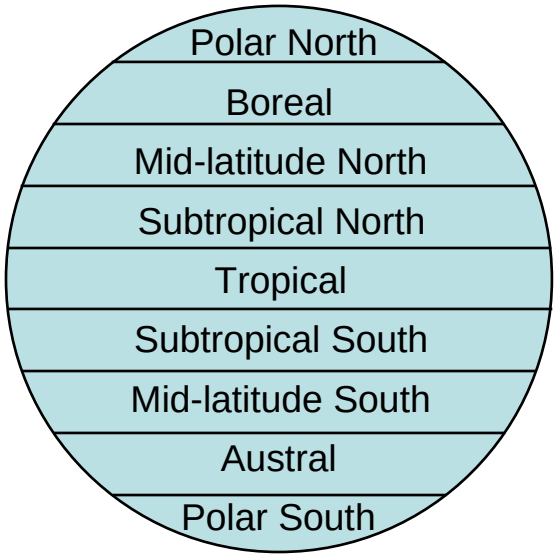
LDCM Cloud Cover Assessment
January 6th, 2009

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

Irish Data Set, now with Cloud Masks

- **212 Landsat 7 L1Gs scenes**
  - 5 scenes were excluded due to visible gain change artifacts.
    - One scene had a gain change but no visible artifacts; it was retained.
- **207 scenes with manual cloud masks**
  - Manual masks were created by three operators at EDC.
  - The classifications were clear, thick cloud, thin cloud, and shadow.
    - Shadows were only masked where they were hard-edged and simple.
    - 46 scenes have shadow masking.
  - 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 pscar@usgs.gov for instructions.

Algorithms under review at EDC

- See5 CCA
  - classification algorithm that works by minimizing the entropy of iterative classifications of the training data.
  - COTS software See5/C5.0, by Rulequest.

- ACCA phase 1
  - Only portion of L7 ACCA algorithm that creates a pixel mask.
  - Poor classifier, but good at cloud/clear distinction.

- AT-ACCA
  - Expansion of ACCA using a calculated thermal band (the AT band).
  - AT band algorithm is created using numerical modelling software Cubist, by Rulequest.
  - Expanded AT-ACCA will be AT-ACCA merged with g4 and Cirrus algorithms.

  - g4
    - Based on examination of the histograms of all bands and all two-band combinations.
    - Threshold tests were developed from the histograms using iterative searches for maximum likelihood ratio. (This minimizes type I error and maximizes the selected percentile.)
    - Tests with performance less than the worst non-thermal ACCA tests were excluded.
    - The threshold tests with the lowest type II error in each band or band combination were implemented.
    - Creates cloud/clear/ambiguous classification based on the tally of tests passed.

  - Cirrus
    - Preliminary tests done with Hyperion data, but needs more study.
    - Expected to be a simple threshold test on the OLI cirrus band.

- kmean
  - Basic unsupervised classification algorithm.
  - Results are poor, but the flexibility of the approach deserved study.

See5 Results

Round B2 algorithm, 10/2008

104 Training scenes

Pixels in training set:	7,998,842
Non-saturated pixels:	6,560,940
Saturated pixels:	1,437,902

103 Validation scenes

Total number of validation pixels:	3,955,741,295

Scene level validation results

% error	# of scenes	% of scenes	cum. %
0 - 5	62	60%	60%
5 - 10	24	23%	83%
10 - 15	7	7%	90%
15 - 20	3	3%	93%
20 - 25	1	1%	94%
25 - 30	0	0%	94%
30 - 35	3	3%	97%
35 - 40	1	1%	98%
40 - 45	1	1%	99%
45 - 50	0	0%	99%
50 - 55	1	1%	100%

Pixel level validation results

Conf. %	Clear	Shadow	Thick	Thin	All Clouds	All Clear
< 35%	92.1%	90.8%	4.2%	28.4%	12.3%	92.1%
35-65%	2.4%	3.5%	1.8%	8.5%	4.1%	2.4%
> 65%	5.5%	5.7%	94.0%	63.1%	83.6%	5.5%

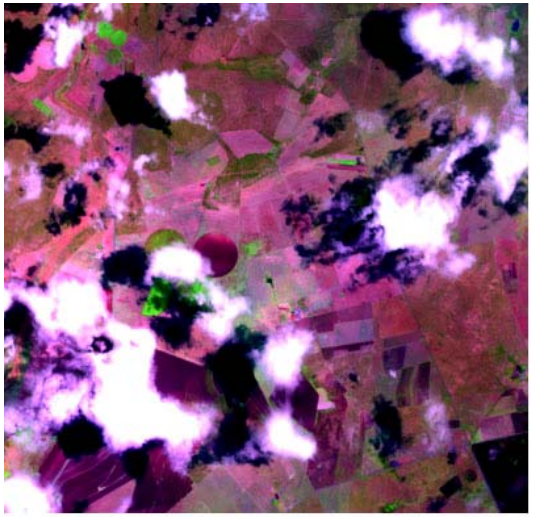
Total Pixels Correct:	89.0%
Total Pixels False:	8.0%
Total Ambiguous:	3.0%

