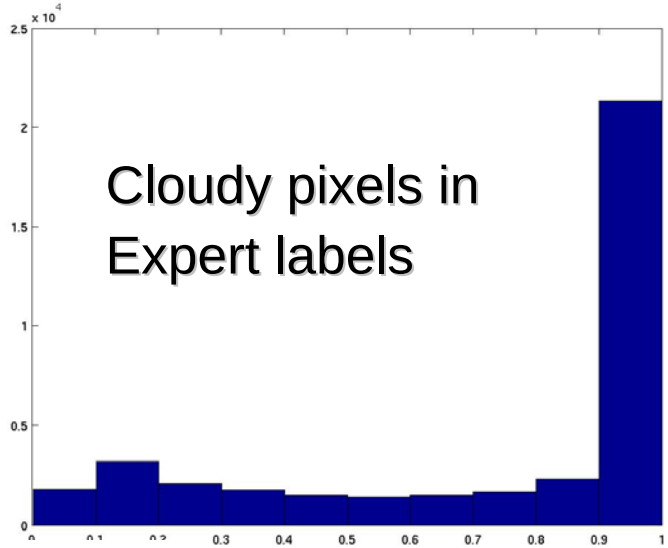


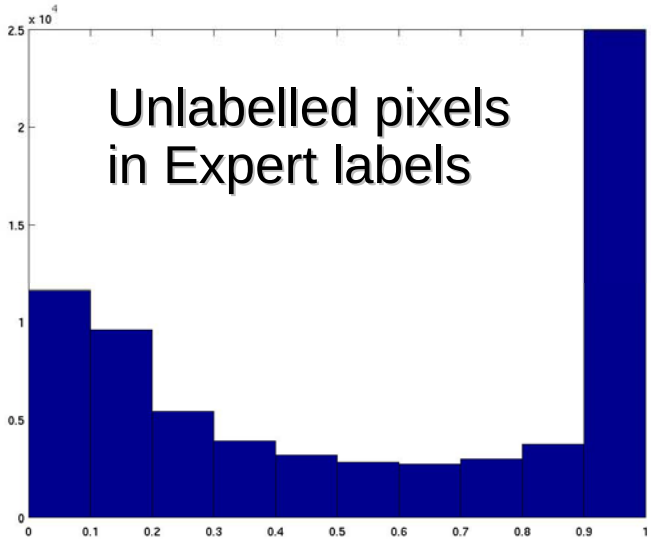
ELCMC -QDA

Green area: probability between 0.3 and 0.7 (those pixels usually fall into the No Expert label areas)

QDA Label Probability Estimates







Summaries of our methods so far

- **ELCMC, ELCMC-QDA online MISR algorithms use three robust features, and are computationally speedy.**
- **ELCMC-QDA provides soft probability labels.**
- **Over 60 test blocks, both ELCMC and ELCMC-QDA give an average accuracy 92% against expert labels.**
- **Over 37 "partly cloudy" blocks, the average accuracy is lowered to around 87%.**
- **ELCMC-SVM gives similar accuracies over some test images, but its computation is too slow.**

