

LDCM Cloud Cover Assessment Update and Overview November, 2010

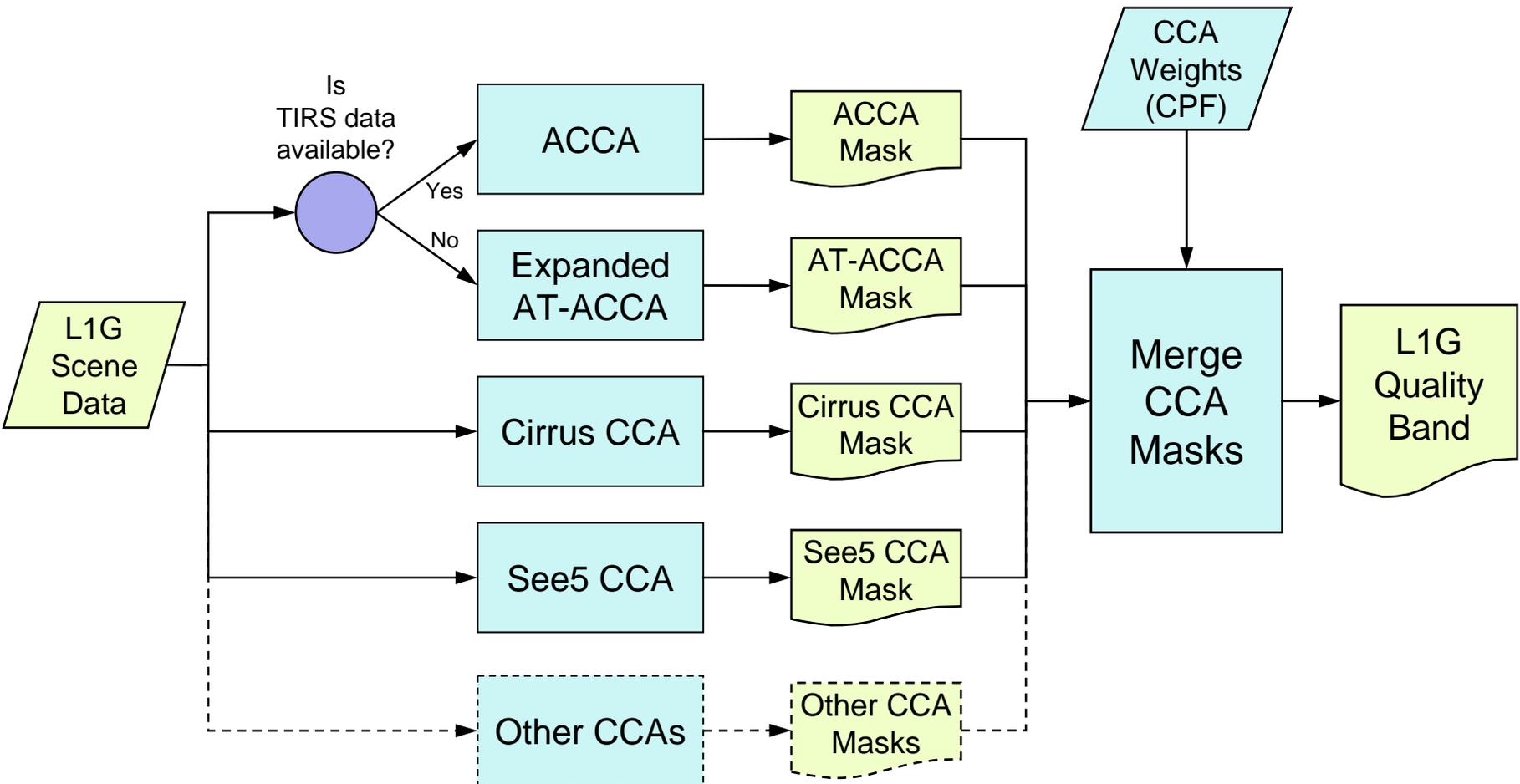
Pat Scaramuzza, SGT
U.S. Geological Survey Center for Earth Resources Observation and Science
Sioux Falls, SD

