



LDCM Cloud Cover Assessment

July 15th, 2008

Pat Scaramuzza, SGT

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

* Work performed under U.S. Geological Survey contract 03CRCN0001

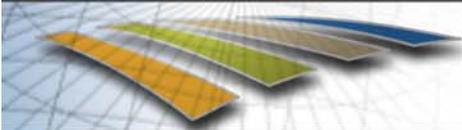
A graphic showing several overlapping satellite images in various colors (yellow, green, blue) on a grid background.

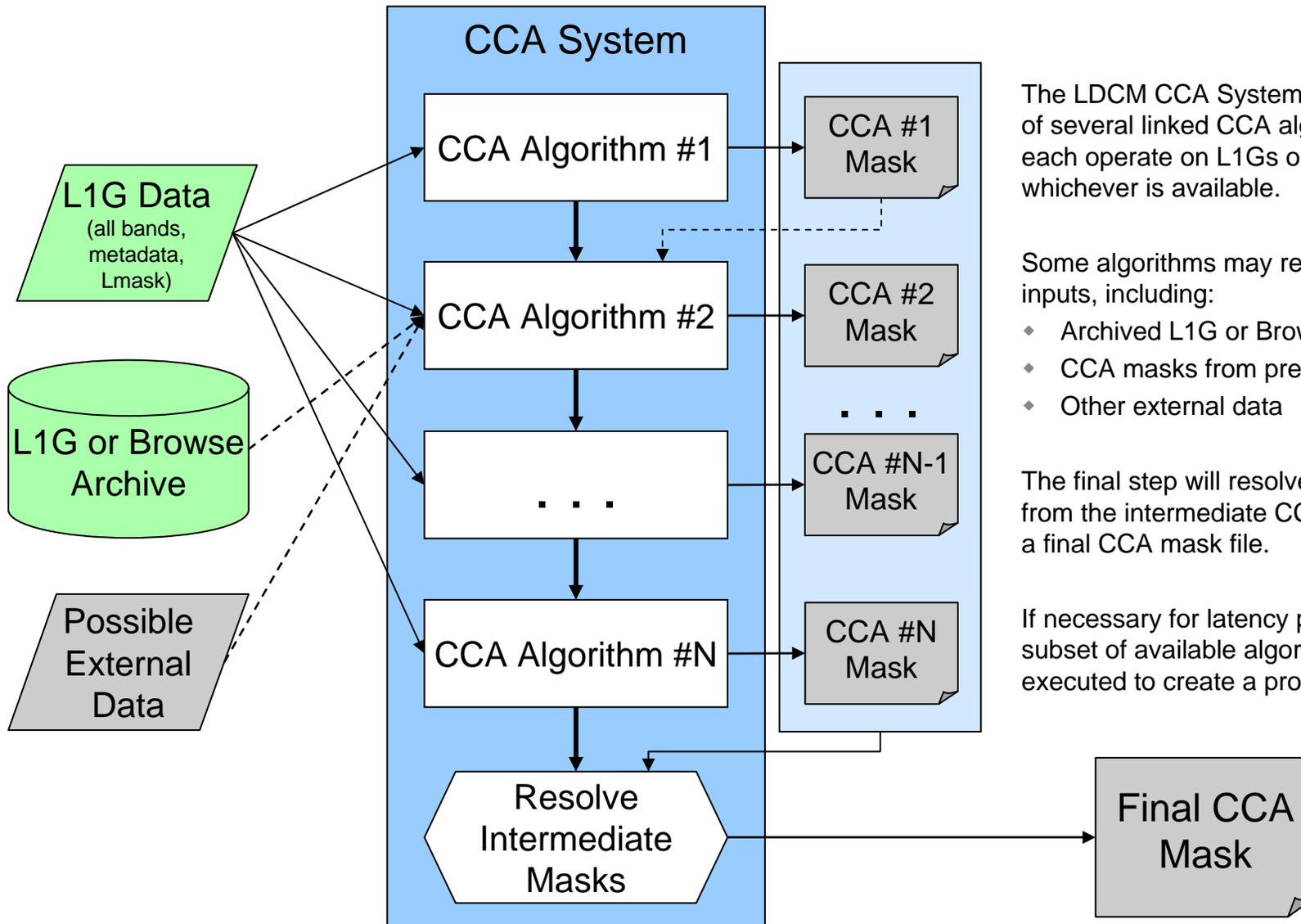
Table of Contents

LDCM

- ◆ Introductions
 - CCA Control System
 - Mask format
 - Irish test data set
- ◆ Algorithms
 - ACCA
 - AT-ACCA
 - See5-ACCA
 - Temporal
 - Cirrus
- ◆ Future work

LDCM CCA System

LDCM



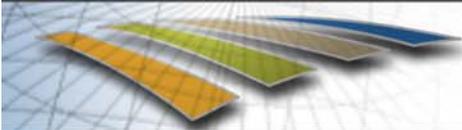
The LDCM CCA System will be comprised of several linked CCA algorithms that will each operate on L1Gs or L1Gt data, whichever is available.

Some algorithms may require additional inputs, including:

- ◆ Archived L1G or Browse imagery
- ◆ CCA masks from previous algorithms
- ◆ Other external data

The final step will resolve all the masks from the intermediate CCA algorithms into a final CCA mask file.

If necessary for latency purposes, a subset of available algorithms can be executed to create a provisional CCA.



LDCM Cloud Mask File

LDCM

Bit	Description
0	Designated Fill
1	Water
2	Water Confidence
3	
4	Snow/Ice
5	Snow/Ice Confidence
6	
7	Cirrus
8	Cirrus Confidence
9	
10	Cloud
11	Cloud Confidence
12	
13	Artifact masks
14	
15	

The proposed format of the final CCA mask file is 16 bit, including 4 classes and an artifact mask.

Two-bit confidence levels:

- 00 = None or Unset.
- 01 = 0-33% confidence
- 10 = 34-66% confidence
- 11 = 67-100% confidence

The artifact mask may contain a similar 'contribution level', marking how much of a given pixel's radiance is from a resampled artifact.

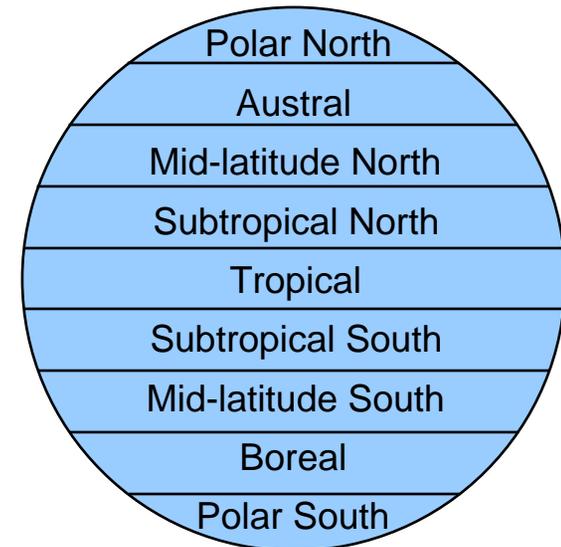
This file format is preliminary and likely to change.

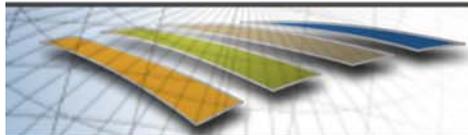
Intermediate CCA masks may use a different format, but will be standardized.

Goddard / Rich Irish Test Data Set

LDCM

- ◆ 212 Landsat 7 L1Gs scenes
 - Scenes acquired in 2000-2001.
 - Divided by latitude into 9 zones.
 - 24 scenes per zone
 - Except Polar South, which has 20.
 - 4 scenes with gain changes
 - Used in most analyses.
 - Discarded for some analyses due to artifacts.
 - Not used as training data.
 - Will not be manually assessed.
 - One stack of 4 scenes with the same path/row in each zone, for temporal testing.
 - When referring to 'Irish' in this presentation, we are referring to the data set. Ireland is not represented.