CCA Algorithms compared

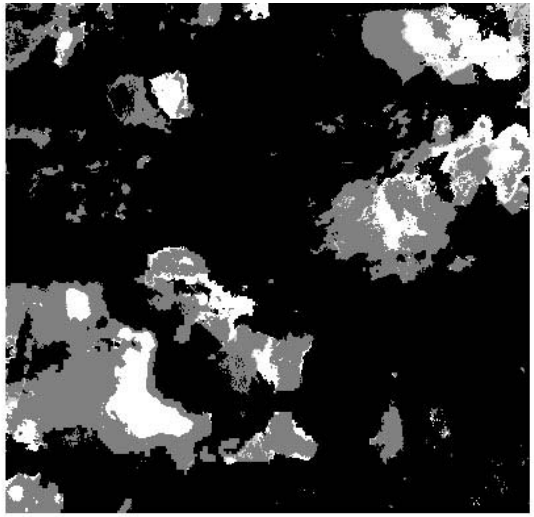
	Total Pixels Correct	Total Pixels False	Total Pixels Ambiguous	Misclassified Clouds	Misclassified Clear
ACCAp1	78.8%	6.0%	15.2%	10.4%	3.6%
See5-rB2	89.0%	8.0%	3.0%	12.3%	5.5%
AT-ACCA	74.6%	7.4%	18.0%	10.3%	5.6%
kmean	48.0%	7.7%	44.3%	19.0%	1.3%
$g4_{0,8}$	69.6%	5.4%	25.0%	4.5%	5.9%
$ATg4_{0,8}$	82.3%	8.4%	9.3%	11.5%	6.6%

These algorithms are being proposed for implementation.

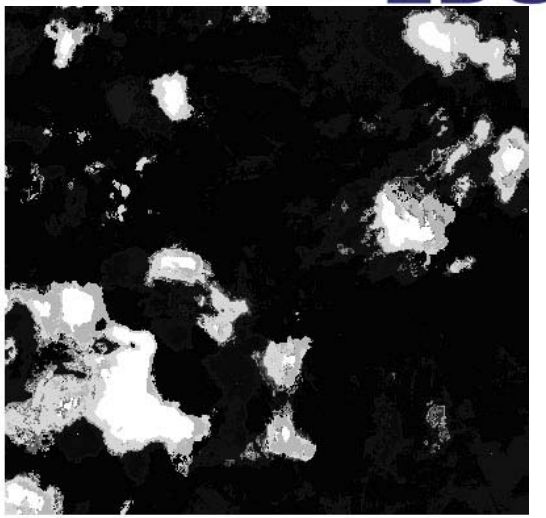
Confidence mask examples



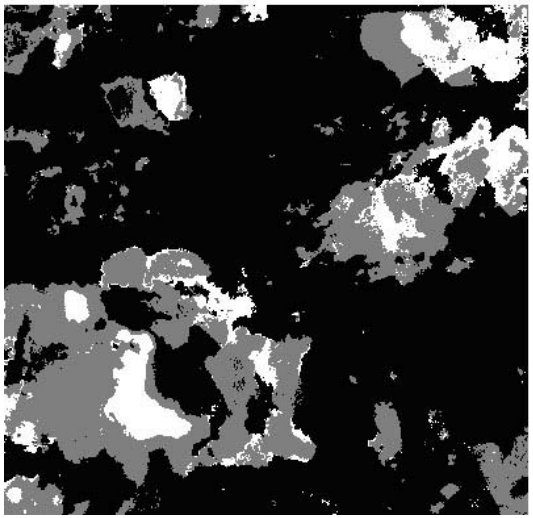
5,4,3 Image



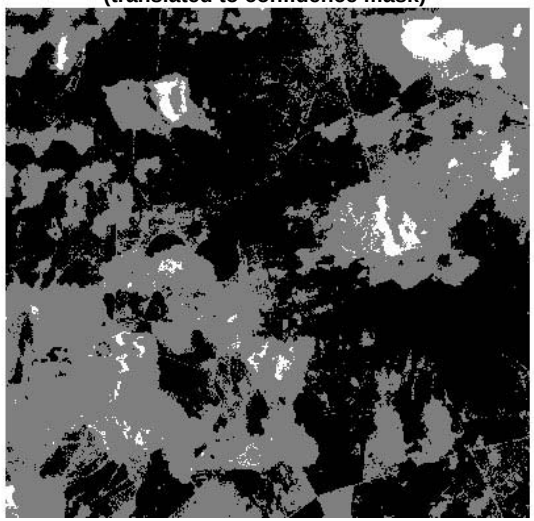
ACCA p1
(translated to confidence mask)