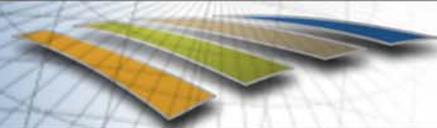
Summary of LDCM CCA Changes in 2010

LDCM

- Minor changes to CCA Control flow.
 - Thermal data switch, no saturation switch.
- Using only non-saturated version of SeaWiFS CCA.
- Quality Band format has been finalized.
- Creation of new ASTER dataset for thermal studies.
 - 15 ASTER scenes as both L1 (TOA reflectance) and L2 (kinetic temperature).
- Addition of Thermal ACCA.
 - Using L7 ACCA on TIRS band 1.
 - Study using ASTER data suggests that surface temperature will offer no advantage over brightness temperature for the ACCA algorithm.
- Addition of Cirrus CCA.
 - Using 1.38 μm band threshold test from MODIS.

CCA Control System Flowchart





CCA Algorithms compared

LDCM

	Pixels Correct	Pixels False	Pixels Ambiguous	Misclassified Clouds	Misclassified Clear
ACCA	79.9%	4.3%	15.8%	7.8%	2.3%
AT-ACCA	89.8%	8.5%	1.7%	12.3%	6.3%
See5-rE	88.5%	7.3%	4.2%	12.1%	4.5%
Cirrus Simple Test (excluding high altitude playas)	91.9%	8.1%	<i>n/a</i>	4.8%	8.7%

Source for these numbers for ACCA, AT-ACCA, and See5-rE is Landsat 7 data from the Irish dataset.

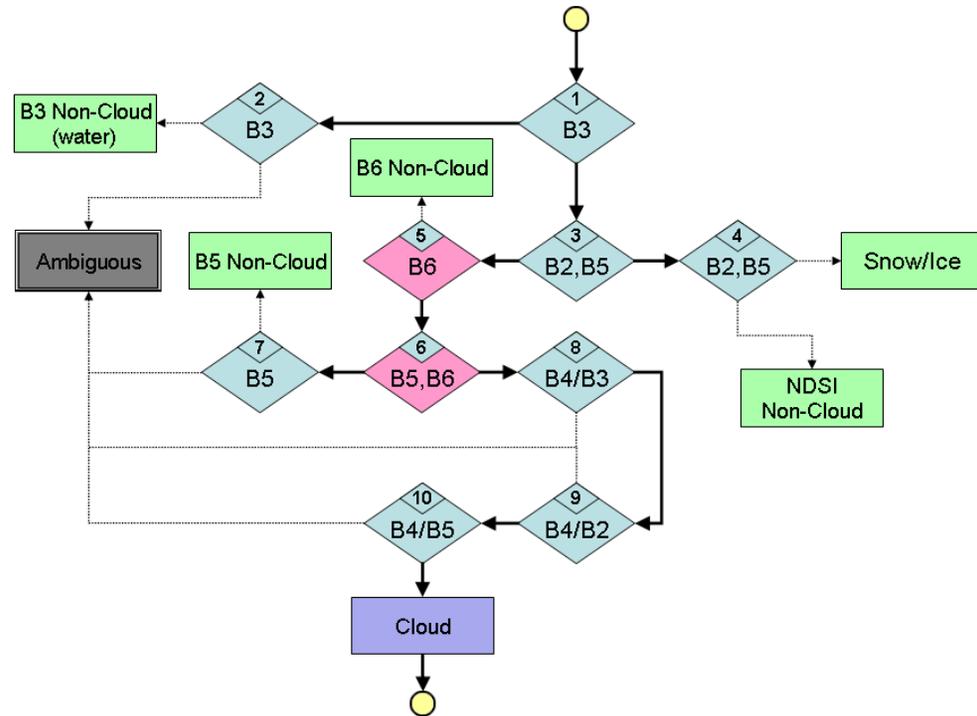
Source for Cirrus test is the Hyperion dataset.

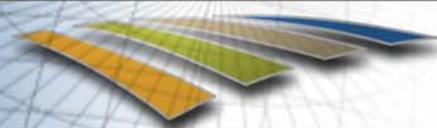
LDCM ACCA

ACCA is a simple decision tree that does crude classification of land cover and clouds, then refines the cloud estimates using statistical data from the thermal band. It was developed by NASA Goddard for the Landsat program.