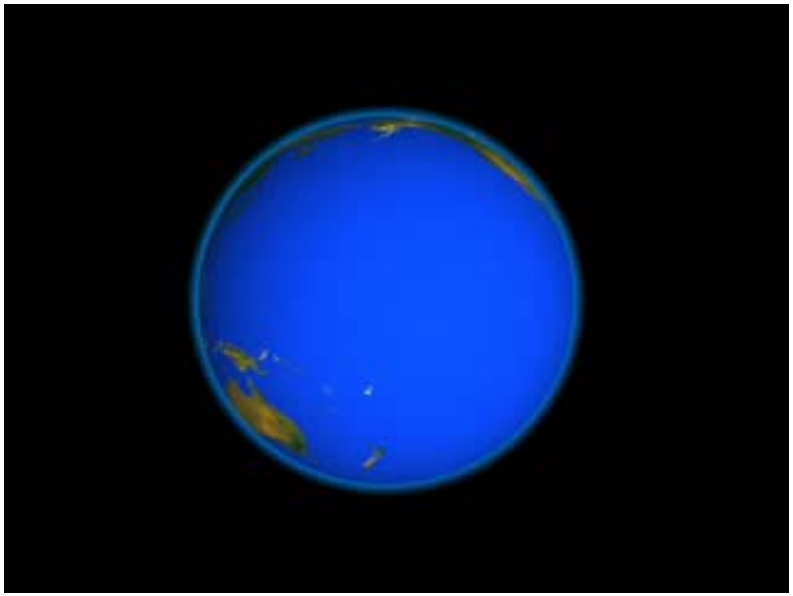
The need for better labels than ELCMC

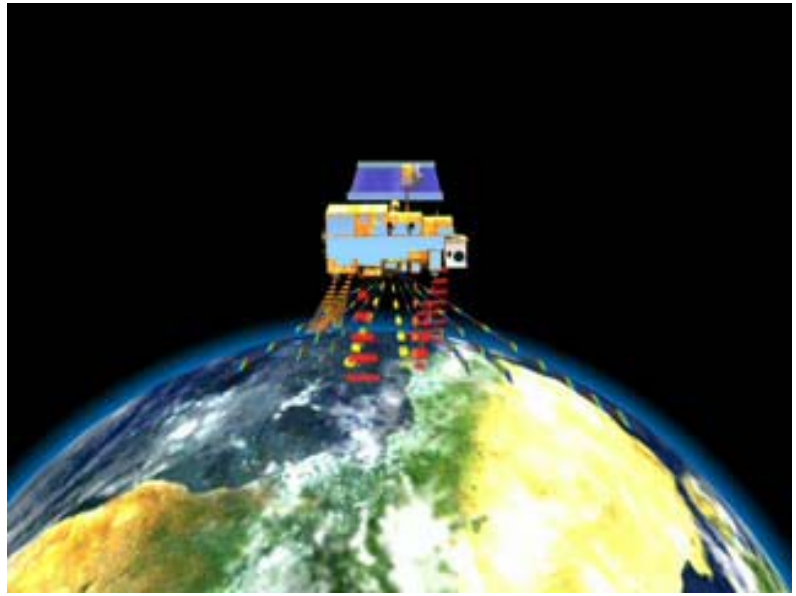
- **IF we use expert labels to train QDA, the accuracy rates are UP quite a few percent (e.g. 5%) from using ELCMC labels.**

  But expert labels are very expensive and impossible to obtain for every block of data even off line, and they could be noisy too.

  Next, we use information from MODIS to get more accurate labels than ELCMC alone.

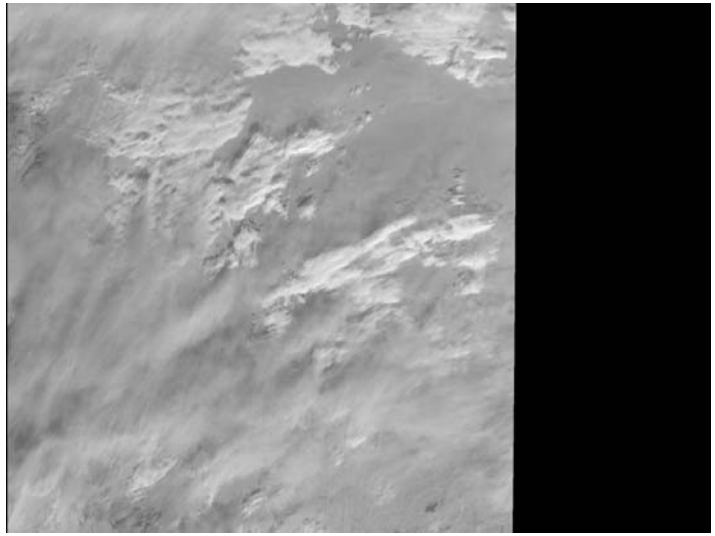
III. Moderate Resolution Imaging Spectroradiometer (MODIS)



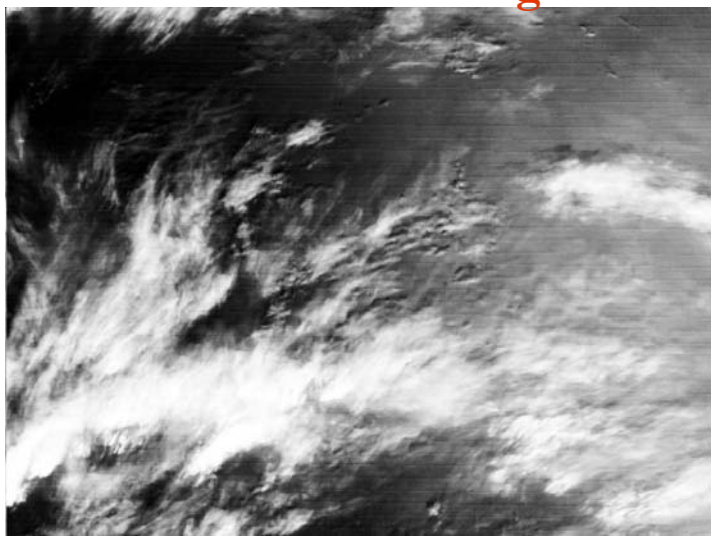


- **MODIS has 36 spectral bands ranging in wavelength from 0.4 μm to 14.4 μm (including visible, near infrared, and infrared).**
- **Spatial Resolution: 250 x 250 m^2 in two bands, 500 x 500 m^2 in five bands, and 1.0 x 1.0 km^2 in other 29 bands.**
- **Combining data from two sensors may provide better results.**