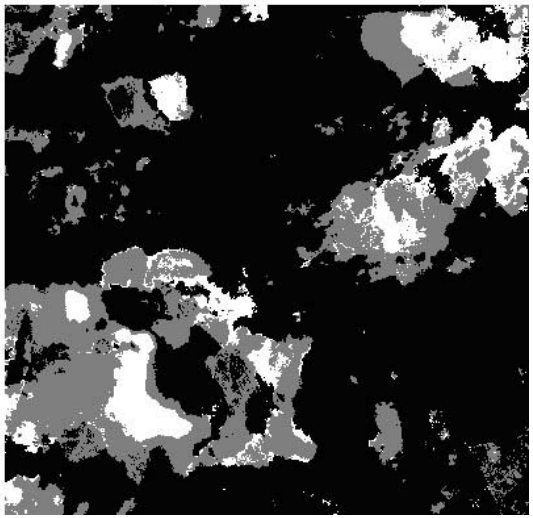
See5 rB2



AT-ACCA

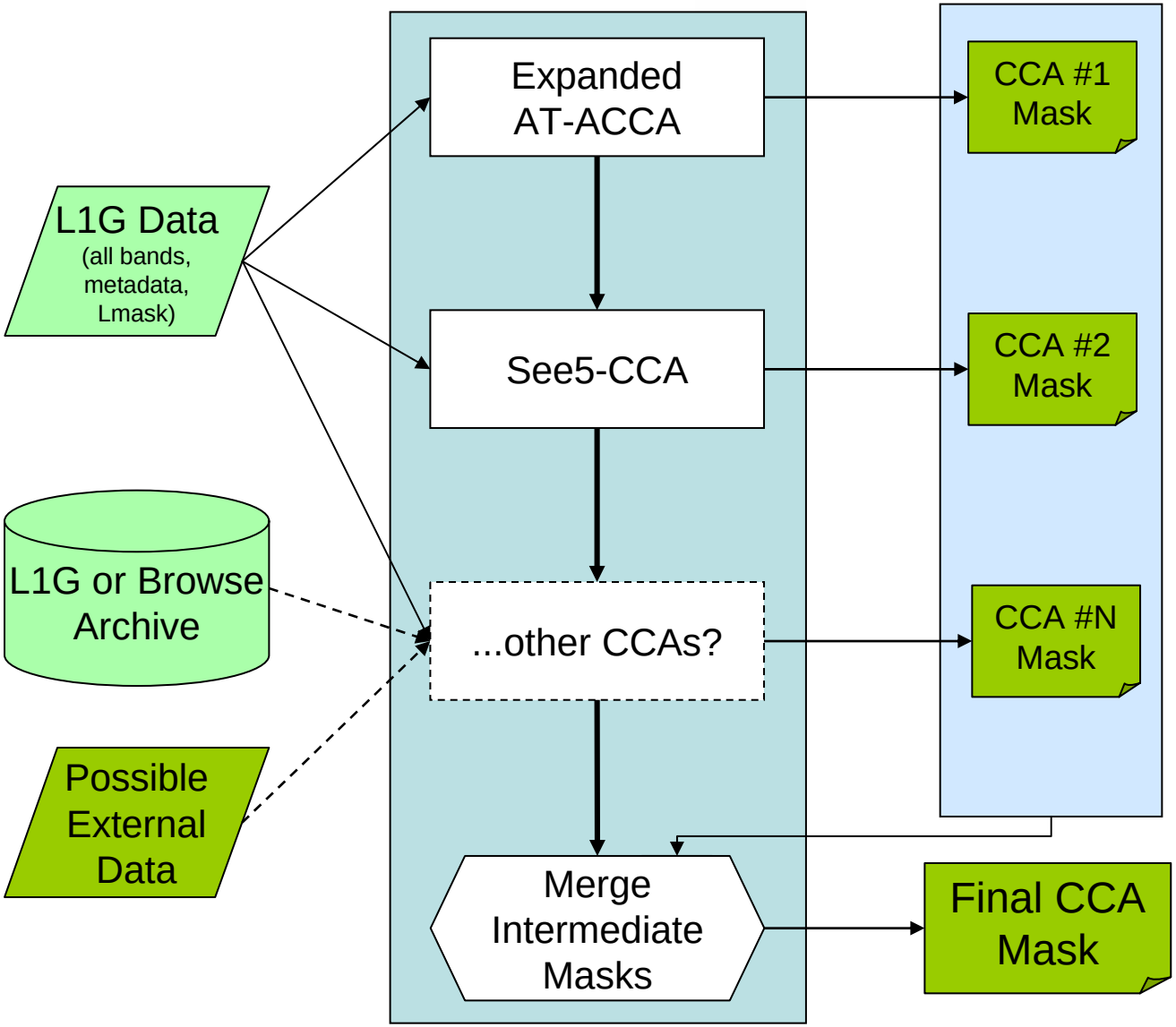


$g4_{0,8}$



Expanded AT-ACCA

Proposed System



AT-ACCA and See-5 CCA are the only algorithms being implemented at this time.