The LDCM ACCA algorithm is only the classification phase of ACCA, using TIRS thermal band 1.

The snow/ice and water classifications reported by the ACCA algorithm will be provided in the Quality Band of the L1 product.





AT-ACCA

LDCM

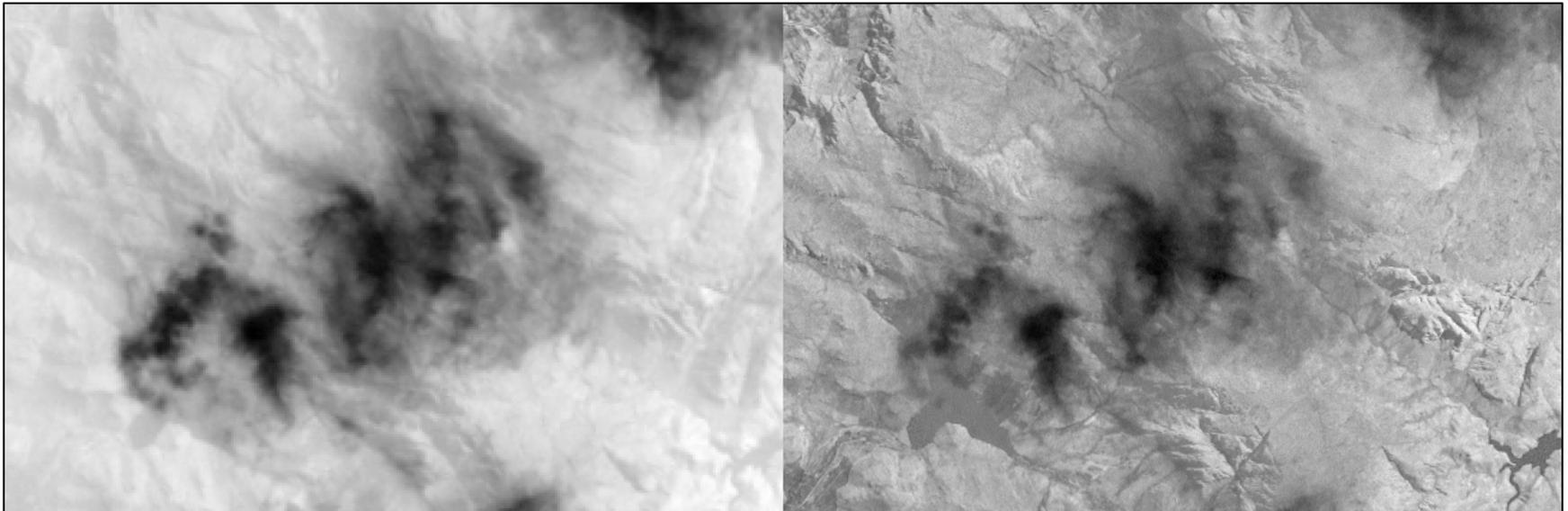
AT-ACCA is a variant of the ACCA algorithm that uses an 'artificial thermal' or AT band. This was developed as a means of running ACCA when thermal data is not available.

The AT band formula was created using the Rulequest Cubist statistical modelling software as a combination of the reflective bands (Landsat bands 1-5 & 7). It was trained on 2 million pixels randomly selected from 103 Landsat 7 scenes, and it has an RMS Average Difference from actual Band 6 brightness temperature of 9.5 K.

To reduce the number of pixels marked as 'ambiguous' by the ACCA algorithm, a disambiguation algorithm was added to AT-ACCA. This is a small McCulloch-Pitts neural network organized in a voting scheme that can re-designate ambiguous pixels as either cloudy or clear terrain.