Registering MODIS Data on MISR Grid



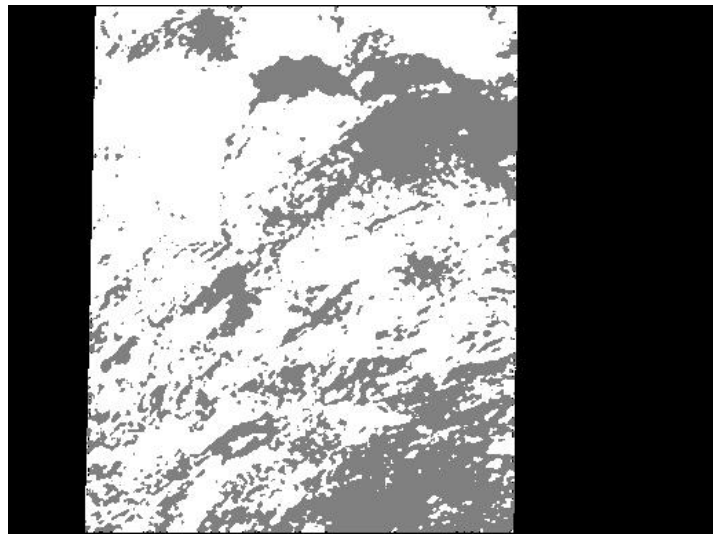
MISR DF red image



MODIS band 26

Data Fusion: ELCMC-MODIS Consensus Pixels

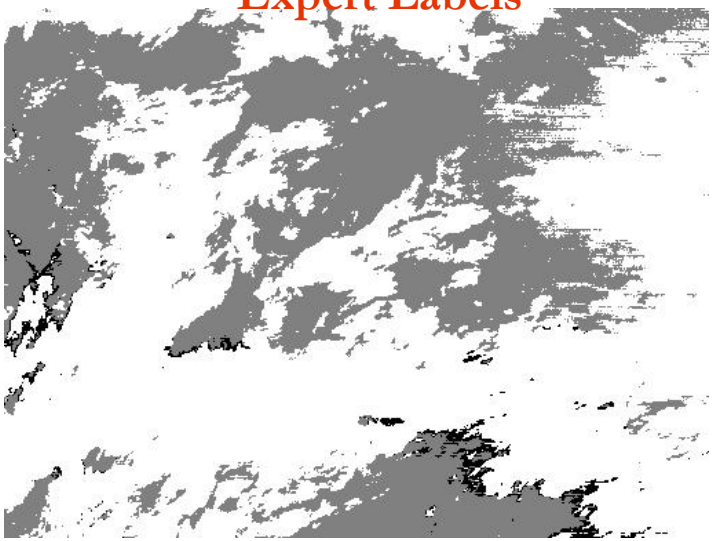
Expert Labels



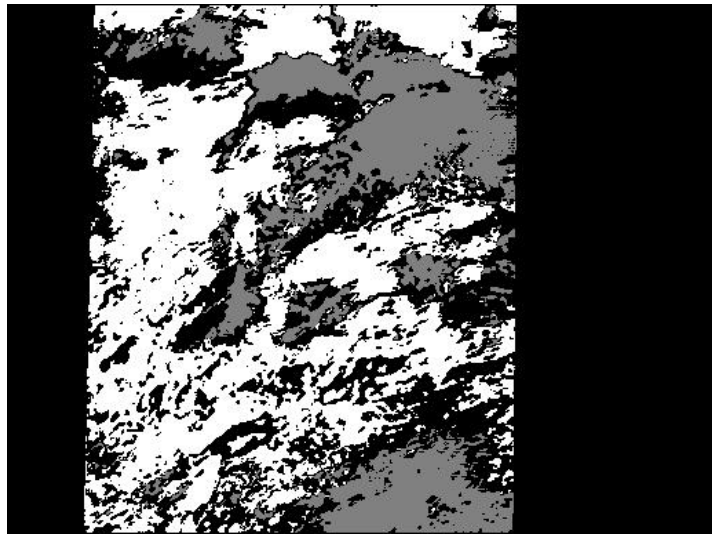
MISR ELCMC cloud mask

91.8%

91.97%



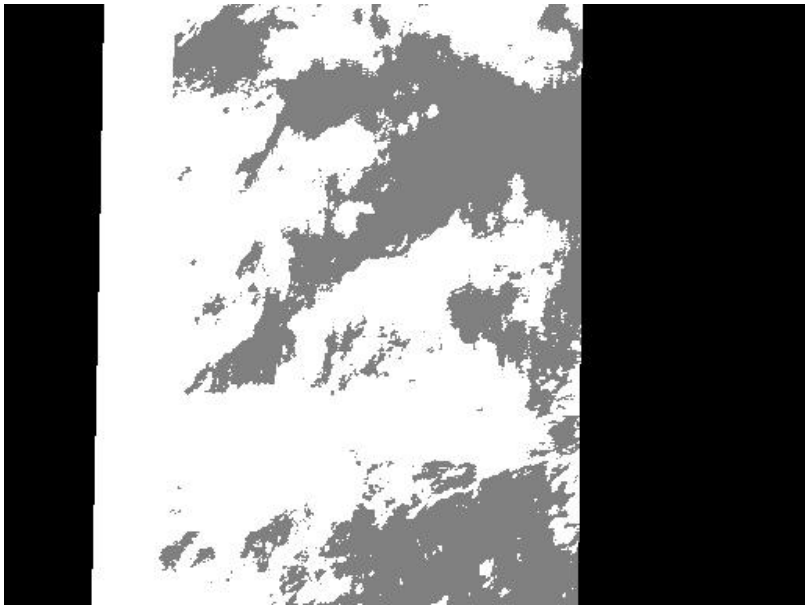
MODIS operational cloud mask



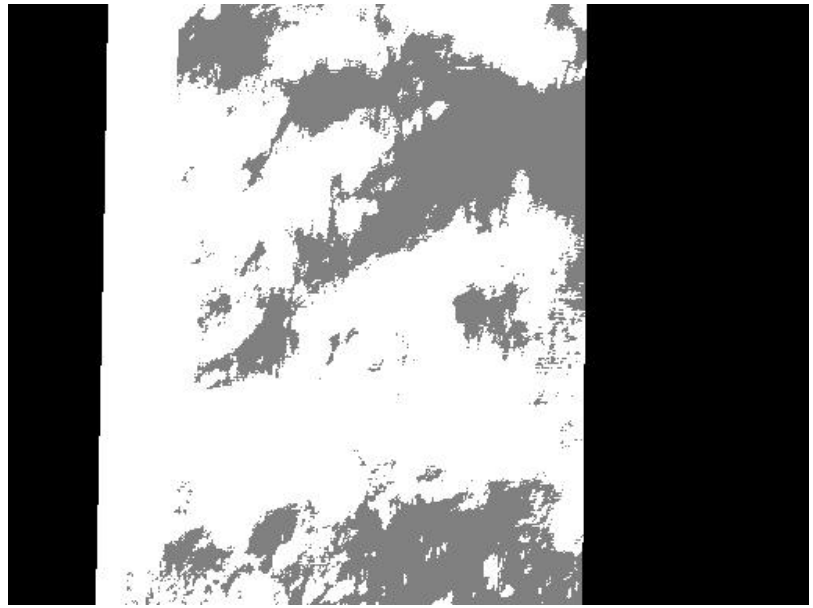
MISR and MODIS consensus pixels

97.75% (72% consensus rate)

Using Consensus pixels and over 37 blocks



QDA using Expert Labels on all features



QDA using consensus Pixels on all features

	QDA (Expert Labels)	QDA (Consensus Pixels)
features (3 MISR+ 5 MODIS)	97% (From 37 blocks of data)	94% (From 37 blocks of data)

Using all features, QDA trained on the consensus pixels (NO EXPERT LABELS USED) provides the best results tested against expert labels.

Reference: Shi et al (2004b, 2004c)

MISR-MODIS Data Fusion Summary

Fusing data from MISR ELCMC and MODIS improves the average accuracy of polar cloud detection to 94% from 87% by ELCMC alone (over 37 blocks).

That is, we basically could reproduce expert performance because experts could differ by a few percents.