(Cirrus and $g4_{0,8}$ are being implemented as part of Expanded AT-ACCA.)

Temporal CCA may be added after OIVP.

Other CCAs can still be added, within project constraints.

Future Plans for CCA

- **Shadows**
  - All attempts at spectral shadow masking have failed. Shadow masking must involve cloud geometry.
  - Known methods do not comply with latency requirements.
  - Shadow masking will not be implemented before launch, but is high priority for after launch.

- **Cirrus**
  - Preliminary study using Hyperion data suggests that cirrus can be detected with a simple threshold test on the OLI cirrus band. (With some false positives over high altitude terrain.)
  - Further study will be performed once OLI spectral characteristics are known.

- **Other Classification Masks**
  - Would be useful to have masks of water, snow/ice, terrain type, etc. as input to CCA.
  - Not currently within scope, so unlikely to be implemented before launch.

- **Intermediate Mask Merge**
  - Currently testing a weighted vote methodology.
  - Further study ongoing.

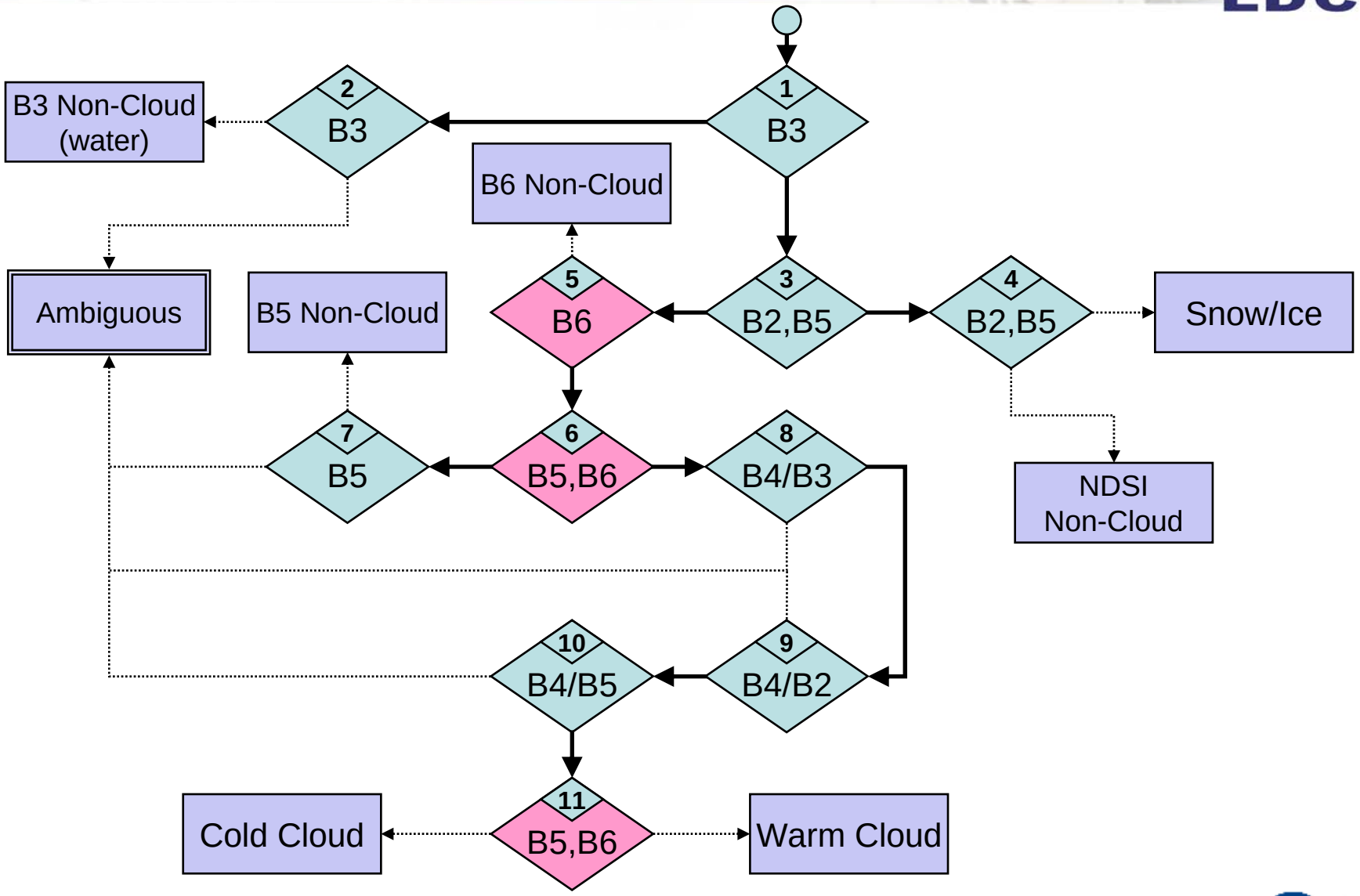
- **Expansion of Study**
  - Utilize data from Hyperion, ALI, and other satellites, converted to OLI-like bands.
  - Involves creation of another set of training and validation scenes *with manual masks*.
  - May not have the resources to do this, but will if possible.