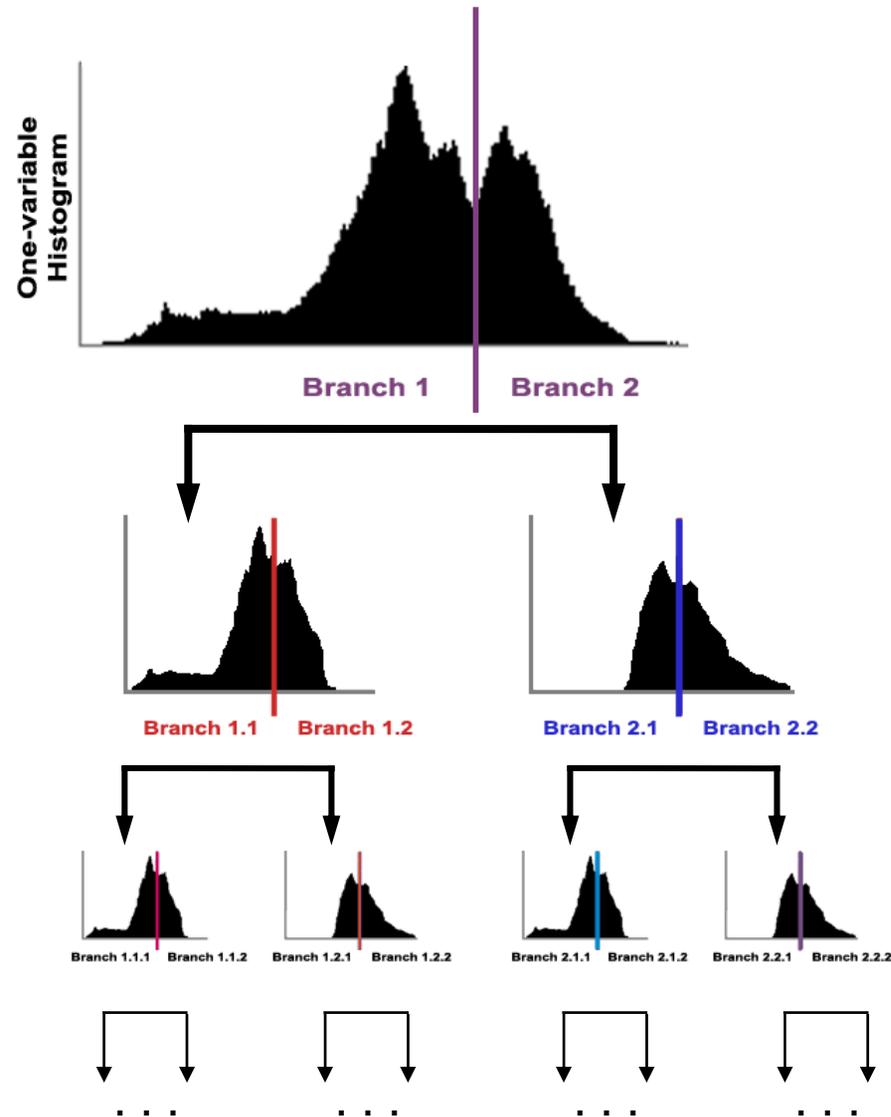
L7 Band 6

AT band



See5 CCA

LDCM



The See5 CCA algorithm was created using the Rulequest C5.0 statistical classifier software.

The Rulequest C5.0 software creates a decision tree by finding the input variable which can best be divided by a threshold test. It selects the threshold value that minimizes the entropy in the dataset.

Once the dataset is divided into subsets, C5.0 then iterates over the subsets. It may select a different variable for each iteration. The user specifies how many iterations and what minimal amount of data each branch must contain. The end result is a decision tree of user-specified size.

The LDCM See5 CCA decision tree has 245 branches, and was trained using 6.4 million pixels randomly selected from 104 Landsat 7 scenes.

LDCM Quality Band Format

LDCM

The Quality Band for LDCM Level 1 products is a 16-bit image of the same dimensions as the L1G scene. Each pixel describes certain processing flags and land/cloud classifications.

The existing CCA algorithms contribute to some of these classifications:

ACCA -- Cloud, Snow/Ice, Water.

AT-ACCA -- Cloud, Snow/Ice, Water.

See5 CCA -- Cloud.

Cirrus CCA -- Cirrus.

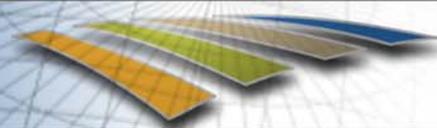
Future algorithms may be developed to improve or expand the classification of these land cover categories.