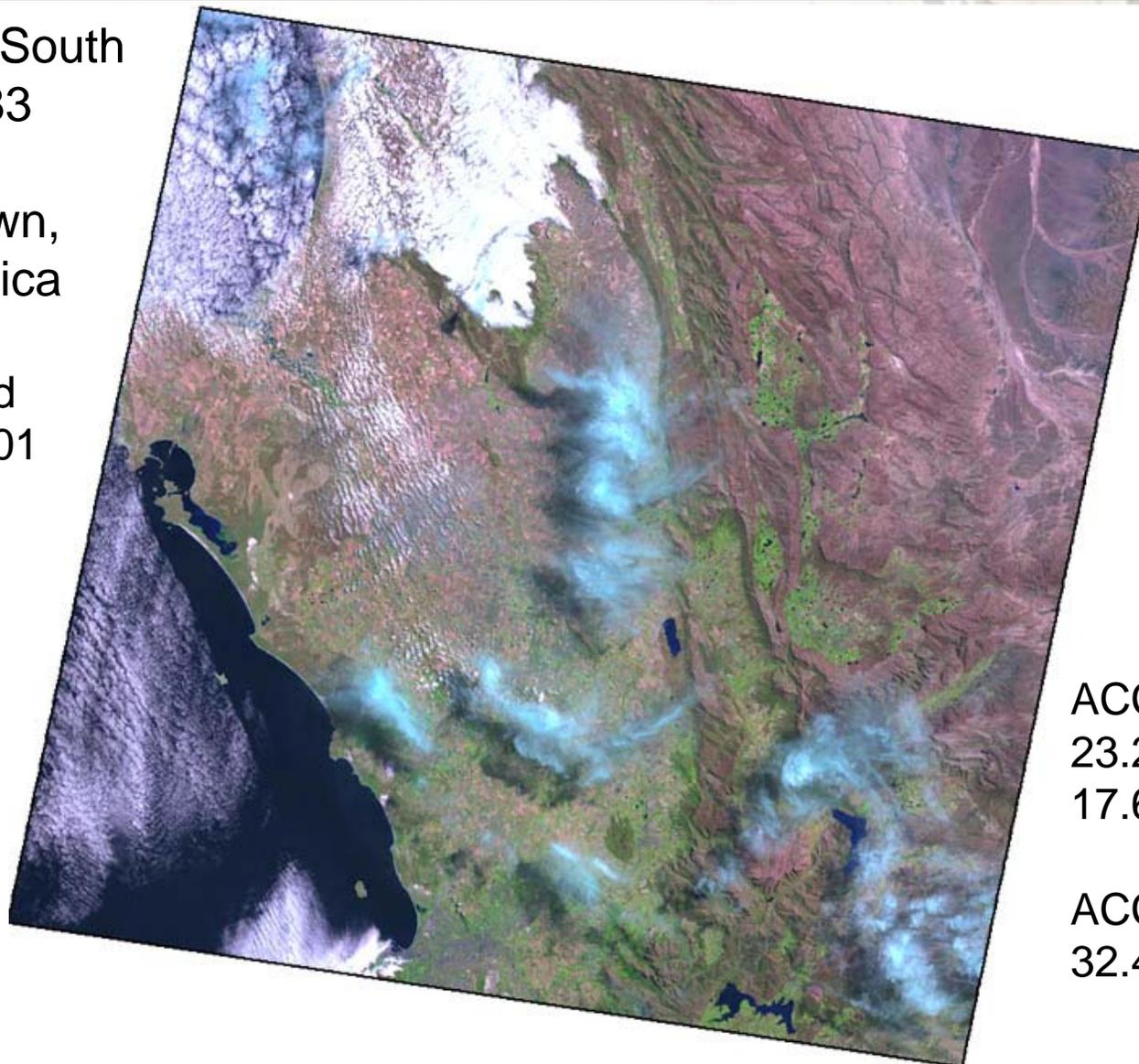
Irish Set -- Example Scenes

LDCM

Mid-latitude South
p175_r83

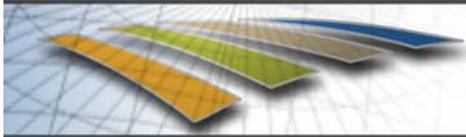
Cape Town,
South Africa

Acquired
10/22/2001



ACCA phase1
23.2% clouds
17.6% ambiguous

ACCA full score
32.4% clouds

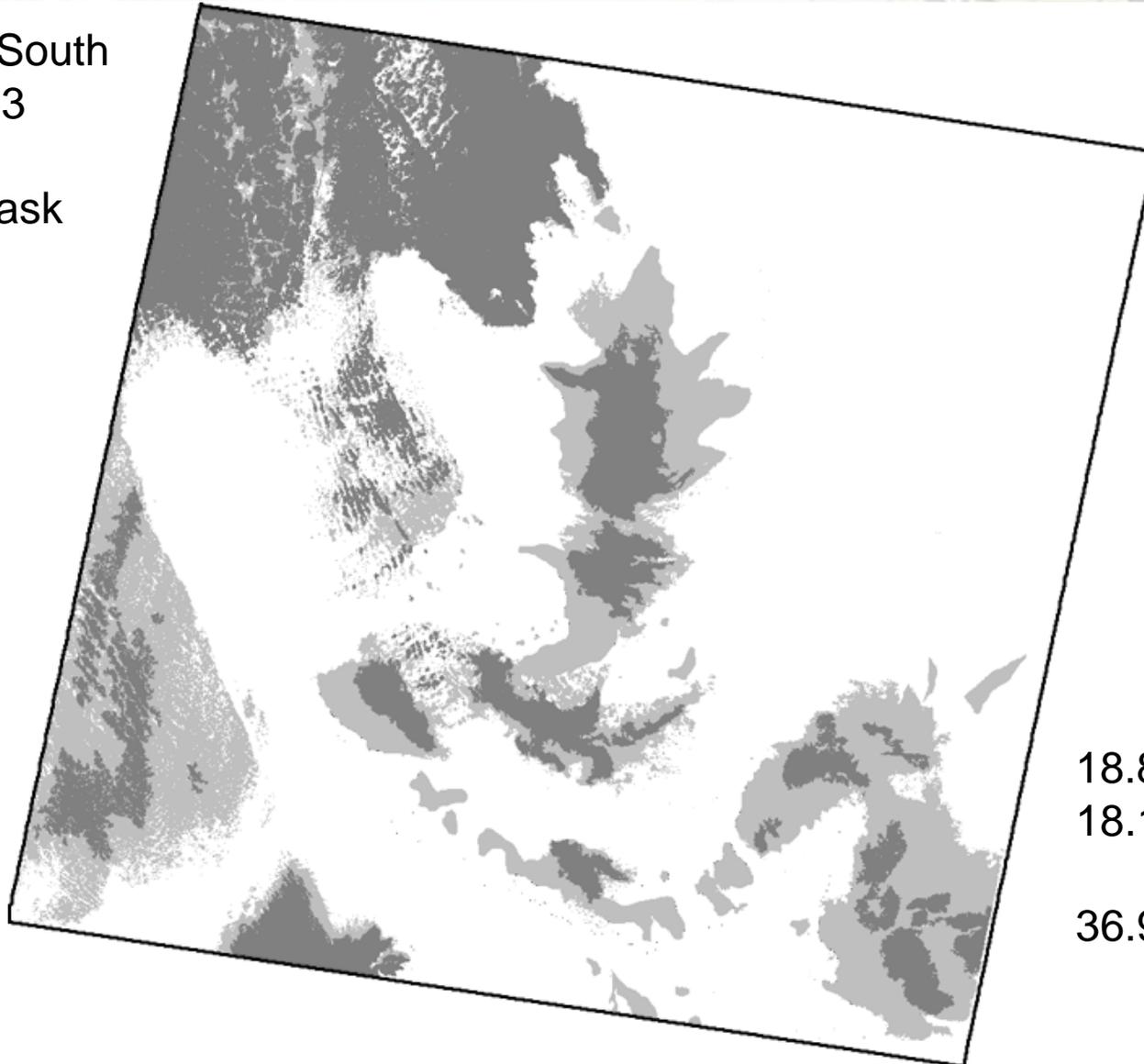


Irish Set -- Example Scenes

LDCM

Mid-latitude South
p175_r83

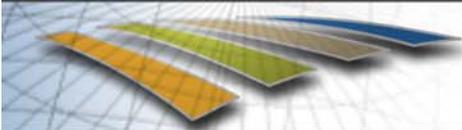
Manual Mask



Thick Cloud
Thin Cloud
Clear

18.8% Thick
18.1% Thin

36.9% Total clouds

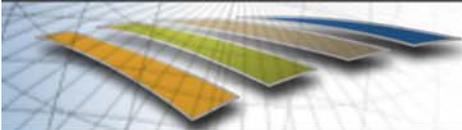
 AT-ACCA

Almost identical to the phase 1 ACCA algorithm.

- Instead of Band 6, AT-ACCA uses an AT ('Artificial thermal') band.
- One pure thermal test removed.
 - Created false negatives with original AT bands.
 - May be reintroduced if final AT band is accurate enough.
- Later phases of ACCA are not implemented.
 - They are mostly scene-wide statistical tests, and do not create per-pixel masks.

Our Goals:

- **Algorithm should be as sensor-neutral as possible.**
 - Works with TOA Reflectance.
 - Trained and tested on L7, ALI, Hyperion, and OLI simulated data.
 - Should work on any sensor with Landsat-like bands.
- **Empirical approach.**
 - We have large quantities of data with which to generate empirical models.
- **ACCA-like accuracy.**
 - Our goal is not to improve on L7 ACCA, but to have a CCA method that works as well without a thermal band.
 - If we improve on ACCA, it will be with the modular boosting of the CCA system.
- **For CCA use only.**
 - AT band is not intended as a replacement for Band 6 for any other purposes!

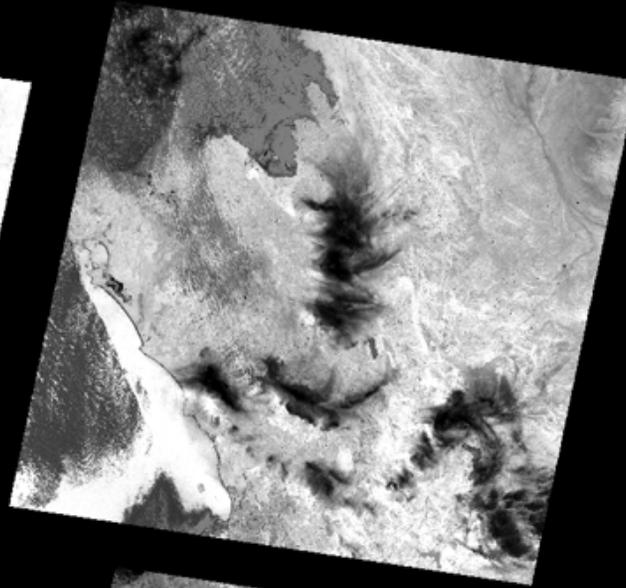
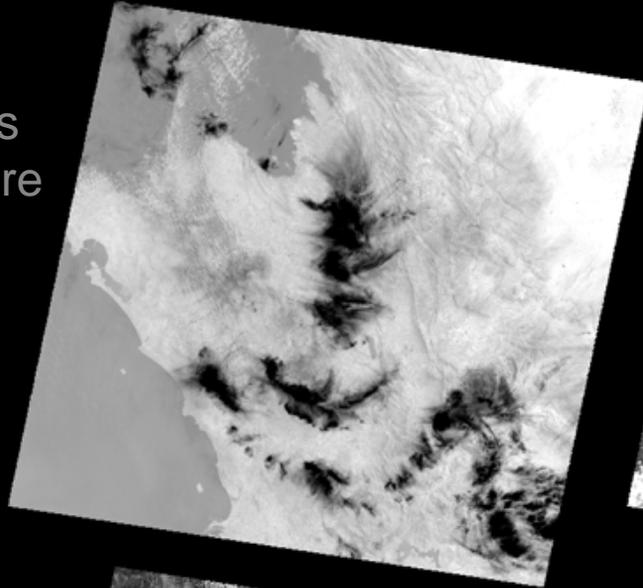
 AT (Artificial Thermal) Bands

- ◆ ATv2
 - Simple equation using only bands 2 and 5.
 - Arrived at via genetic approximation over 90 random IAS scenes.
 - Tested on Irish data set.
 - Average difference from Band 6 brightness temperature = **16.2 ° K.**
 - Well-studied, but not likely to be the final AT algorithm.
- ◆ ATv3
 - Form of equation is similar to ATv2, but with coefficients optimized via Irish data set.
 - Av. Diff = **13.8 ° K.**
 - Not well-studied (Cubist followed soon after).
- ◆ Cubist
 - Model derived using Cubist software. (Cubist will not be necessary in the operational system.)
 - Training data is 1.1 million points selected at random from all 212 Irish scenes (~5,200 points per scene.)
 - 100 Rule version is slow and discontinuous.
 - Av. Diff = **5.4 ° K.**
 - 1 Rule version is faster and continuous.
 - Av. Diff = **8.8 ° K.**

Example of AT bands (full)

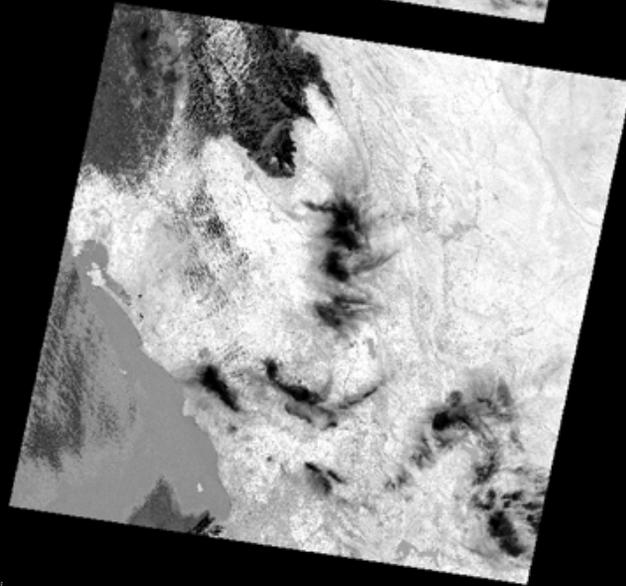
LDCM

Band 6
Brightness
Temperature

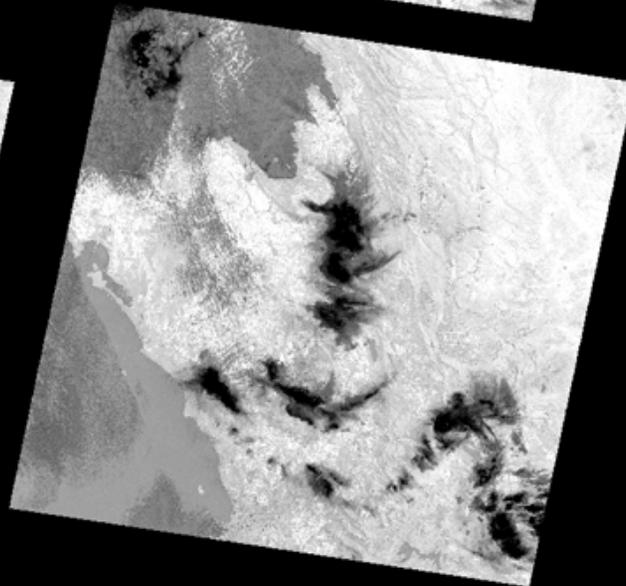


ATv2

Cubist
1 rule

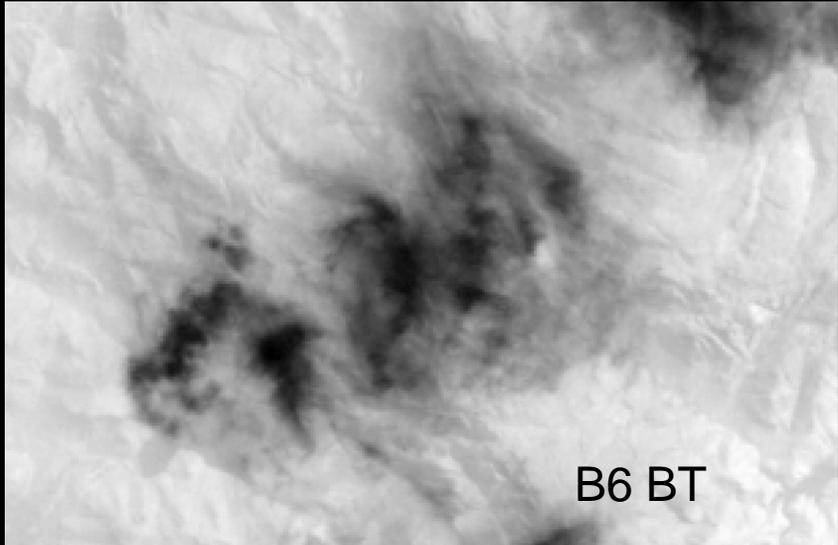


Cubist
100 rules

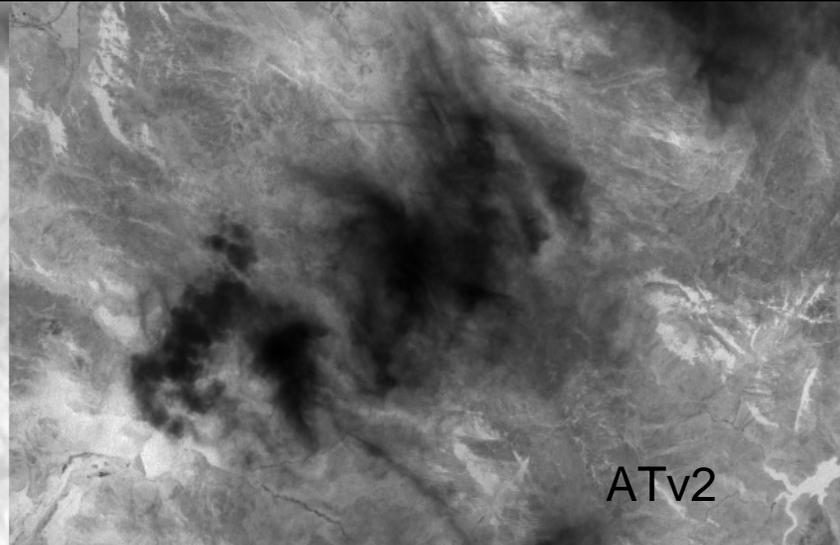


Example of AT bands (zoom)