Comparing with MISR operational SDCM and ASCM

	ELCMC	SDCM	ASCM (new)
accuracy	94%	80%	83%
Coverage	100%	27%	70%

Current Directions

- **Use ELCMC labels to exclude ground pixels and fractional Gaussian spatial model to estimate cloud heights on remaining pixels**
- **Cloud height validation data from Leica Photogrammetry Suite (stereo visualization from images of two angles)**
- **Produce average cloud coverage over months or years for the polar region, needed input to climate models**

Acknowledgements

- Financial support from NSF, ARO and Miller Institute at U. C. Berkeley

- MISR data were obtained at the courtesy of the NASA Langley Research Center Atmospheric Sciences Data Center

- Helpful discussions with MISR team members L. Di Girolamo, R. Davies, D. Diner, R. Kahn

- D. Mazzoni at JPL for his MISRLEARN Package for labeling the pixels by an expert with MISR and MODIS images

Gaussian Kernel Support Vector Machines

With labels $y_i \in \{-1, 1\}$, SVMs find $f \in RKHS_K$ to minimize

$$\sum_{i=1}^{n} (1 - y_i f(x_i))^+ + \lambda ||f||_K$$

The solution is in the form of

$$\hat{f}(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x)$$

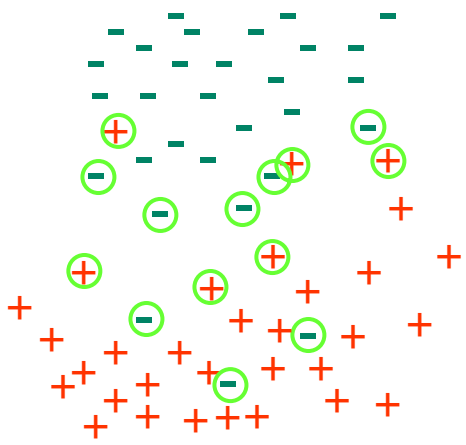
with most of α's being 0, the non-zero α's are called ***support vectors***. The most common used kernel in machine learning literature is the ***Gaussian kernel***.

The computation of training SVM is $O(n^2)$ empirically.

Gaussian Kernel SVMs

The solution is in the form of

$$\hat{f}(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x)$$



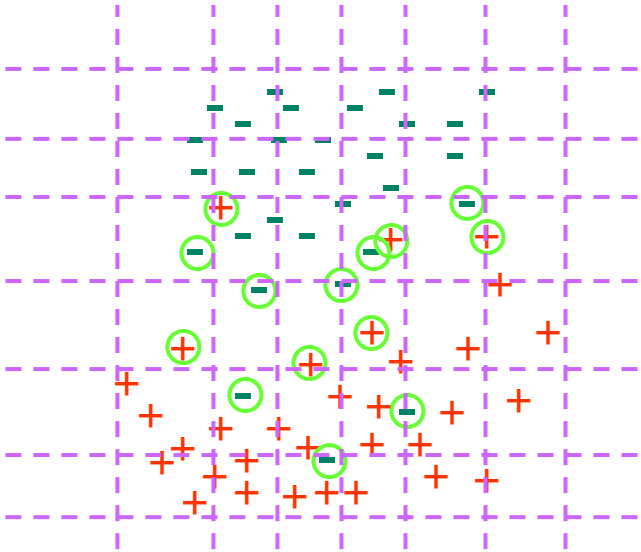
Many α's being 0

The non-zero α's are called ***support vectors***.

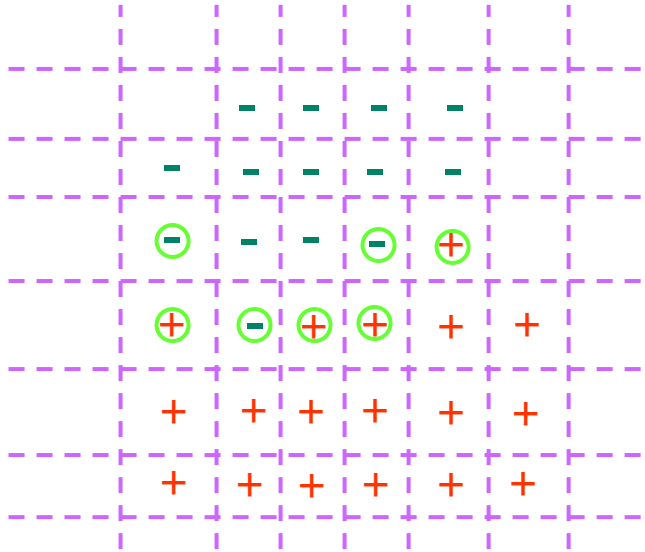
SV's are the misclassified points and the points within the "margin".