- **Journal Article**

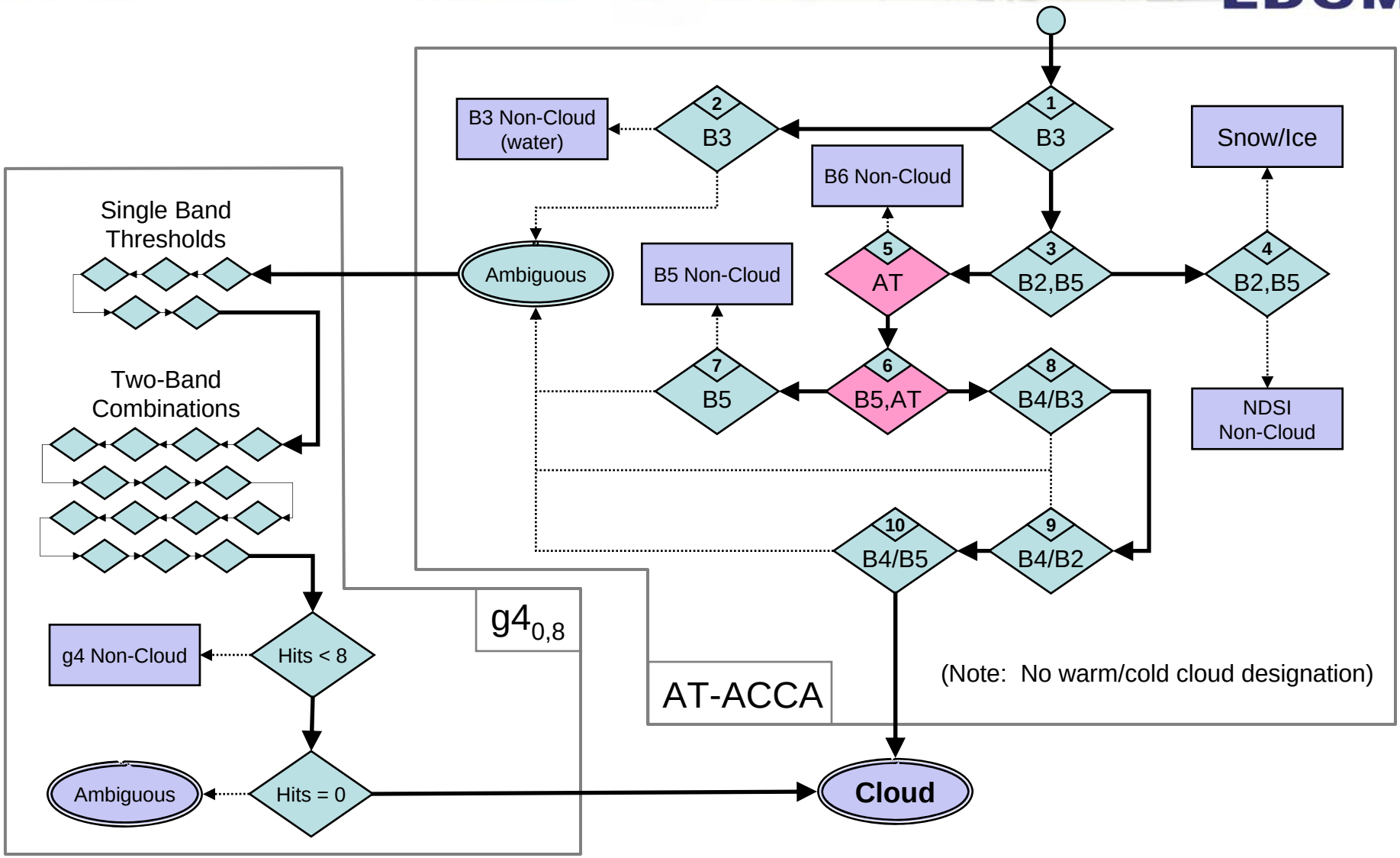
End of planned slides

(after this are supporting slides)

ACCA Phase 1 Flowchart



AT-ACCA w/$g4_{0,8}$ Flowchart



LDCM Cloud Mask File

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** = 0-35% confidence
- **10** = 36-64% confidence
- **11** = 65-100% 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 may change.

Water and snow masks are currently only being developed in the AT-ACCA algorithm, and they have no confidence values.