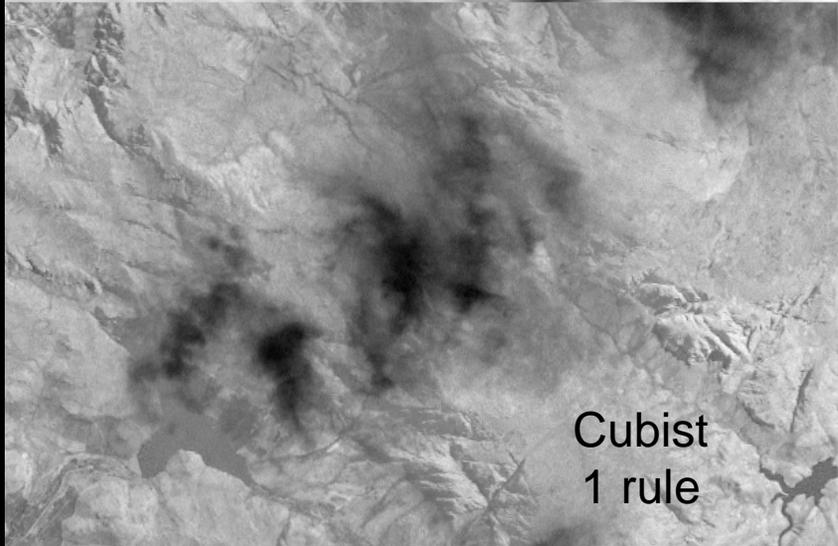
LDCM



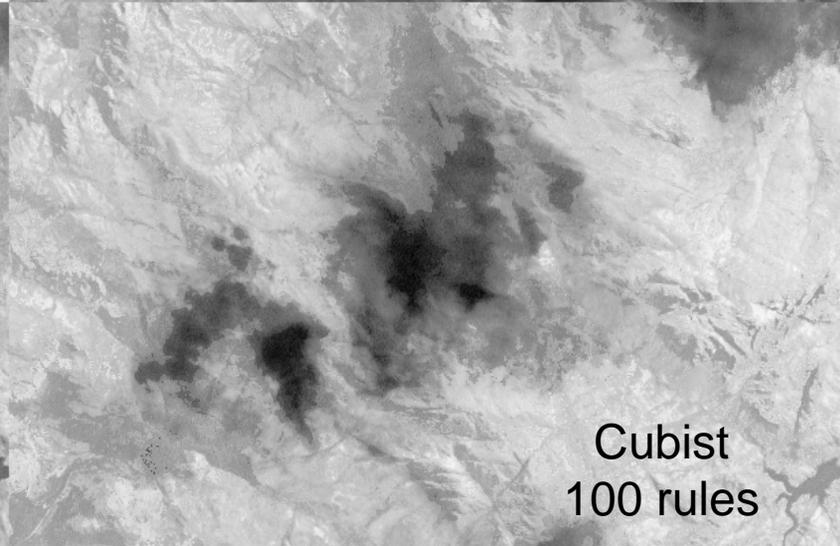
B6 BT



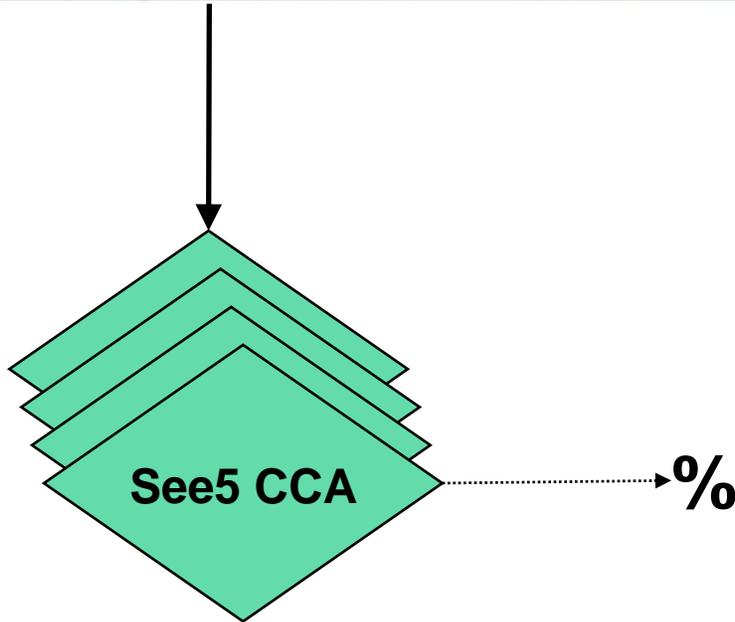
ATv2



Cubist
1 rule



Cubist
100 rules



Decision tree created by Rulequest's See5 classifier software.

Trained with 1 million pixels selected at random from all 212 Irish scenes (~4,700 points per scene.)

Current implementation is 252 'if...then...else' statements, totalling 1526 lines of C++ code.

Returns two classes (cloud or clear), with confidence scores.

Will continue to develop this algorithm to increase accuracy and decrease complexity. Will also include ALI and Hyperion in future training data.

Percentage of pixels tested by band or equation:
(SA = solar elevation angle)

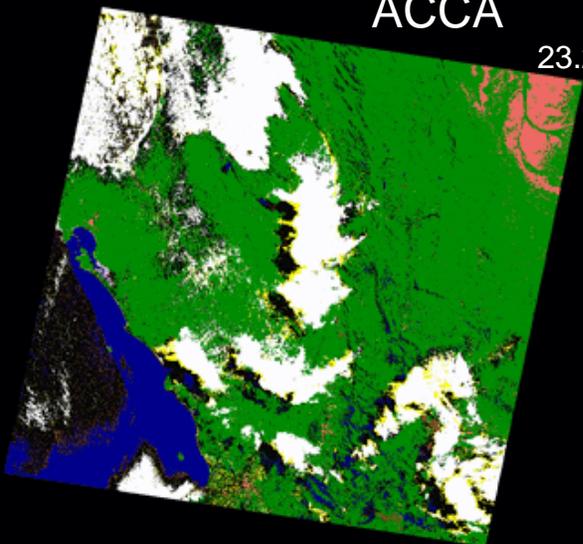
100%	B1	50%	B3
100%	B2	45%	B4/B3
100%	$\cos(\text{SA}) * \text{B1}$	35%	B4/B5
92%	B5	32%	$\cos(\text{SA}) * \text{B5}$
92%	B4	28%	$\cos(\text{SA}) * \text{B4}$
87%	B4/B2	11%	$\cos(\text{SA}) * \text{B2}$
83%	B5/B2	10%	$\cos(\text{SA}) * \text{B3}$
58%	B7	4%	$\cos(\text{SA}) * \text{B7}$
52%	B7/B2		

CCA phase 1 Mask examples (full)

DCM
See5-CCA

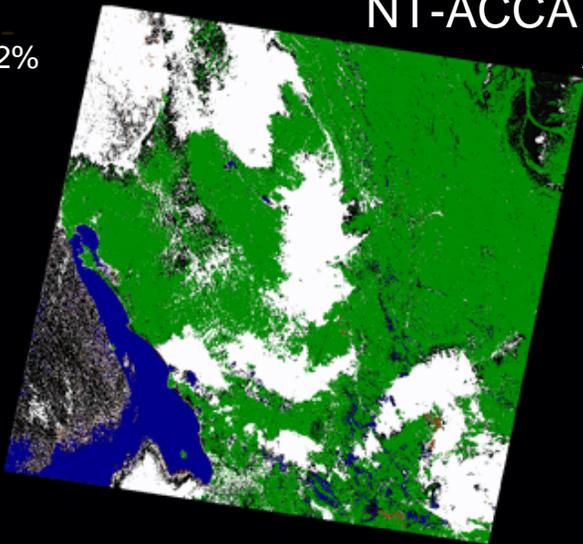
ACCA

23.2%



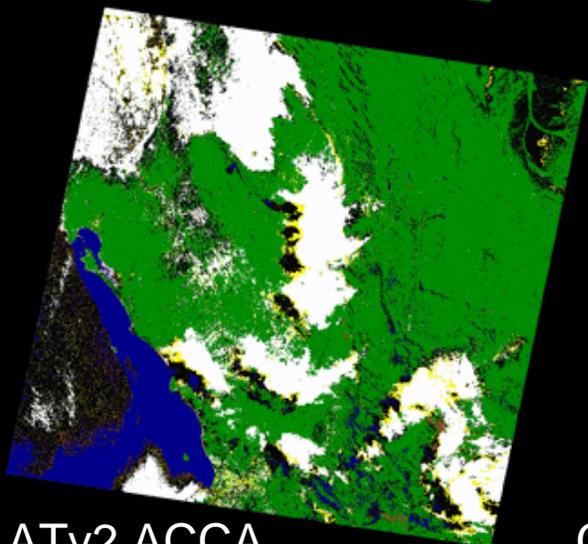
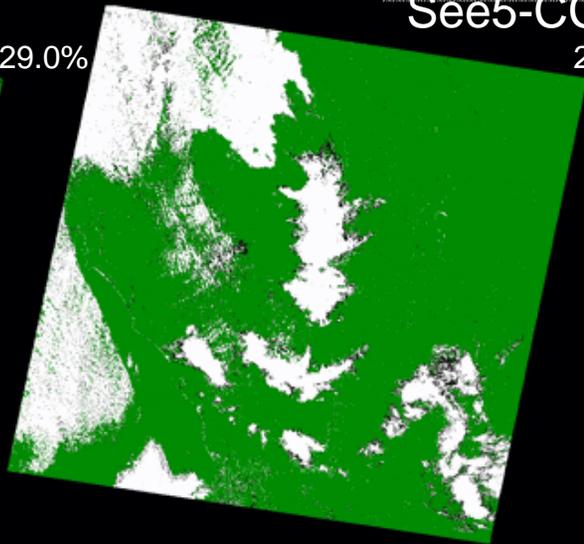
NT-ACCA

29.0%



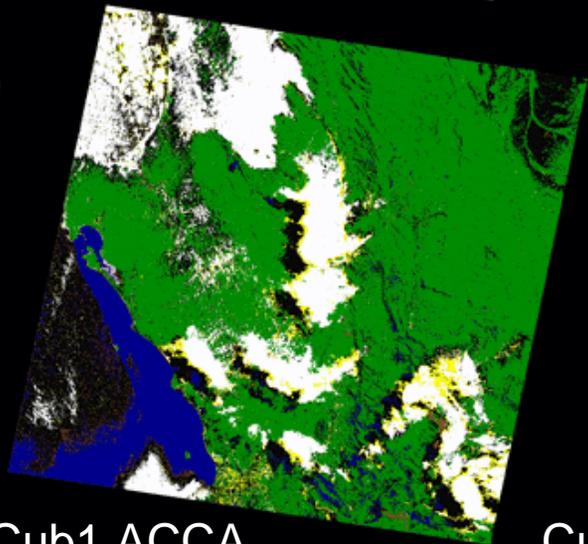
See5-CCA

28.1%



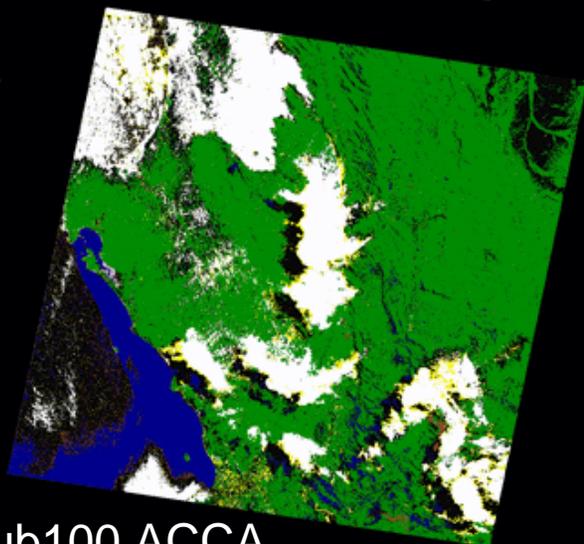
ATv2 ACCA

24.9%



Cub1 ACCA

22.9%



Cub100 ACCA

23.0%

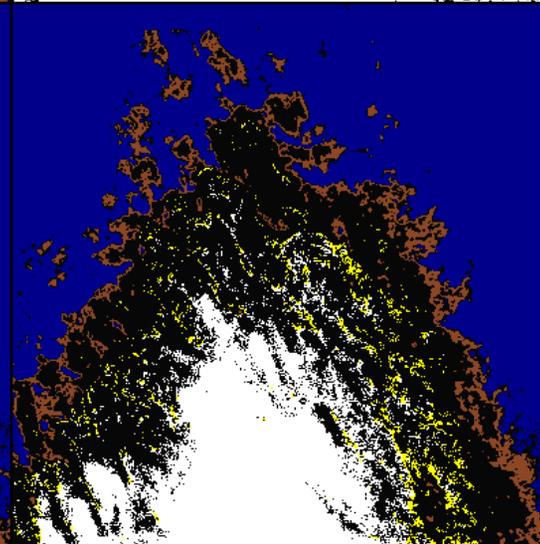
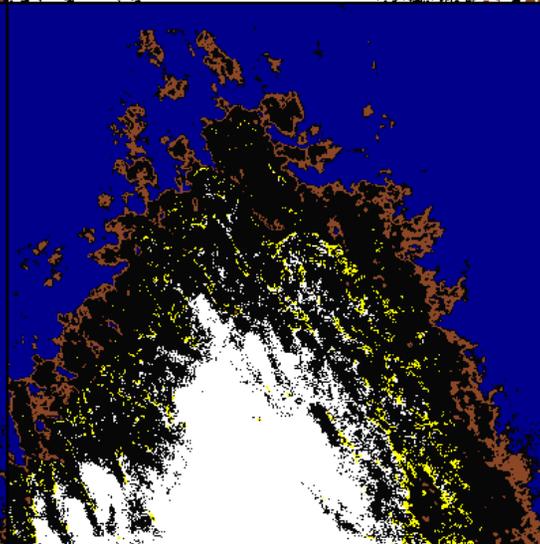
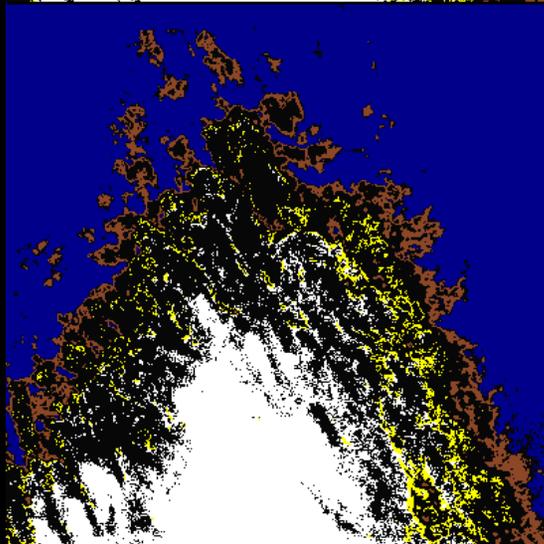
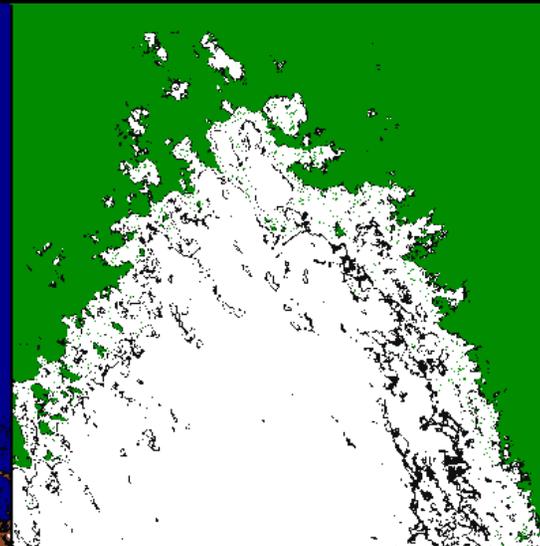
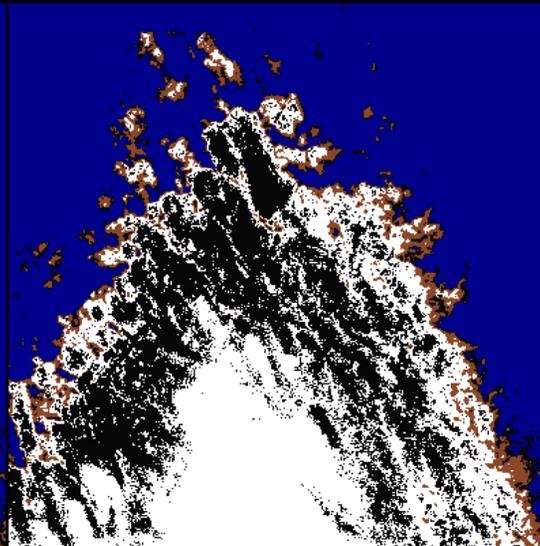
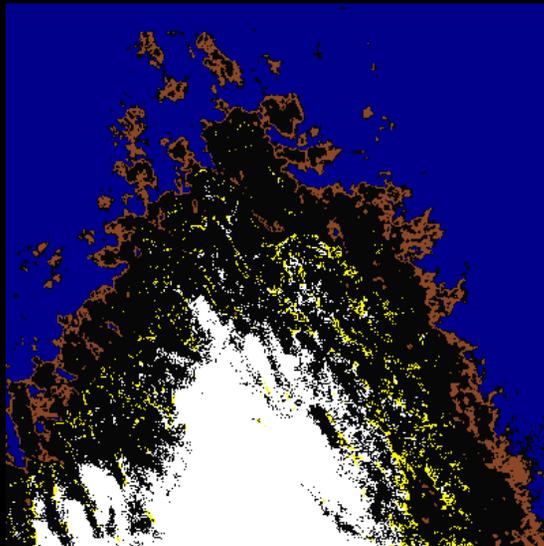