Bit	Description	Bit	Description
0	Designated Fill	8	Vegetation Confidence
1	<i>Reserved</i>	9	
2	Terrain Occlusion	10	Snow/Ice Confidence
3	<i>Reserved</i>	11	
4	Water Confidence	12	Cirrus Confidence
5		13	
6	<i>Reserved</i>	14	Cloud Confidence
7		15	

Two-bit confidence levels:

- 00 = Not set
- 01 = 0-35% confidence
- 10 = 36-64% confidence
- 11 = 65-100% confidence

Planned end of slide show; all slides past this point are supporting information.

A graphic showing several overlapping satellite images in various colors (orange, green, blue) on a grid background.

Cloud Datasets

- **'Irish' Dataset**

- 212 Landsat 7 L1Gs scenes. (207 used)
- Manual truth masks created by three analysts.
- For general cloud studies.
- Available online.

- **Hyperion Dataset**

- 22 Hyperion scenes, spectrally resampled to resemble Landsat 7 bands 1-5 & 7.
- Manual truth masks created by a single analyst.
- For studying Cirrus algorithms.

- **ASTER Dataset**

- 15 ASTER scenes as both L1 (TOA reflectance) and L2 (kinetic temperature).
- Manual truth masks created by a single analyst.
- For studying brightness temperature / surface temperature differences in ACCA.

Goddard / Rich Irish Test Data Set

LDCM

- **212 Landsat 7 L1Gs scenes**

- Provided by Rich Irish at NASA Goddard.
- Divided by latitude into 9 zones, 20-24 scenes per zone.
- Scenes acquired in 2000-2001.
- When referring to 'Irish' in this presentation, we are referring to this data set. Ireland is not represented.

- **207 scenes with manual cloud masks**

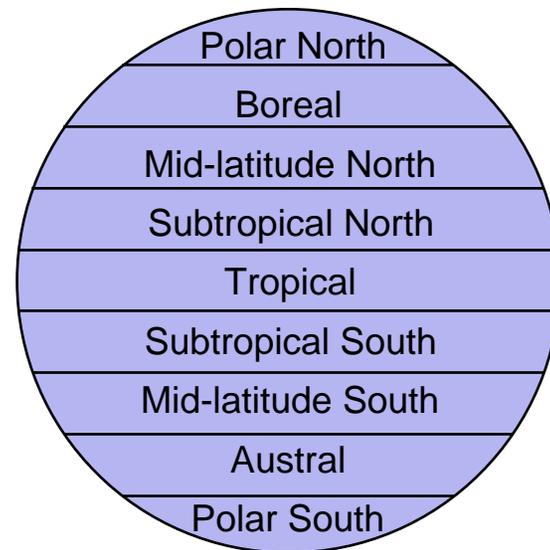
- 5 scenes were excluded due to visible gain change artifacts.
 - One scene had a gain change but no visible artifacts; it was retained.
- Manual masks were created by three operators at EDC.
- Cloud masks should be treated as having ~7% error.
 - Based on 11 masks duplicated by all three operators.

- **104 Training scenes, 103 Validation scenes.**

- Scenes were tagged based on scene content and ACCA performance.
- Tags included heavyclouds, cloudfree, shadows, lowsun, manyambig, falseclouds, and manyprovisionals.
- Division into training and validation sets was performed while attempting to equalize the quantity of each tag in each set, for each latitude zone.

- **Irish data set and the accompanying cloud masks are available for download.**

- Courtesy of EDC. Contact pscaar@usgs.gov for instructions.



See5 Training and Validation Results



Round E algorithm, 1/2009

104 Training scenes

Entire training set:	7,922,160	
Non-saturated pixels:	6,484,262	245 leaves
Saturated pixels:	1,437,898	70 leaves

103 Validation scenes

Total number of validation pixels: 3,955,741,397

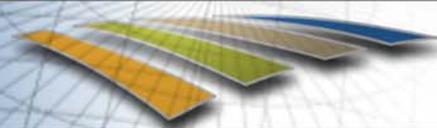
Scene level validation results

Pixel level validation results

% error	# of scenes	% of scenes	cum. %
0 - 5	64	62%	62%
5 - 10	22	21%	83%
10 - 15	9	9%	92%
15 - 20	1	1%	93%
20 - 25	3	3%	96%
25 - 30	0	0%	96%
30 - 35	2	2%	98%
35 - 40	0	0%	98%
40 - 45	0	0%	98%
45 - 50	0	0%	98%
50 - 55	0	0%	98%
55 - 60	1	1%	99%
60 - 65	0	0%	99%
65 - 70	1	1%	100%