LDCM Intermediate Cloud Mask Files

Bit	Description
0	Class
1	
2	Class Confidence
3	
4	Cirrus Confidence
5	
6	Cloud Confidence
7	

Two-bit class flags:

- **00** = Fill
- **01** = Undetermined
- **10** = Water
- **11** = Snow/Ice

Two-bit confidence levels:

- **00** = None or Unset.
- **01** = 0-35% confidence
- **10** = 36-64% confidence
- **11** = 65-100% confidence

ACCA phase 1 Results

103 Validation scenes (same validation data as See5 rB2 results)

Total number of validation pixels: 3,945,094,408

Scene level validation results

% error	# of scenes	% of scenes	cum. %
0 - 5	55	53%	53%
5 - 10	22	21%	75%
10 - 15	12	12%	86%
15 - 20	5	5%	91%
20 - 25	4	4%	95%
25 - 30	0	0%	95%
30 - 35	0	0%	95%
35 - 40	1	1%	96%
40 - 45	0	0%	96%
45 - 50	0	0%	96%
50 - 55	0	0%	96%
55 - 60	1	1%	97%
60 - 65	0	0%	97%
65 - 70	0	0%	97%
70 - 75	0	0%	97%
75 - 80	1	1%	98%
80 - 85	2	2%	100%

Pixel level validation results

Class	Clear	Shadow	Thick	Thin	All Clouds	All Clear
Non-Cloud	80.0%	74.4%	4.4%	22.3%	10.4%	80.0%
Ambig.	16.4%	22.0%	6.8%	25.0%	12.9%	16.4%
Cloud	3.6%	3.6%	88.8%	52.6%	76.7%	3.6%

Total Pixels Correct: 78.8%

Total Pixels False: 6.0%

Total Ambiguous: 15.2%

Provisional clouds are included in the cloud results above.
Separate results for provisional clouds are:

Class	Clear	Shadow	Thick	Thin	All Clouds	All Clear
Provisional Cloud	1.1%	1.1%	7.2%	11.2%	8.5%	1.1%

Total Pixels classed as Provisional Clouds: 3.8%

AT-ACCA Results

103 Validation scenes (same validation data as See5 rB2 results)

Total number of validation pixels: 3,880,460,393

Scene level validation results

% error	# of scenes	% of scenes	cum. %
0 - 5	51	50%	50%
5 - 10	16	16%	65%
10 - 15	12	12%	77%
15 - 20	8	8%	84%
20 - 25	7	7%	91%
25 - 30	2	2%	93%
30 - 35	0	0%	93%
35 - 40	1	1%	94%
40 - 45	2	2%	96%
45 - 50	0	0%	96%
50 - 55	0	0%	96%
55 - 60	1	1%	97%
60 - 65	0	0%	97%
65 - 70	0	0%	97%
70 - 75	1	1%	98%
75 - 80	0	0%	98%
80 - 85	1	1%	99%
85 - 90	1	1%	100%

Pixel level validation results

Class	Clear	Shadow	Thick	Thin	All Clouds	All Clear
Non-Cloud	74.7%	68.5%	4.4%	22.2%	10.3%	74.6%
Ambig.	19.8%	24.4%	8.3%	27.9%	14.9%	19.8%
Cloud	5.6%	7.1%	87.3%	49.9%	74.8%	5.6%

Total Pixels Correct: 74.6%

Total Pixels False: 7.4%

Total Ambiguous: 18.0%

(This version of AT-ACCA does not distinguish between Clouds and Provisional clouds.)

$g4_{0,8}$ Results

103 Validation scenes (same validation data as See5 rB2 results)

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

Scene level validation results

% error	# of scenes	% of scenes	cum. %
0 - 5	49	48%	48%
5 - 10	12	12%	59%
10 - 15	16	16%	75%
15 - 20	8	8%	83%
20 - 25	2	2%	84%
25 - 30	3	3%	87%
30 - 35	3	3%	90%
35 - 40	0	0%	90%
40 - 45	1	1%	91%
45 - 50	1	1%	92%
50 - 55	1	1%	93%
55 - 60	2	2%	95%
60 - 65	1	1%	96%
65 - 70	1	1%	97%
70 - 75	1	1%	98%
75 - 80	1	1%	99%
80 - 85	1	1%	100%

Pixel level validation results

Class	Clear	Shadow	Thick	Thin	All Clouds	All Clear
Non-Cloud	71.1%	49.1%	1.3%	10.7%	4.4%	70.9%
Ambig.	23.0%	48.6%	23.9%	37.0%	28.3%	23.2%
Cloud	5.9%	2.3%	74.8%	52.3%	67.3%	5.9%

Total Pixels Correct: 69.6%

Total Pixels False: 5.4%

Total Ambiguous: 25.0%

$ATg4_{0,8}$ Results

103 Validation scenes (same validation data as See5 rB2 results)

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

Scene level validation results

% error	# of scenes	% of scenes	cum. %
0 - 5	60	58%	56%
5 - 10	17	17%	75%
10 - 15	12	12%	86%
15 - 20	5	5%	91%
20 - 25	3	3%	94%
25 - 30	1	1%	95%
30 - 35	0	0%	95%
35 - 40	1	1%	96%
40 - 45	0	0%	96%
45 - 50	0	0%	96%
50 - 55	0	0%	96%
55 - 60	1	1%	97%
60 - 65	0	0%	97%
65 - 70	1	1%	98%
70 - 75	1	1%	99%
75 - 80	1	1%	100%