CCA phase 1 Mask examples (zoom)

ACCA

NT-ACCA

See5-CCA

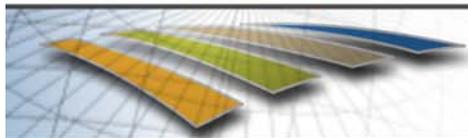


ATv2 ACCA

Cub1 ACCA

Cub100 ACCA





CCA Comparisons to ACCA p1

Difference from ACCA p1 scene-wide cloud score for the 212 scene Irish data:

Error vs. ACCA p1	NT-ACCA		ATv2-ACCA		Cub1-ACCA		Cub100-ACCA		See5-CCA	
0-5%	118	56%	179	84%	186	88%	200	94%	131	62%
0-10%	170	80%	198	93%	201	95%	207	98%	161	76%
0-15%	189	89%	203	96%	204	96%	210	99%	183	86%
0-20%	195	92%	207	98%	208	98%	211	99.5%	189	89%
0-25%	203	96%	210	99%	209	99%	211	99.5%	195	92%
0-30%	205	97%	211	99.5%	211	99.5%	211	99.5%	199	94%
0-35%	207	98%	211	99.5%	211	99.5%	211	99.5%	203	96%
0-40%	209	99%	212	100%	212	100%	212	100%	204	96%
0-45%	211	99.5%							205	97%
0-50%	212	100%							207	98%
0-55%									208	98%
0-60%									210	99%
0-65%									211	99.5%
0-70%									212	100%
Avg. Diff:	6.7%		2.9%		2.2%		1.3%		7.9%	

Caveats:

These results are differences between phase 1 algorithms, and thus do not count ambiguous pixels.

The 'truth' in these comparisons is ACCA, which is known to be imperfect.

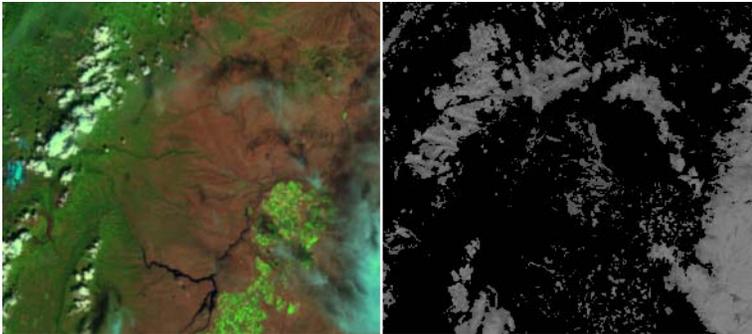
See5-CCA is not related to ACCA. Its errors in this table are compounded with the errors of all ACCA-like algorithms.

Other CCA Algorithms

LDCM

◆ Cirrus CCA

- Simple threshold test using the 1.38 micron band.
- Tested with Hyperion data, looks reliable.



◆ Temporal Change Algorithm

- Compares scene to previously acquired imagery, and marks pixels that have changed.
- Will require an archive of data, thus will not be implemented until after OIVP period for LDCM.
- Might function with full resolution Browse.

◆ Parallax CCA

- Theorized, but no useful methodology yet exists.
- May be redundant with Cirrus CCA.



- ◆ Project is open to consideration of other CCA modules.



Future Work

LDCM

- ◆ Create Training and Test sets.
 - Create manual masks of all Irish test scenes.
 - Create set of OLI simulated data with manual masks, from Hyperion and ALI data.
 - Connect to L7 archive systems for large supply of test data.
 - Already done for level 0 data, through IAS.
 - Portion all scenes (Irish, OLI-sim, L7 archive) into training and test sets.

- ◆ Decide upon and implement final suite of algorithms.
 - Select an AT band for use in an AT-ACCA algorithm.
 - Train a final See5 algorithm.
 - Possibly add a Land/Water lookup mask.
 - Develop mask resolver algorithm.
 - Currently have a supervised k-mean classification that works, but could be better.
 - Implement algorithms.
 - When instrument RSR is available, revisit all algorithms and tune thresholds.

- ◆ Develop methodology for monitoring CCA performance.
 - IAS-based CCA monitoring and parameter adjustment system.

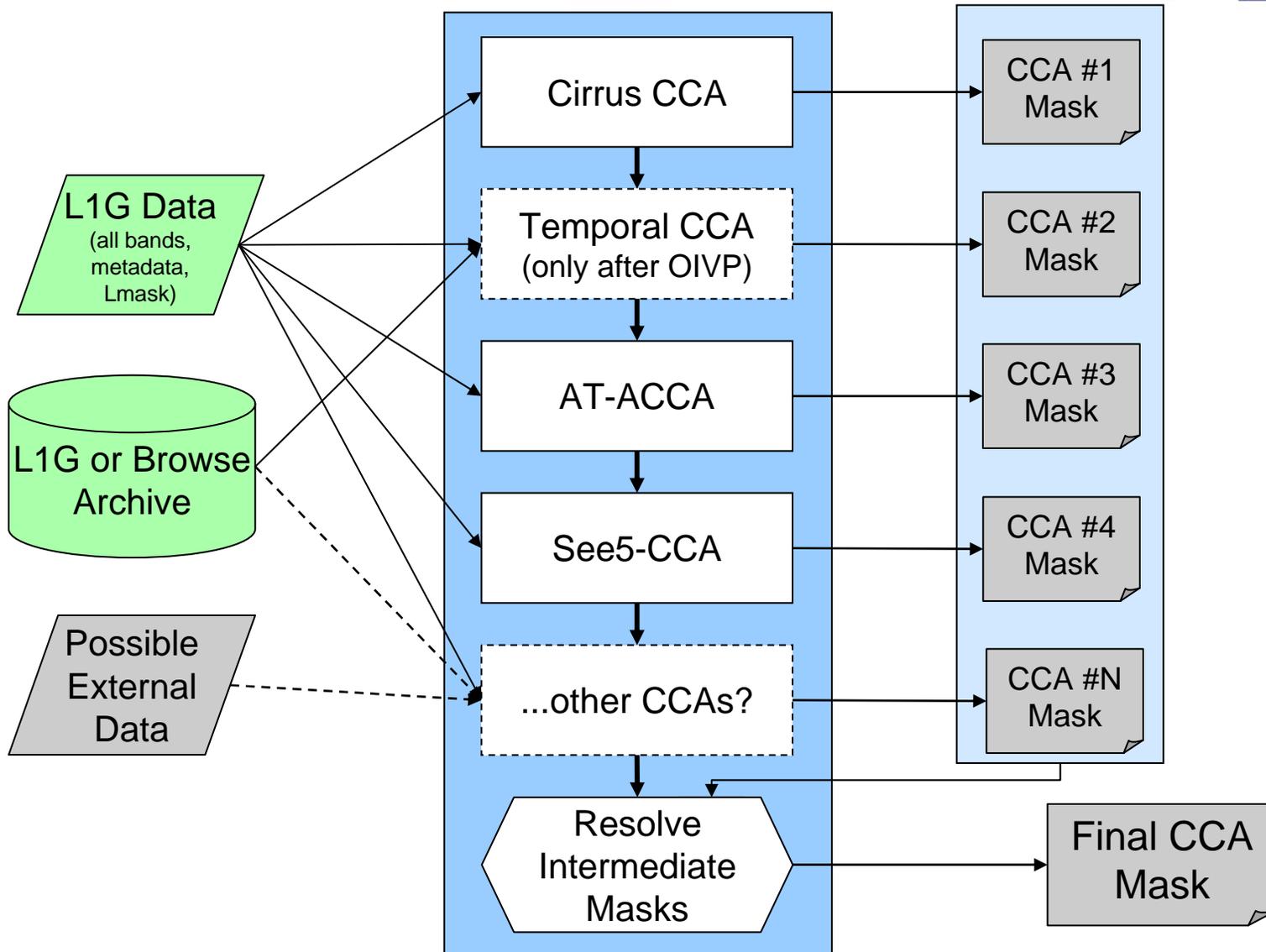
- ◆ Launch.

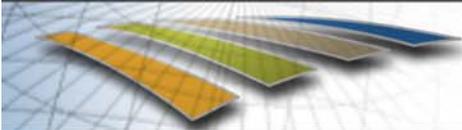
- ◆ Activate post-launch algorithms. (Temporal)

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

Proposed System

LDCM





CCA Comparisons to Manual Masks

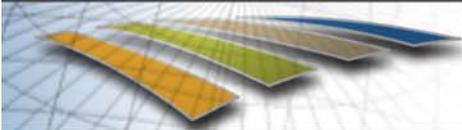
LDCM

- ◆ 19 scenes from the Irish test set were analyzed with manually-created cloud masks, and the p1 CCA algorithms were tested against them.

% of scene pixels	ACCA*	NT ACCA	ATV2 ACCA*	Cub1 ACCA*	Cub100 ACCA*	See5 CCA
Correct	70.42%	65.06%	56.62%	56.15%	58.84%	91.33%
False	11.96%	15.18%	16.03%	14.56%	13.40%	6.05%
Ambiguous	17.62%	19.76%	27.35%	29.30%	27.76%	2.62%

- ◆ **BIG Caveats:**

- (*) -- For this comparison, the provisional cloud class was treated as the full cloud class. This increased errors for all the algorithms except NT-ACCA, which has no provisional class, and See5-CCA, which reports confidence level for all pixels. Our methodology for manual mask comparison will be improved.
- Very small sample size. The scenes selected were known to be difficult, with one or two 'easy' scenes thrown in as controls. Thus, the results are probably not representative of results over the entire Irish test set.
- 11 scenes were done in triplicate, with three different operators. Their final CCA scores differed by an average of 6.75% (maximum of 24%). So manual assessment for this small test set should be considered to have ~6-7% error.



AT v2 and v3 equations

LDCM

$$AT_{v2} = \frac{1}{C_1 B2 + C_2} + C_3 B5 + C_4$$

$C_1 = 0.1366$

$C_1 = 36.5441$

$C_2 = 0.0087$

$C_2 = 235.4310$

$$AT_{v3} = \frac{1}{(C_1 B2 + C_2)^2} + \frac{1}{(C_3 B5 + C_4)} + C_5$$

$C_1 = 11.4294$

$C_4 = 0.11149$

$C_2 = 0.40292$

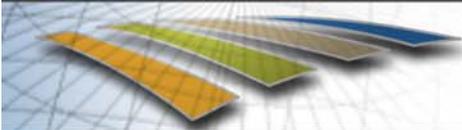
$C_5 = 269.768$

$C_3 = -4.29418$

Other forms were considered, with the goals of:

- Eliminating bright cloud bloom due to B5.
- Approaching a physically explicable equation.