IV. Binning for Gaussian Kernel SVMs

- Bin each predictor by its inverse marginal CDF
- Find the data points in each bin
- For non-empty bins, average the predictors and take a majority vote of labels



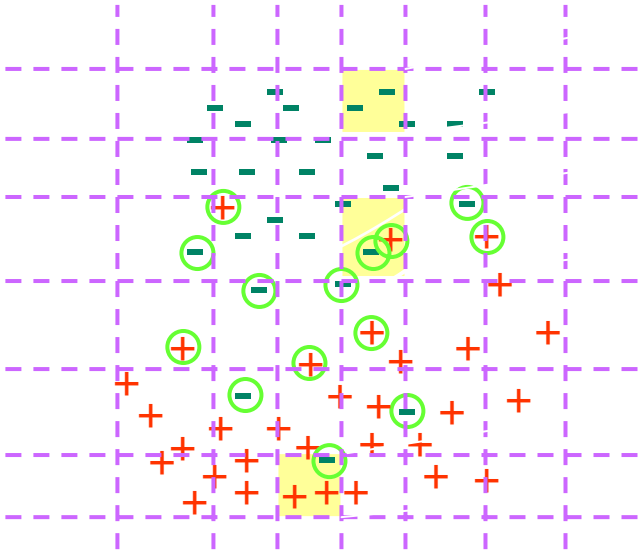
Original Data

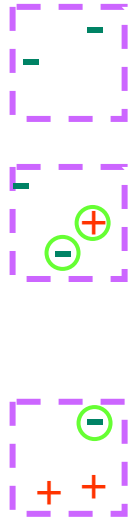


Binned data

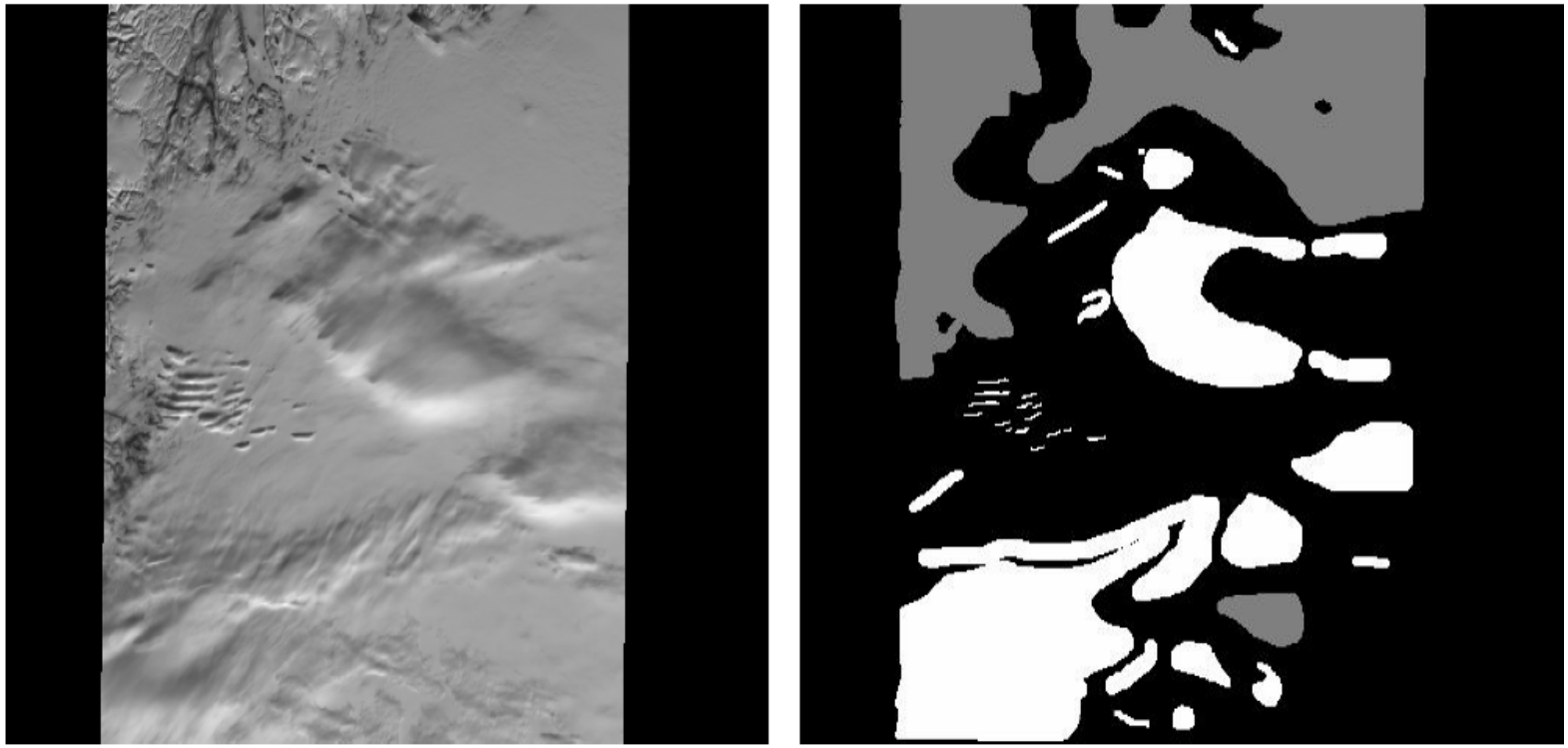
Binning for Gaussian Kernel SVMs (cont)