Pixel level validation results

Class	Clear	Shadow	Thick	Thin	All Clouds	All Clear
Non-Cloud	83.4%	75.4%	4.6%	25.2%	11.5%	83.4%
Ambig.	10.0%	16.9%	4.5%	14.9%	8.1%	10.0%
Cloud	6.6%	7.8%	90.8%	59.9%	80.4%	6.6%

Total Pixels Correct: 82.3%

Total Pixels False: 8.4%

Total Ambiguous: 9.3%

Comparison of all CCA algorithms under review at EDC

	Total Pixels Correct	Total Pixels False	Total Pixels Ambiguous	Misclassified Clouds	Misclassified Clear
See5-rB2	89.0%	8.0%	3.0%	12.3%	5.5%
ACCAp1	78.8%	6.0%	15.2%	10.4%	3.6%
AT-ACCA	74.6%	7.4%	18.0%	10.3%	5.6%
kmean	48.0%	7.7%	44.3%	19.0%	1.3%
$g4_5$	53.4%	3.4%	43.2%	9.4%	n/a
$g4_{0,8}$	69.6%	5.4%	25.0%	4.5%	5.9%
$g4_{0,8}AT$	81.5%	9.1%	9.4%	8.9%	9.3%
$ATg4_{0,8}$	82.3%	8.4%	9.3%	11.5%	6.6%
$ATg4_{0,8}km$	87.7%	10.4%	1.9%	16.9%	6.7%
$g4_5AT$	84.5%	7.6%	7.9%	14.5%	3.7%
$g4_5kmAT$	86.4%	11.6%	2.0%	28.3%	2.1%
ATkm	84.7%	10.9%	4.3%	20.2%	5.6%

These algorithms are being proposed for implementation.

AT equations

All AT algorithms in this presentation use the 'ATv4 Cubist r1' formulation, as presented in the June 2008 LST.

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 * NDVI	- 17.2 * NDxI
	+ 5.1 * (B4/B3)	- 3.7 * (B4/B2)
	+ 302.2927	

where cS = cos(solar elevation angle)
NDVI = (B4-B3)/(B4+B3)
NDxI = (B1-B7)/(B1+B7)

g4 Parameter study example -- Combinations of B3 & B4

				Type I Error	Type II Error	Accuracy	
Parameter Combination	Low Threshold	High Threshold	Likelihood Ratio	Mis-classified clouds	Mis-classified clear	1- Average Error	Percentile
B4/CB3	1.812	9.058	1.223	98.451%	1.144%	50.203%	98.731%
NDC43	-0.288	0.287	8.372	5.061%	57.543%	68.698%	30.858%
ND43	-0.064	0.514	12.550	2.580%	65.000%	66.210%	24.944%
CB4/B3	0.480	1.866	14.534	2.117%	66.286%	65.799%	23.913%
B4/B3	0.763	2.870	19.149	1.469%	68.474%	65.029%	22.203%
B4/B3	1.035		0.729	36.778%	72.085%	45.569%	29.779%
ND4C3	0.000	0.707	12.458	1.494%	77.633%	60.437%	15.892%
B4/B3		2.397	7.310	2.552%	77.854%	59.797%	16.068%
worst non-thermal ACCA tests			*> 3.800*	*< 8.100%*	*< 79.900%*	*> 60.000%*	*> 15.800%*

Various combinations of B3 and B4 were examined. A set of thresholds for each combination was derived by an iterative maximum likelihood ratio test. The combinations were then excluded if they did not meet the criteria set by the worst non-thermal ACCA tests. The remaining combinations were ranked by Type II error, and the best (NDC43) was implemented in the g4 decision tree.

(Ranking by Likelihood ratio was attempted, but the results were inferior to the TII ranking method.)

This examination was performed for 284 combinations -- including all two-band combinations, the AT band, and the ACCA-based parameter (1-B5)*AT. A total of 22 threshold tests were implemented in this fashion.