(Stephan-Boltzmann law predicts $T \sim (1-B2)^{1/4}$)

 AT v4 equations

$$\begin{aligned} \text{ATv4 Cubist r1} = & + 442 * B1 & - 895 * cS * B1 \\ & - 405 * B2 & + 714 * cS * B2 \\ & - 147.3 * B3 & + 331 * cS * B3 \\ & + 38.3 * B4 & - 141 * cS * B4 \\ & - 197.1 * B5 & + 549 * cS * B5 \\ & + 430.1 * B7 & - 960 * cS * B7 \\ & - 15.9 * \text{NDVI} & - 17.2 * \text{NDxI} \\ & + 5.1 * (B4/B3) & - 3.7 * (B4/B2) \\ & + 302.2927 & \end{aligned}$$

where $cS = \cos(\text{solar elevation angle})$

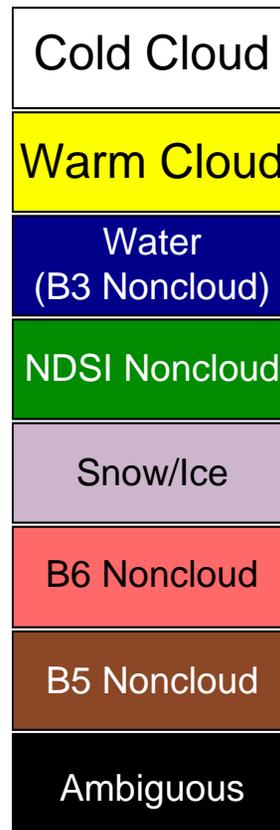
$\text{NDVI} = (B4 - B3) / (B4 + B3)$

$\text{NDxI} = (B1 - B7) / (B1 + B7)$

ATv4 Cubist r100 is too involved for presentation display. (790 lines of C++ code)

CCA Examples -- Legend

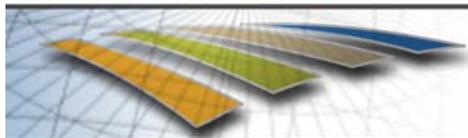
The 'Water' class is loosely defined. Cloud shadows also fall into this class, as may some types of dark terrain.



The 'Warm Cloud' class is for provisional clouds. Pixels in this class may be reverted back to the ambiguous class based on Band 6 scene-wide tests in ACCA phase 2.

In ACCA algorithms with confidence scoring, pixels in the 'Warm cloud' class will be marked as clouds with lower confidence.

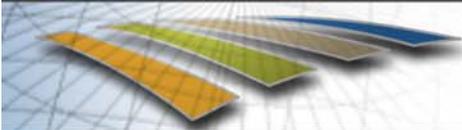
This class only exists for ACCA. It has been disabled in all versions of AT-ACCA.



CCA phase 1 Ambiguous Pixels

Scene percentage of ambiguous pixels for the Irish data:

Ambig. pixels	ACCA		NT ACCA		ATv2 ACCA		Cub1 ACCA		Cub100 ACCA		See5 CCA	
	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
0-5%	52	25%	78	37%	46	22%	45	21%	43	20%	155	73%
0-10%	87	41%	122	58%	79	37%	77	36%	79	37%	193	91%
0-15%	120	57%	146	69%	104	49%	99	47%	104	49%	204	96%
0-20%	154	73%	174	82%	139	66%	132	62%	137	65%	207	98%
0-25%	177	83%	184	87%	158	75%	157	74%	162	76%	210	99%
0-30%	189	89%	192	91%	175	83%	171	81%	176	83%	211	99.5%
0-35%	198	93%	201	95%	185	87%	184	87%	187	88%	211	99.5%
0-40%	204	96%	203	96%	195	92%	190	90%	193	91%	211	99.5%
0-45%	208	98%	206	97%	203	96%	199	94%	200	94%	211	99.5%
0-50%	210	99%	207	98%	206	97%	203	96%	205	97%	211	99.5%
0-55%	210	99%	210	99%	207	98%	206	97%	207	98%	211	99.5%
0-60%	210	99%	210	99%	209	99%	207	98%	207	98%	211	99.5%
0-65%	211	99.5%	210	99%	209	99%	209	99%	209	99%	211	99.5%
0-70%	211	99.5%	210	99%	210	99%	209	99%	209	99%	212	100%
0-75%	212	100%	210	99%	210	99%	210	99%	210	99%		
0-80%			212	100%	212	100%	212	100%	212	100%		

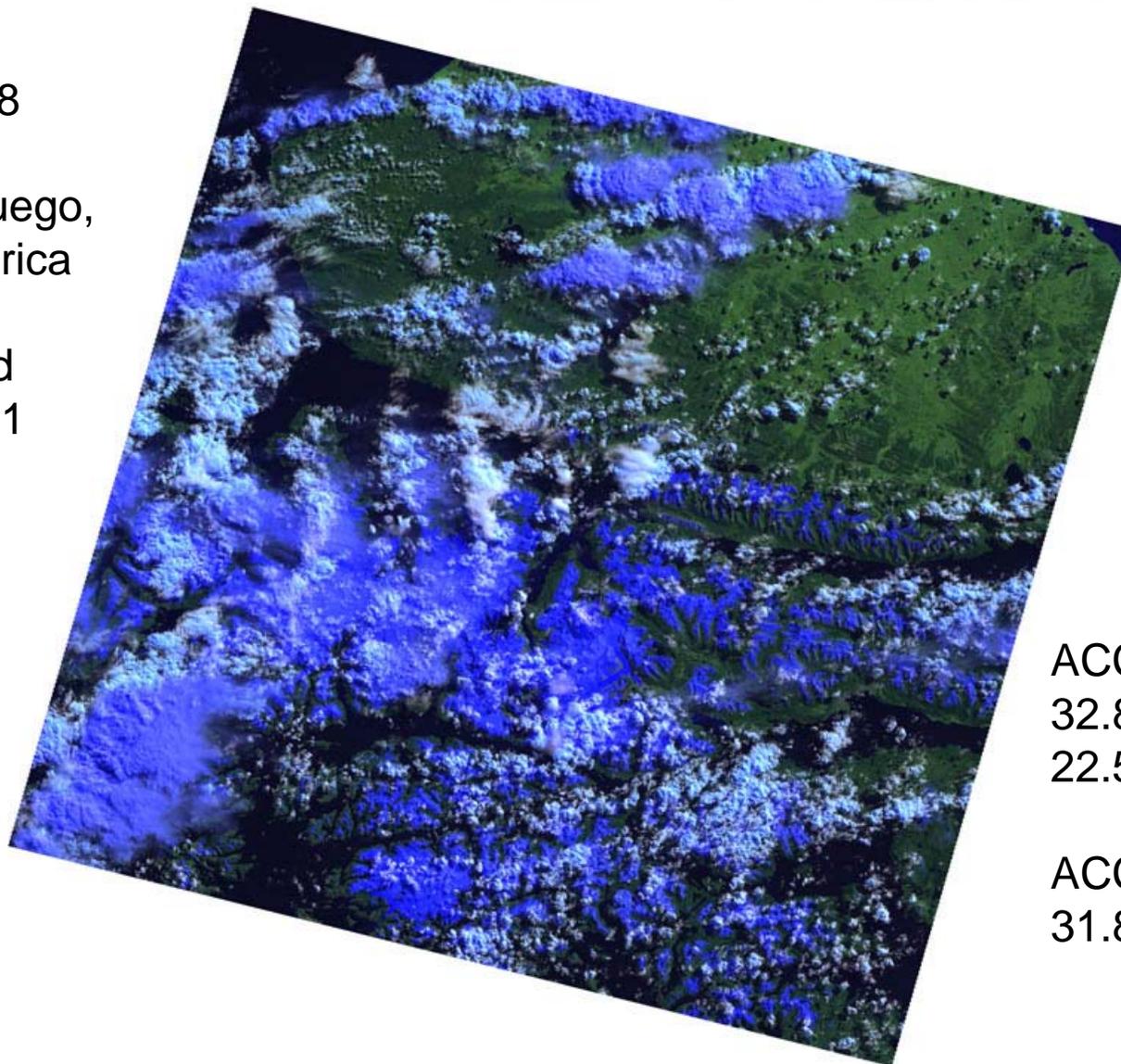
 Irish Set -- Example Scenes

LDCM

Austral
p227_r98

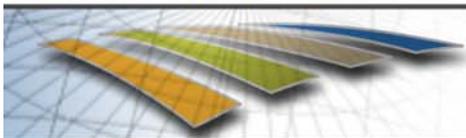
Tierra del Fuego,
South America

Acquired
11/3/2001



ACCA phase1
32.8% clouds
22.5% ambiguous

ACCA full score
31.8% clouds

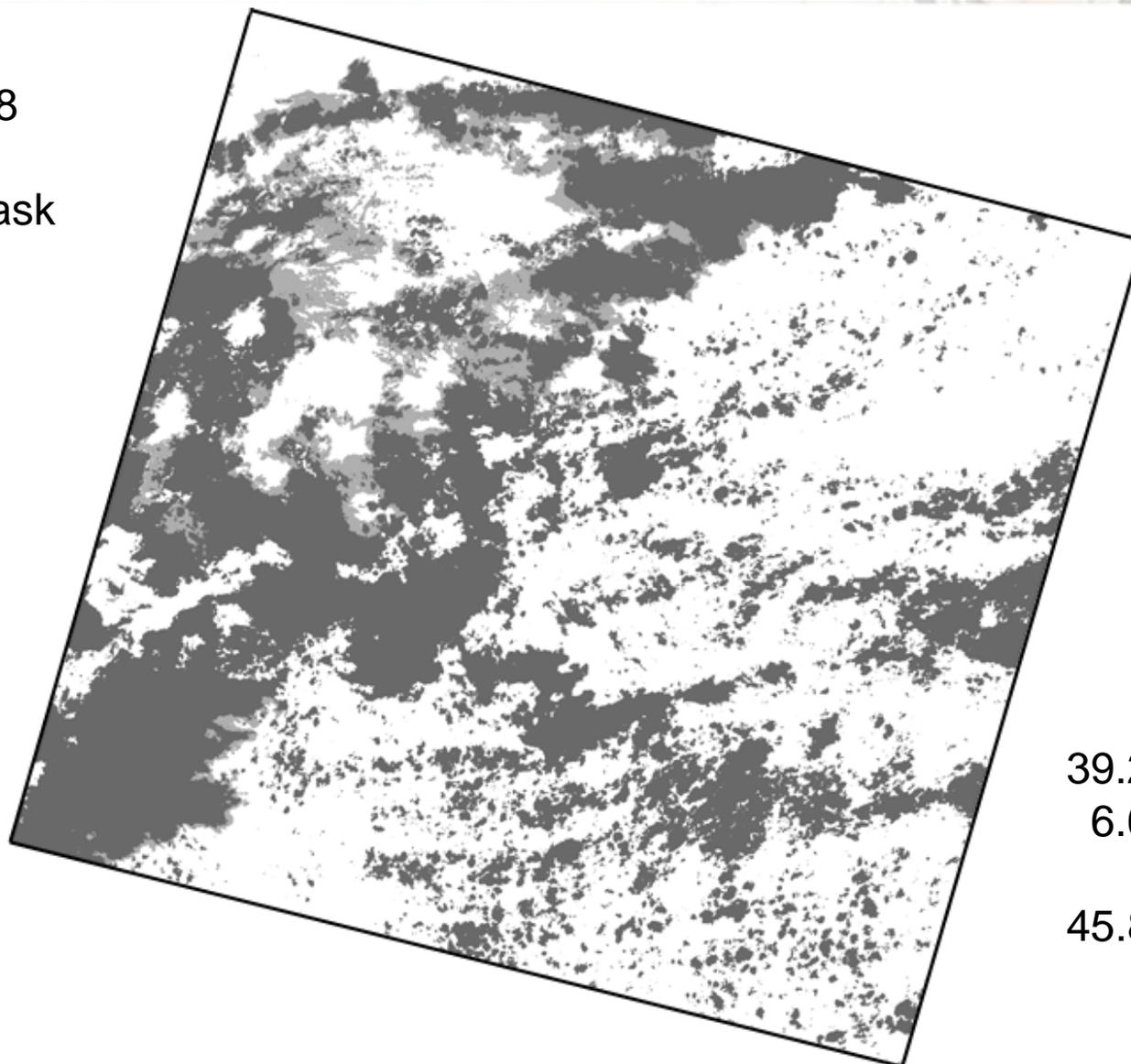


Irish Set -- Example Scenes

LDCM

Austral
p227_r98

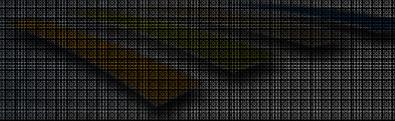
Manual Mask



Thick Cloud
Thin Cloud
Clear

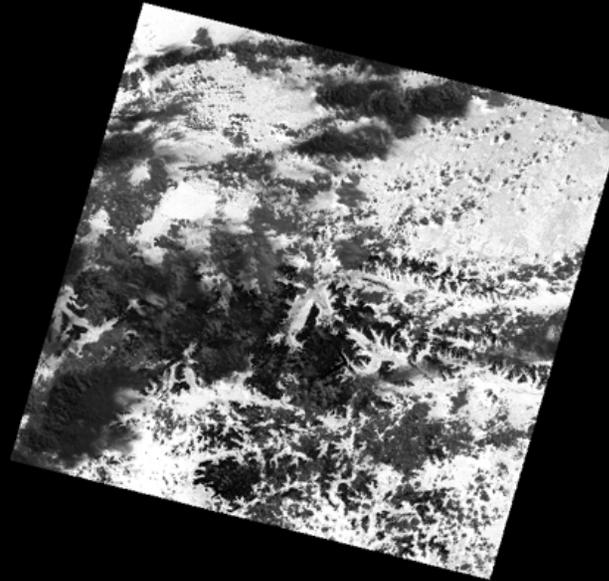
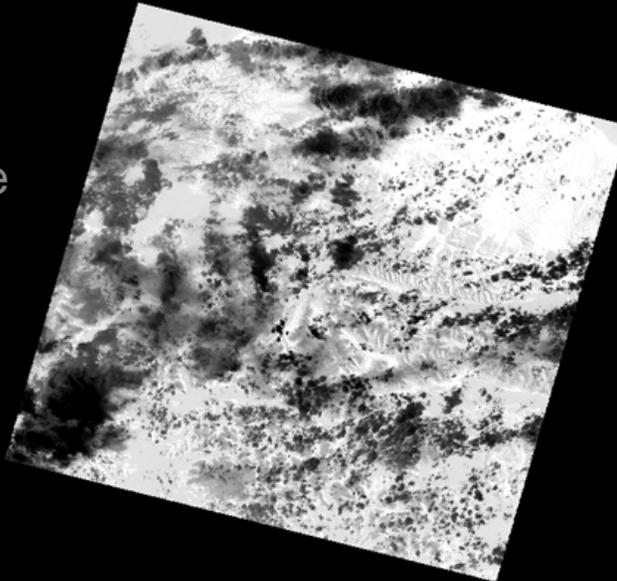
39.2% Thick
6.6% Thin