- Reduces the variance of labels (see illustration below)
- Maintains the marginal distribution of predictors
- Reduces the training sample size and the number of SV's
- Keeps the optimal minimax rate over Sobolev spaces in regression case





One example: Expert label-QDA gives 77% accuracy on a separate test label set



SVMs for Polar Cloud Detection: Expert labels as training data

We train Gaussian Kernel SVMs on three features in 4 setups:

1. SVMs on random sample ~ 966 data points from 54K points

2. Bagged SVMs (repeat training SVM on ~ 966 random sample 21 times; then average the prediction of 21 runs)

3. SVMs on bin centers (split each predictor to 10 bins and use the resulted 996 bin centers and majority vote of labels)

4. SVMs on half of the sample (27K training points)

All parameters of SVM are tuned by cross-validation.

Comparisons

	random sample size 966		SVM on 966	SVM
	SVM	Bagged SVM	bin centers	size 27179
Accuracy	*85.09%	86.07%	86.08%	86.46%
Comp Time (seconds)	81×1.85 = 2.5 minutes	$21 \times 81 \times 1.85$ = 52.11 minutes	$3.87 + 81 \times 1.85$ = 2.56 minutes	81×266.06 = 5.99 hours
# Support Vectors	350	~ 7350	210	8630

λ and w are chosen by CV over a 9×9 grid.

1. SVM on bin centers give the closest rate to the “full” SVM

2. Binning step itself (3.87sec) is fast compared to training SVMs

3. SVM on bin centers have the fewest SV’s, which leads to fast computation in the prediction step.

Summary on Binning

- Binning on Gaussian Kernel Regularization keeps the accuracy and reduce the computation significantly.
- Binning on Gaussian kernel SVMs speeds up both the training and testing speeds.
- The computation of binning is faster than other training sample size reduction methods, such as clustering and bagged SVMs.

Conclusions and Future Work

- **Computationally fast cloud mask generation algorithms: ELCMC, ELCMC-QDA, ELCMC-MODIS-QDA**

- **Excellent accuracy when compared with expert labels**

- **Binned SVMs compuationally feasible for off-line cloud mask generation with much improved accuracy in hard cases.**

Conclusions and Future Work

- **Use ELCMC labels to help estimate cloud heights**

- **Produce average cloud coverage over months or years for the polar region as input to climate models (possible collaboration with LBL)**