Cloud Conf. %	Manual Designation				All Clouds	All Clear
	Clear	Shadow	Thick	Thin		
< 35%	91.8%	93.5%	3.8%	28.7%	12.1%	91.8%
35-65%	3.7%	2.3%	1.8%	11.5%	5.1%	3.7%
> 65%	4.5%	4.2%	94.4%	59.8%	82.8%	4.5%

Total Pixels Correct: 88.5%
 Total Pixels False: 7.3%
 Total Ambiguous: 4.2%



AT equations

Final AT algorithm is the 'ATv6 Cubist r1' formulation:

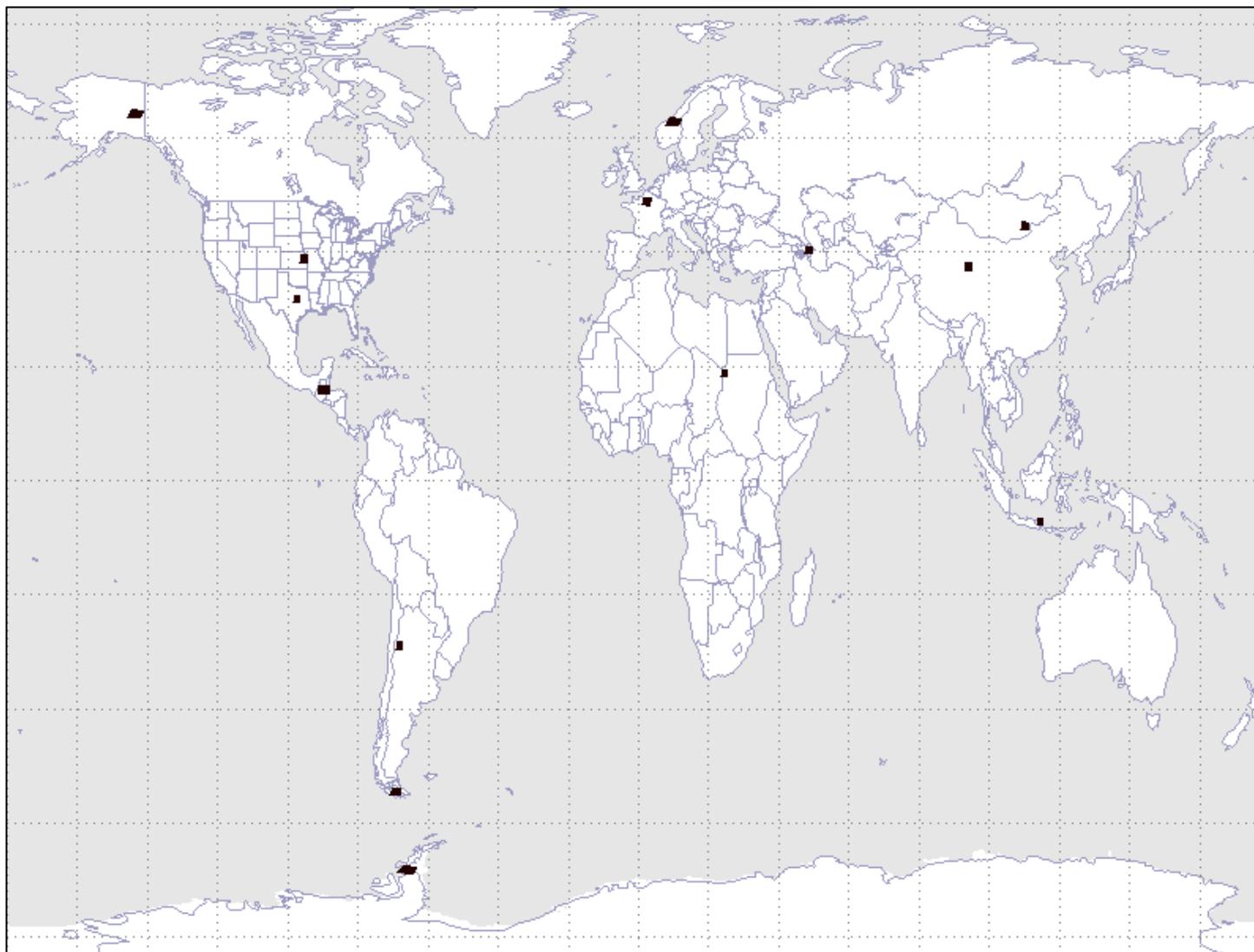
$$\begin{aligned}
 \text{AT} = & \quad - 92.7 \text{ ND}(\text{B3},\text{B5}) & \quad + 261.4 \text{ ND}(\text{B2},\text{B7}) \\
 & - 48.8 \text{ ND}(\text{B2},\text{B5}) & - 17.5 \text{ ND}(\text{B4},\text{B2}) \\
 & - 146.9 \text{ ND}(\text{B1},\text{B7}) & + 58.7 \text{ ND}(\text{B3},\text{B1}) \\
 & - 117 * \text{ND}(\text{B2},\text{B1}) & + 172 * \text{CSA} * \text{B5} \\
 & + 76 * \text{CSA} * \text{B4} & + 151 * \text{CSA} * \text{B3} \\
 & - 951 * \text{CSA} * \text{B2} & + 539 * \text{CSA} * \text{B1} \\
 & + 28 * \text{B7} & - 132 * \text{B5} \\
 & - 106.2 * \text{B4} & - 22.4 * \text{B3} \\
 & + 633.1 * \text{B2} & - 443.6 * \text{B1} \\
 & & + 302.0986
 \end{aligned}$$