45.8% Total clouds



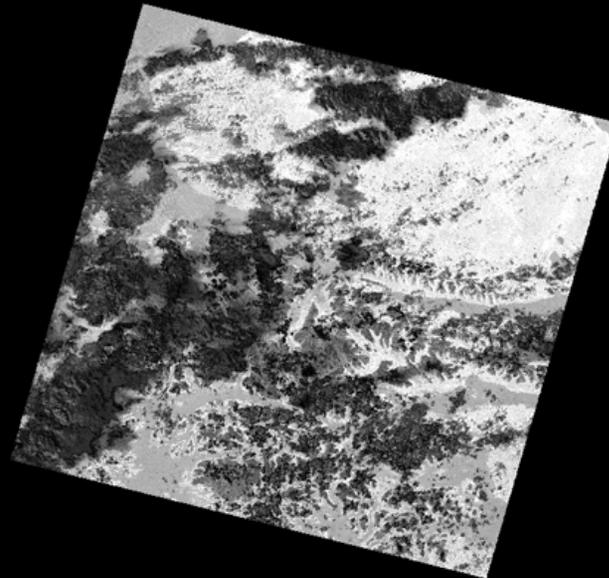
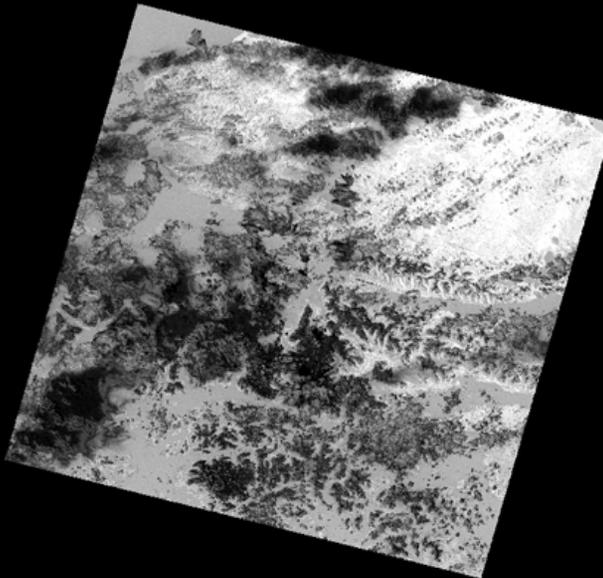
Example of AT bands (full)

Band 6
Brightness
Temperature



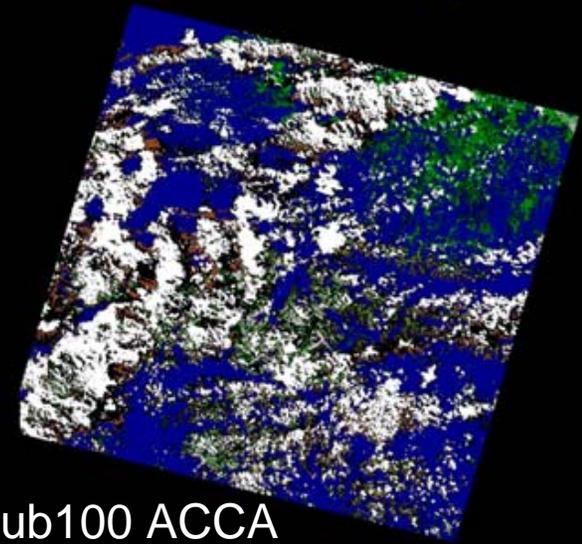
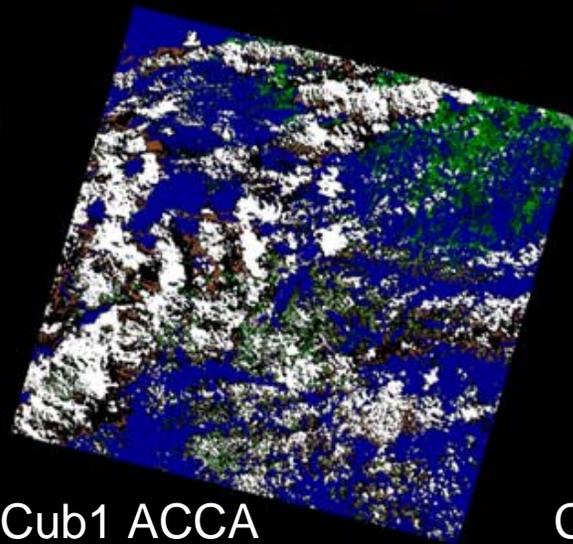
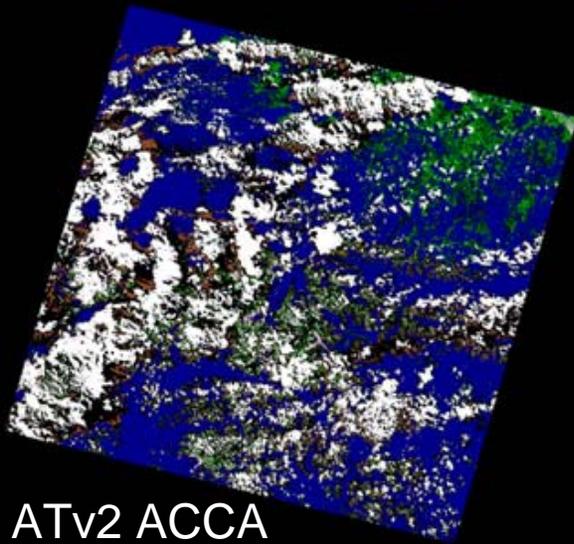
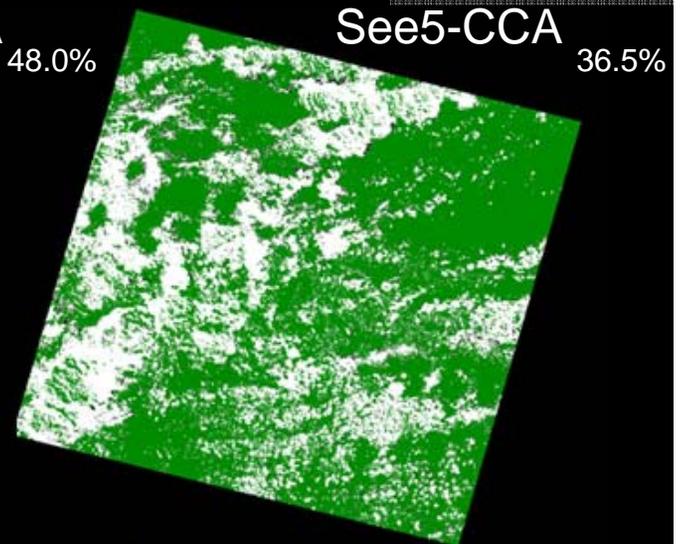
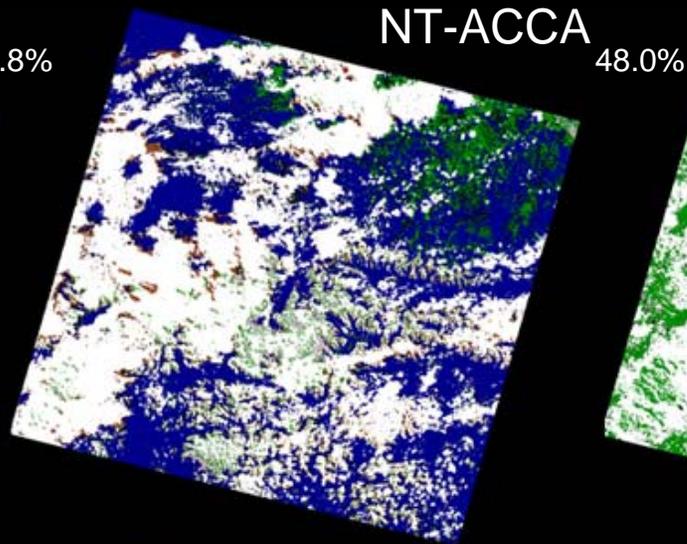
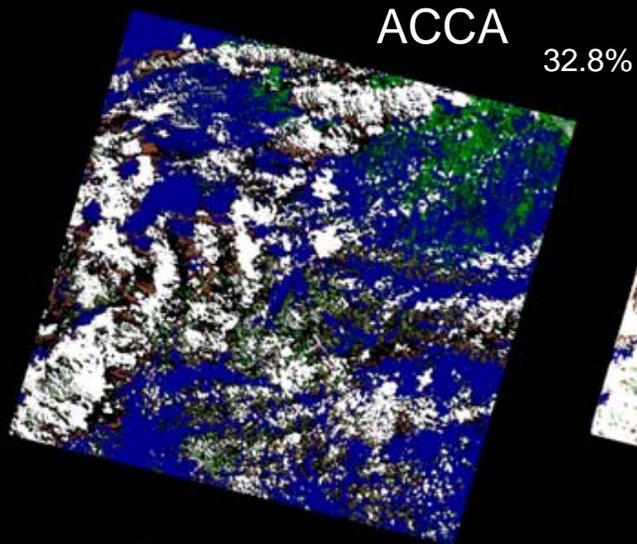
ATv2

Cubist
1 rule



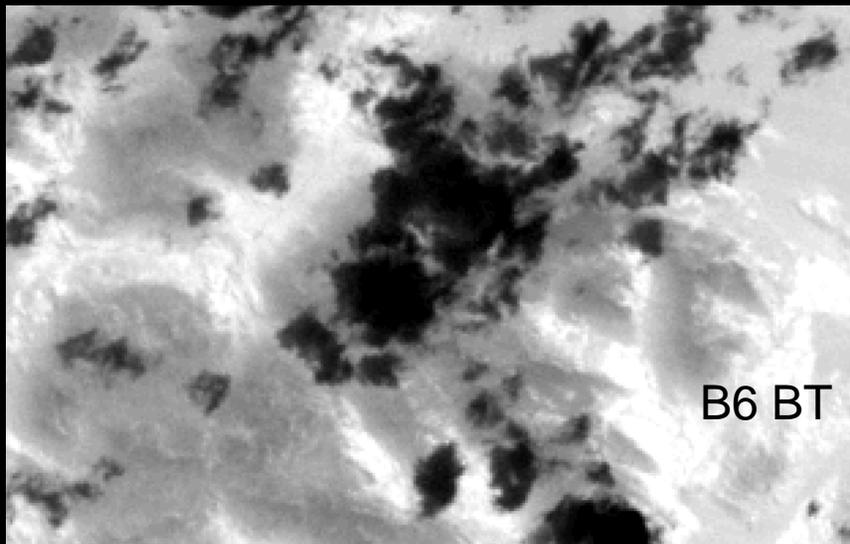
Cubist
100 rules

CCA Mask examples (full)





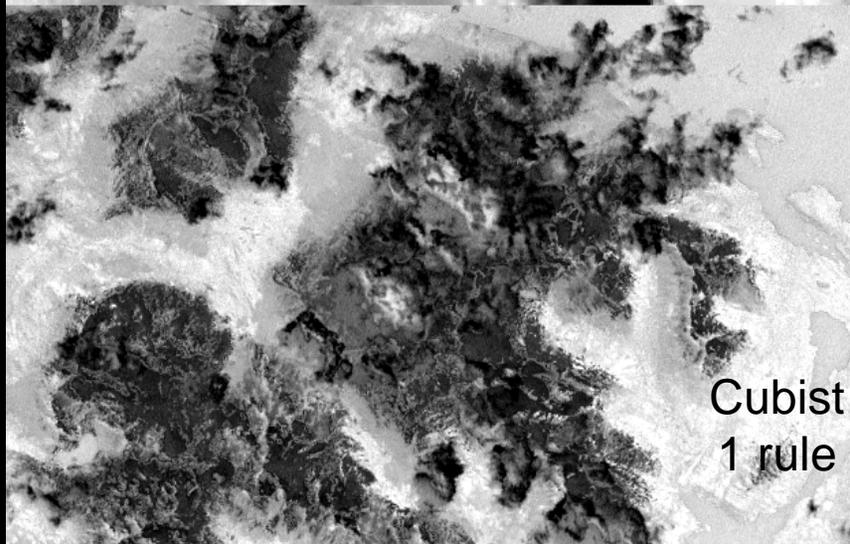
Example of AT bands (zoom)