Acknowledgements

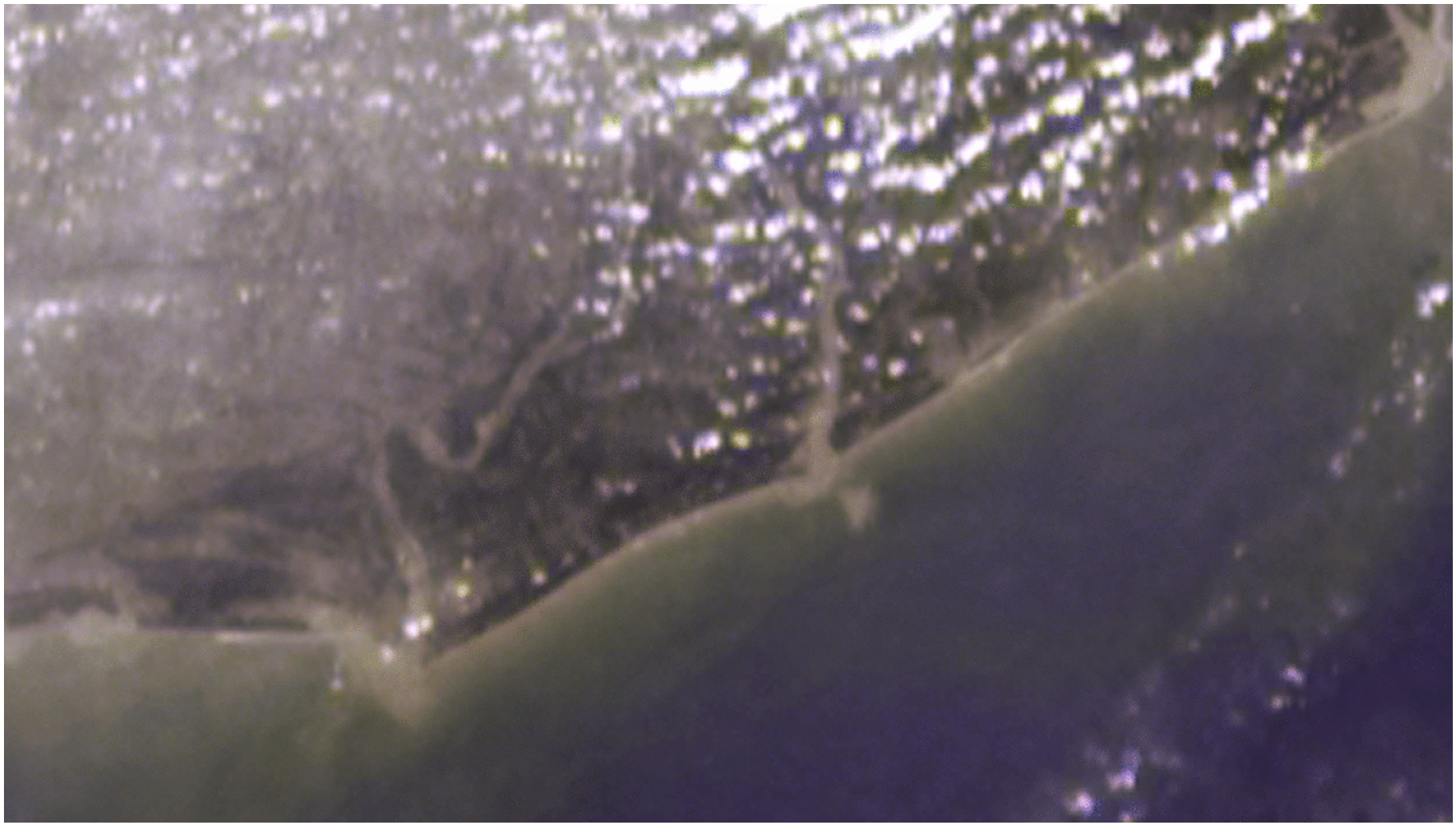
- Financial support from NSF, ARO and Miller Institute at U. C. Berkeley

- MISR data were obtained at the courtesy of the NASA Langley Research Center Atmospheric Sciences Data Center

Statistics is in an exciting era faced with massive amounts of data due to information technology advancements.

When taking into account subject matter, computation, storage and transmission, we are extracting new knowledge from these data to push science forward and at the same time forge new statistics frontiers.

MISR Capturing Asian Tsunami



Summary Accuracies Against Expert Label Test Set

On this image block,

ELCMC: 93.95%
MISR Operational: 46.76%

Using the ELCMC outputs as training data:

QDA: 95.91%
SVM: 96.16%

QDA is much faster than OSU-SVM so more suitable for MISR on-line processing until we can speed up SVM sufficiently.

Binning for PGKR in Regression

Let $x_1, \cdots, x_n$ be equally space in (0, 1] and assume n = mp. Binning the data as:

$$\bar{x}_j = (x_{(j-1)\times p+1} + \cdots + x_{(j-1)\times p+p})/p$$

$$\bar{y}_j = (y_{(j-1)\times p+1} + \cdots + y_{(j-1)\times p+p})/p$$

Use PGKR on the binned data, we get:

$$\hat{y} = G^{(n,m)}(G^{(m)} + \lambda_B I)^{-1} B^{(m,n)} y$$

with $G_{i,j}^{(n,m)} = K(x_i, \bar{x}_j)$ $G_{i,j}^{(m)} = K(\bar{x}_i, \bar{x}_j)$

and $B_{i,j}^{(m,n)} = I\{\lfloor j/p \rfloor = i\}/p$

Asymptotic Properties of Binned PGKR

Theorem: (Shi and Yu, 2005) The binned PGKR achieves the same minimax rate as the unbinned PGKR does in any finite order Sobolev space $\boldsymbol{H^k(Q)}$, when bin size and regularization parameter λ_B are properly chosen.

NOTE: the binned estimator only requires inverting matrix

$$(G^{(m)} + \lambda_B I)$$

The computation complexity is $O(m^3)$. For estimator k-th order Sobolev space $\boldsymbol{H^k(Q)}$, we need $m = O(kn^{1/(2k+1)})$ to achieve the optimal rate.

For example: $O(n)$ for $k = 1$, $O(n^{3/5})$ for $k= 2$, and …

Regularization methods find a function ***f*** that minimizes

$$L(f, \text{data}) + \lambda J(f)$$

--- *L* is an empirical loss.
--- *J(f)* is a penalty functional, which is usually a norm or semi-norm of a Reproducing Kernel Hilbert Space (RKHS).
For example: cubic spline corresponds to

$$J(f) = \int [f''(x)]^2 dx$$