where B_x = TOA Reflectance in Landsat Band x.
(Landsat Bands 1-5 & 7 correspond to OLI bands 2-7.)

CSA = $\cos(\text{solar zenith angle})$

$ND(x,y)$ = Normalized difference between x and y.
= $(x-y)/(x+y)$

ASTER Test Scenes

LDCM

15 test scenes, each containing both L1 spectral reflectance and ASTER Kinetic temperature products.

The test set has representatives of partially cloudy scenes from regions covered with desert, snow, vegetation and water.

Scenes were selected from the years 2000-2006, when ASTER had all bands functioning.

Comparison of ACCA Thermal options

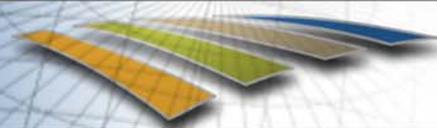
LDCM

	Pixels Correct	Pixels False	Pixels Ambiguous	Misclassified Clouds	Misclassified Clear
Brightness T	55.40%	8.79%	35.81%	11.72%	7.38%
Kinetic T	55.27%	8.67%	36.06%	13.63%	6.28%
Fixed T	49.25%	7.37%	43.38%	11.69%	5.29%

Total number of pixels in the 15 scene ASTER test set: 258,947,180

Conclusions:

- ASTER data does not make a good test set for ACCA.
 - High number of ambiguous pixels suggest that spectral differences in Band 5 are the problem in comparing ASTER to Landsat.
- KT may have a slight (1.1%) advantage in classifying clear terrain, while BT may have a slight (1.9%) advantage in classifying clouds.
- There is no clear advantage to using surface temperature in the ACCA algorithm. Brightness temperature is as good (or slightly better), and is a less complex algorithm.



Hyperion Cirrus results

LDCM

	Cirrus/Not-Cirrus only			Misclassified Cirrus	Misclassified Not-Cirrus
	Pixels Correct	Pixels False	Mid-confidence pixels		
Simple Test Cirrus > 0.02	76.8%	23.2%	<i>n/a</i>	4.4%	26.3%
Simple Test Cirrus > 0.02 (no playas)	91.9%	8.1%	<i>n/a</i>	4.8%	8.7%
See5 hC-6 (parameterized)	74.0%	8.1%	18.0%	44.7%	5.9%

In the absence of high-altitude playas, the Simple (MODIS) threshold test is the best algorithm yet studied, and it easily surpasses the ACCA accuracy goal.