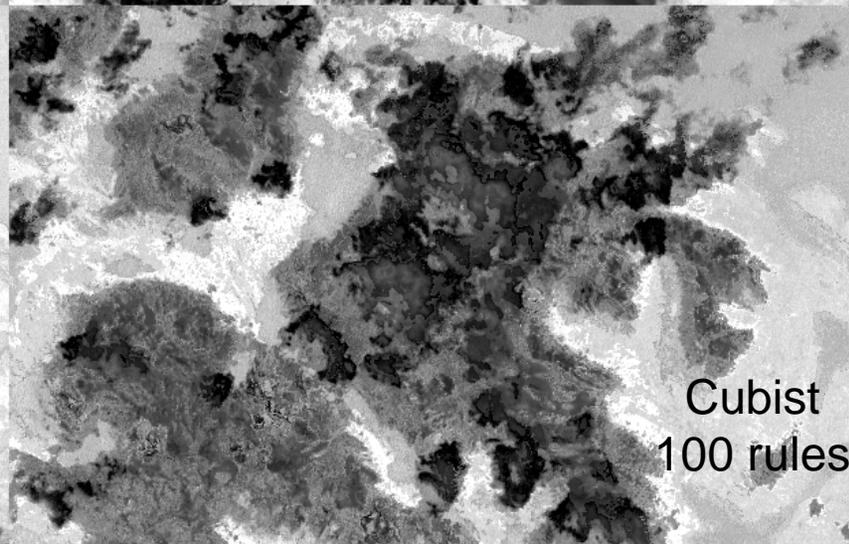
B6 BT



ATv2



Cubist
1 rule



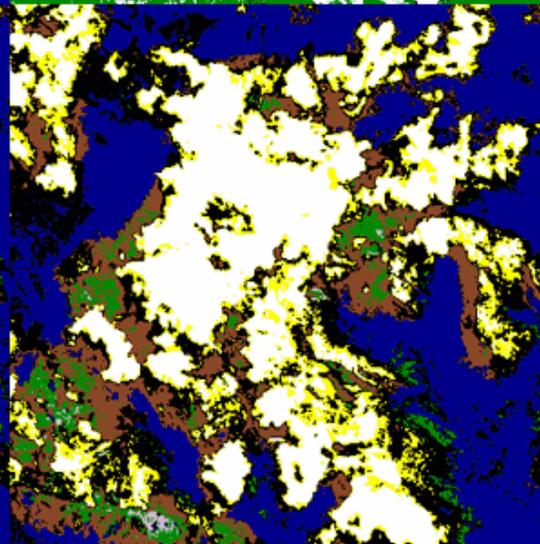
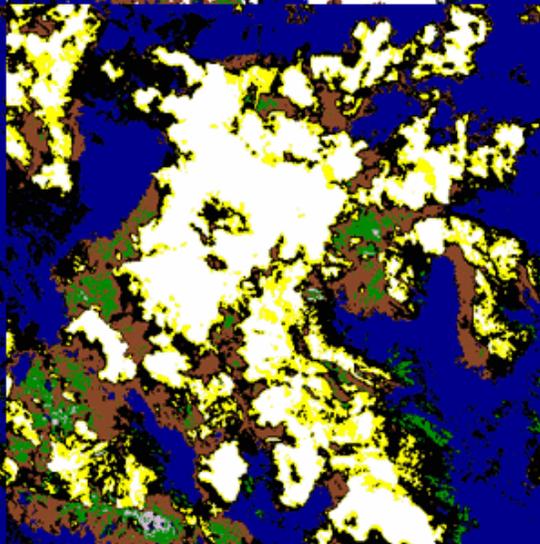
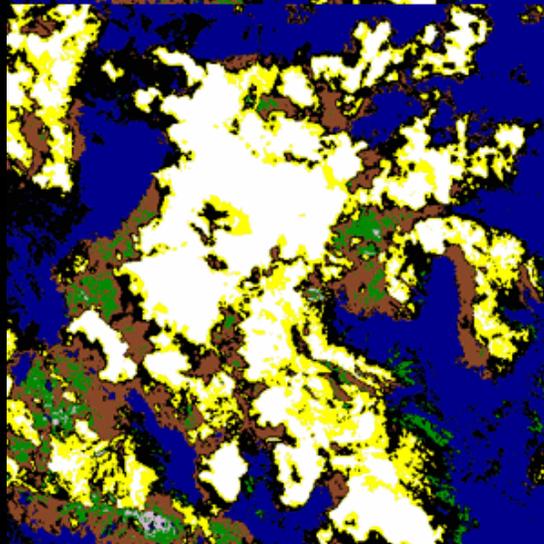
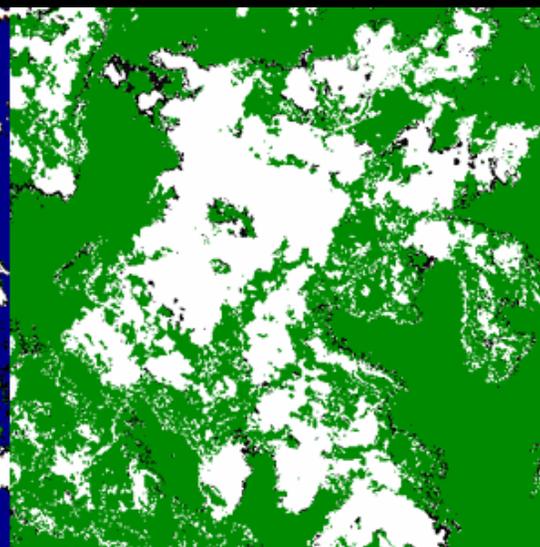
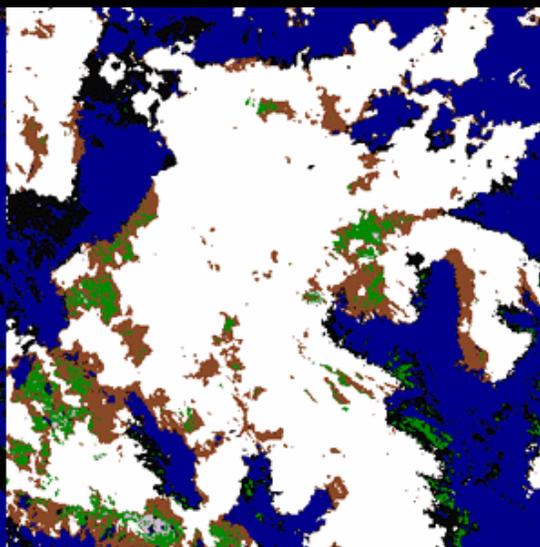
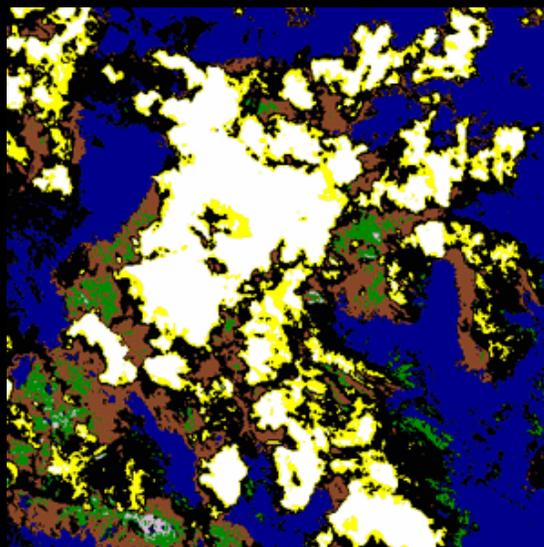
Cubist
100 rules

CCA Mask examples (zoom)

ACCA

NT-ACCA

See5-CCA



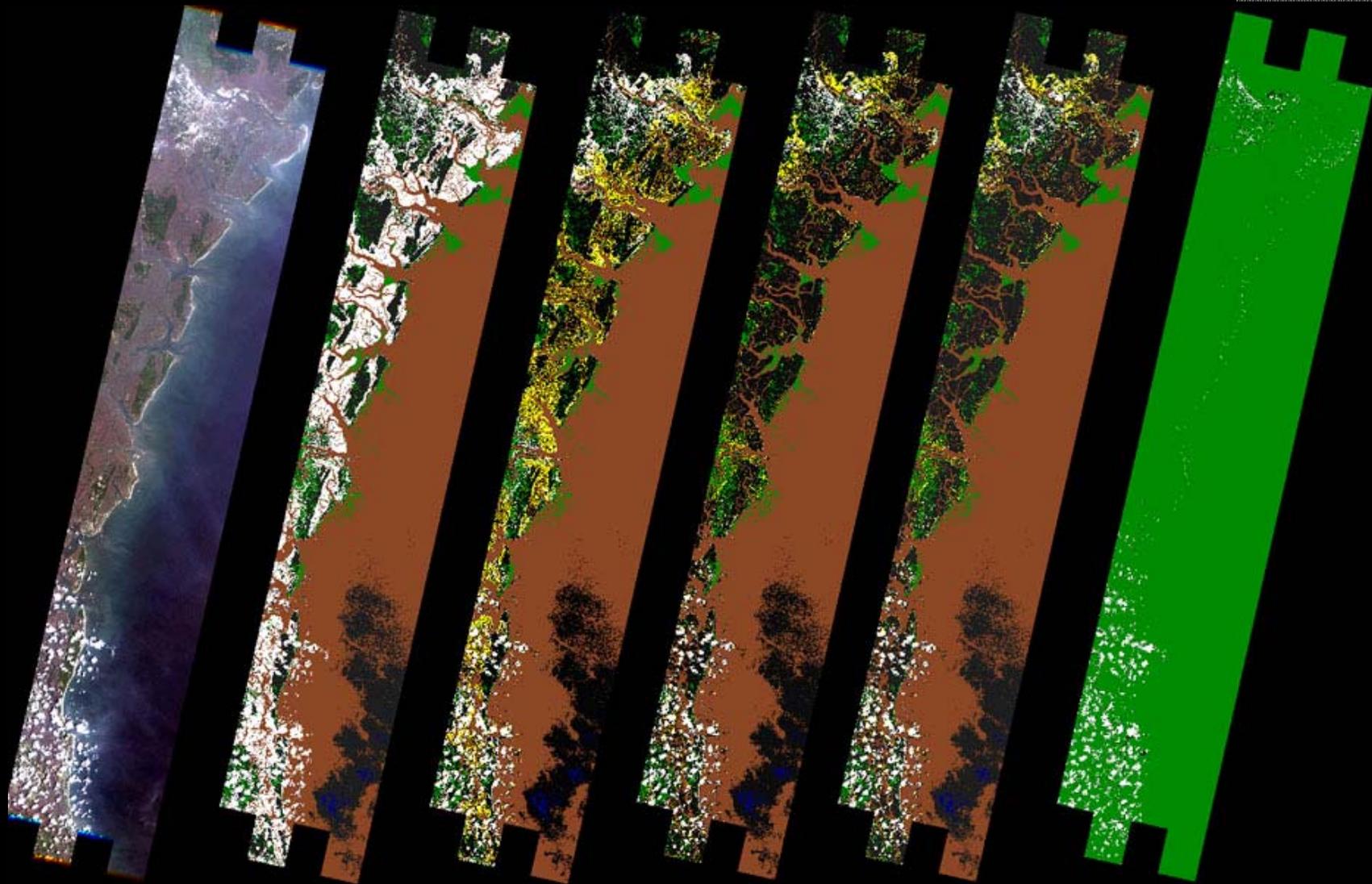
ATv2 ACCA

Cub1 ACCA

Cub100 ACCA



ALI Mask examples



NT-ACCA

ATv2 ACCA

Cub1 ACCA

Cub100 ACCA

See5-CCA



Hyperion Mask examples

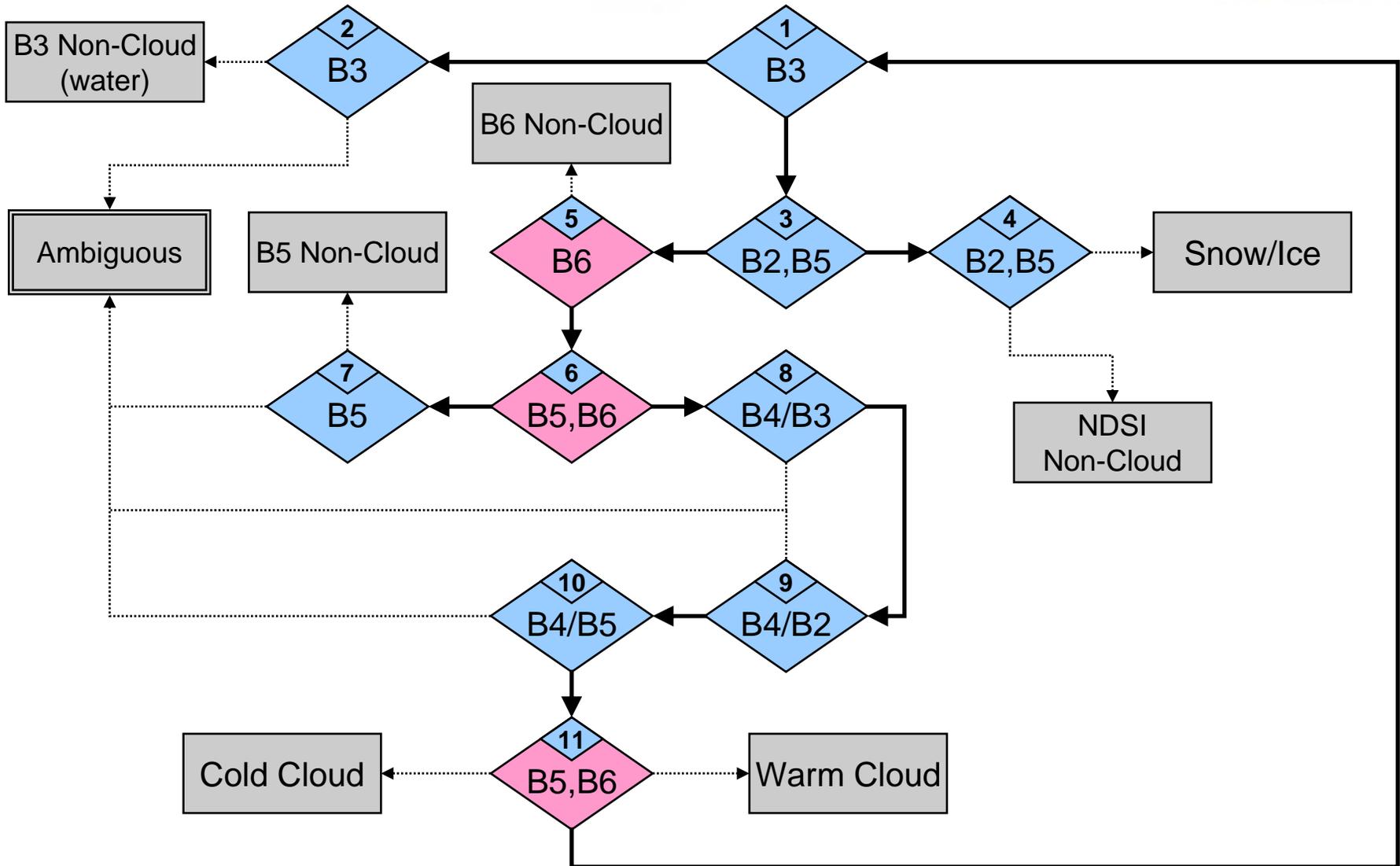


(*) = Includes additional Cirrus band test, index color is 

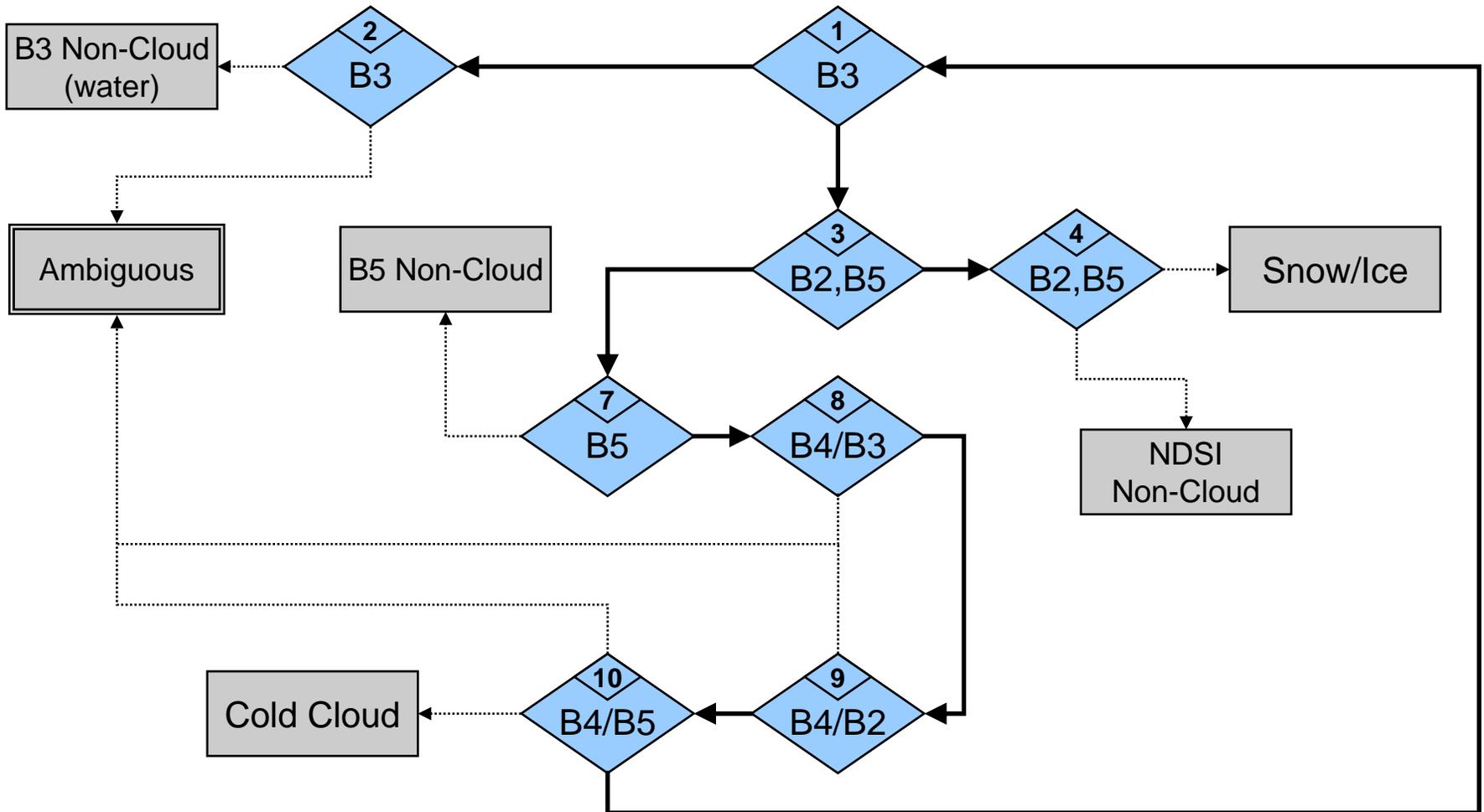


Landsat 7 ACCA Phase 1

LDCM

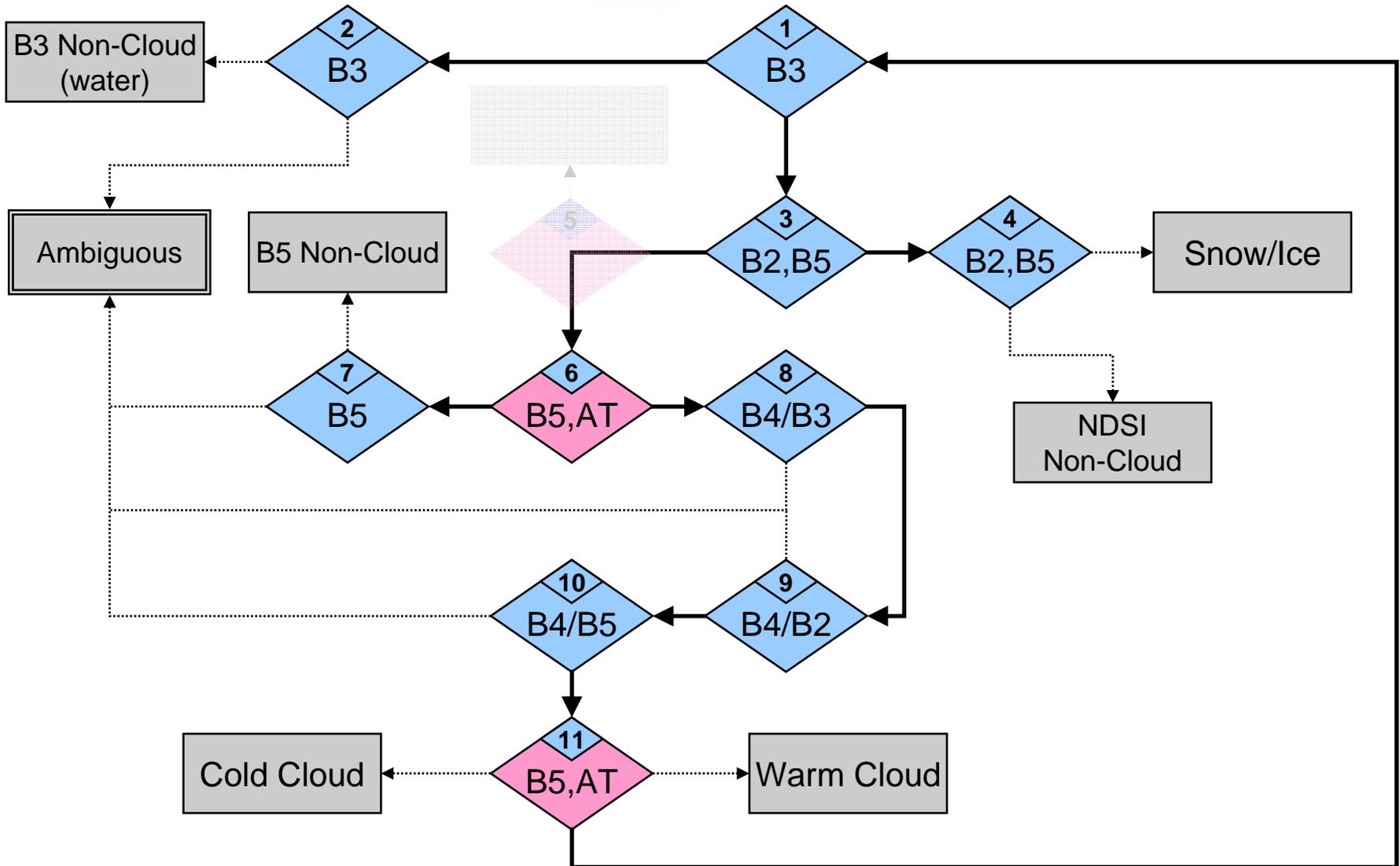


NT-ACCA (for testing)



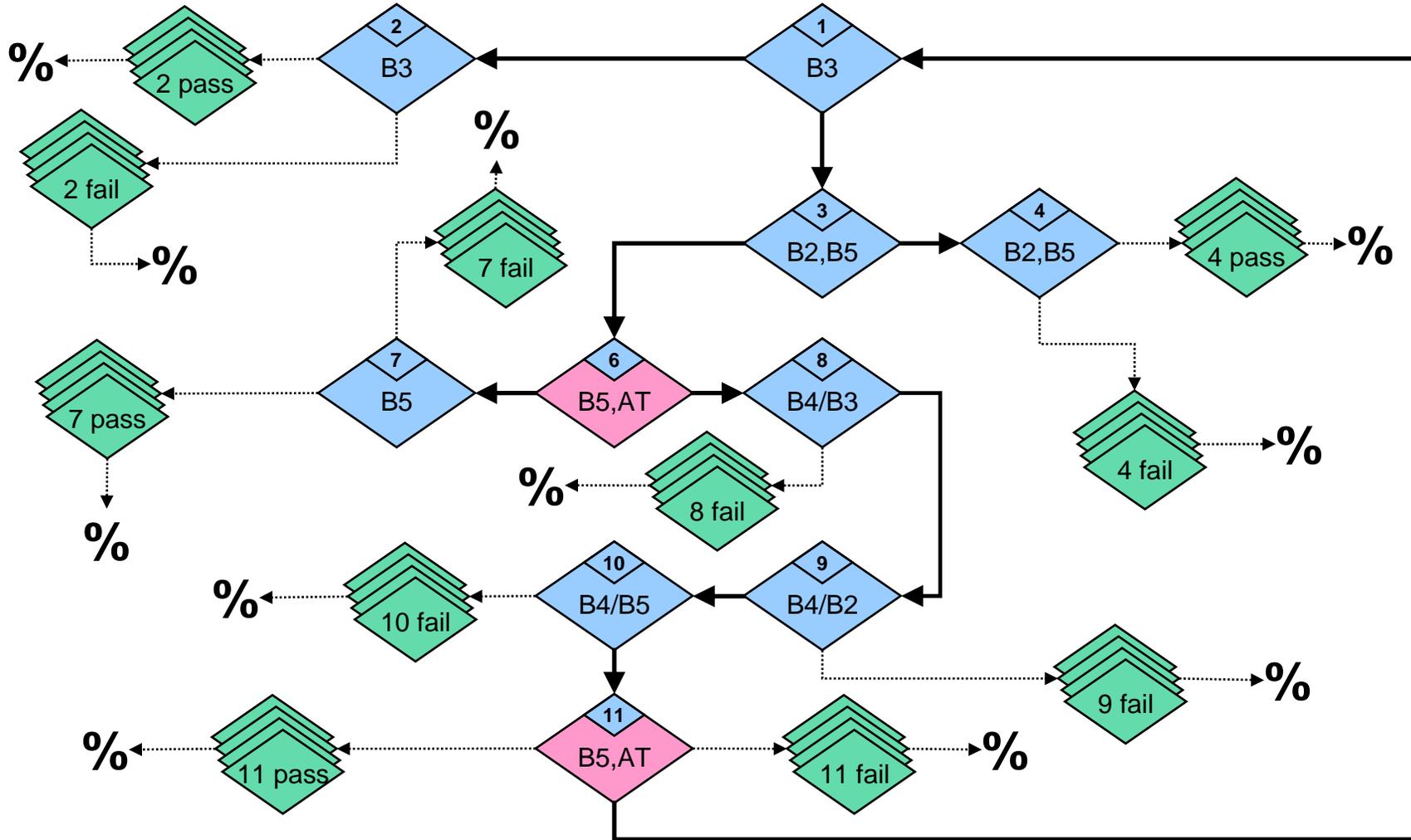
AT-ACCA Flowchart

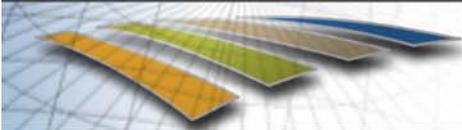
LDCM



AT-ACCA p1 with Confidence

LDCM





Glossary

ACCA	The CCA algorithm as implemented in L7. Often used to refer only to phase 1 of ACCA, the decision tree processing.
AT Band	An 'artificial thermal' band, <i>intended only for use in CCA algorithms</i> .
ATv2	One candidate AT band equation, derived with genetic analysis.
Cubist r100	Another candidate AT band equation, derived using Cubist software with 100 rules. Sometimes labelled 'Cub100'.
Cubist r1	Another candidate AT band equation, derived using Cubist constrained to 1 rule. Sometimes labelled 'Cub1'.
AT-ACCA	An ACCA-like algorithm that uses an AT band.
NT-ACCA	An ACCA-like algorithm, in which all thermal tests have been removed. For comparison purposes.
See5-CCA	A CCA algorithm derived using See5 classification software.
Temporal CCA	A CCA algorithm that detects changes in terrain over time.
Cirrus CCA	A CCA algorithm using the OLI cirrus band to detect cirrus clouds.
Parallax CCA	A fictional CCA algorithm intended to detect high clouds.
Irish Set	The 212 scene test set from NASA Goddard's Rich Irish.