$g4_{0,8}$ Threshold Tests

cS*B1	< 0.101		B5/nfac	< 0.064	
B2	< 0.144		B7/nfac	< 0.048	
B3	< 0.140		AT	> 301.8	
ND(B2,B1)	< -0.108	or > 0.049	ND(B2,B7)	< -0.021	or > 0.838
ND(B3,B1)	< -0.192	or > 0.029	ND(cS*B4,B3)	< -0.288	or > 0.287
ND(B4,B1)	< -0.140	or > 0.408	ND(cS*B3,B5)	< -0.490	or > 0.786
ND(B1,cS*B5)	< 0.024	or > 0.834	ND(B3,B7)	< -0.040	or > 0.842
ND(B1,B7)	< -0.056	or > 0.854	ND(cS*B4,B5)	< -0.280	or > 0.777
ND(B3,B2)	< -0.096	or > 0.046	ND(cS*B4,B7)	< -0.200	or > 0.819
ND(B4,B2)	< -0.070	or > 0.420	ND(cS*B5,B7)	< -0.210	or > 0.210
ND(B2,B5)	< -0.160	or > 0.801	(1-B5)*AT	> 262.3	

where cS = cos(solar elevation angle)

ND(x,y) = Normalized difference = $\frac{x-y}{x+y}$

nfac = Normalization factor = $\sqrt{\sum_n (Band_n)^2}$

Variant $g4_{0,8}$ --

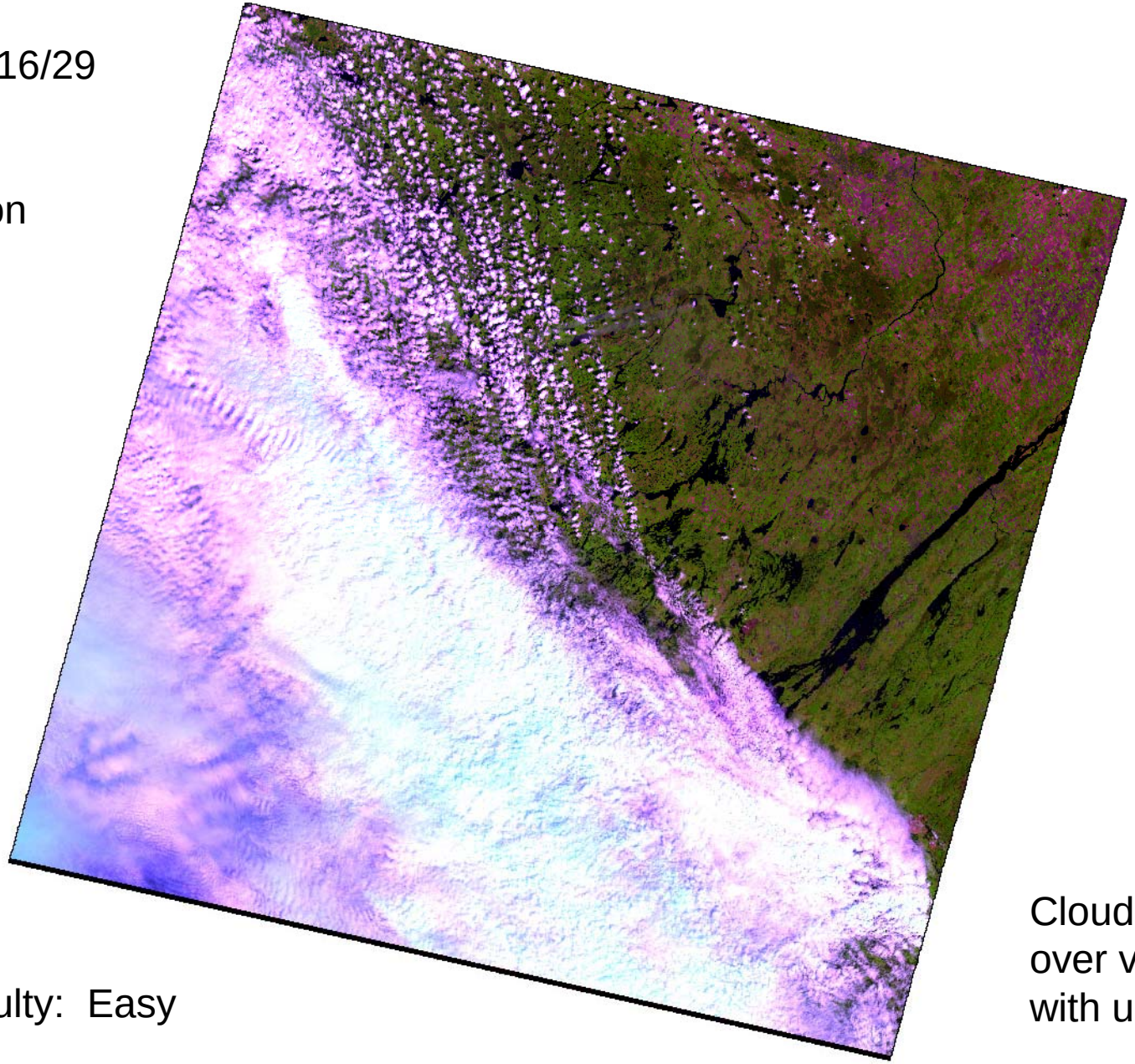
If 0 tests are true, Cloud.

If 8 or more tests are true, Clear.

All other outcomes are ambiguous.

p16 r29 -- Image

Path/Row 16/29
5/21/2001
5,4,3 band
combination



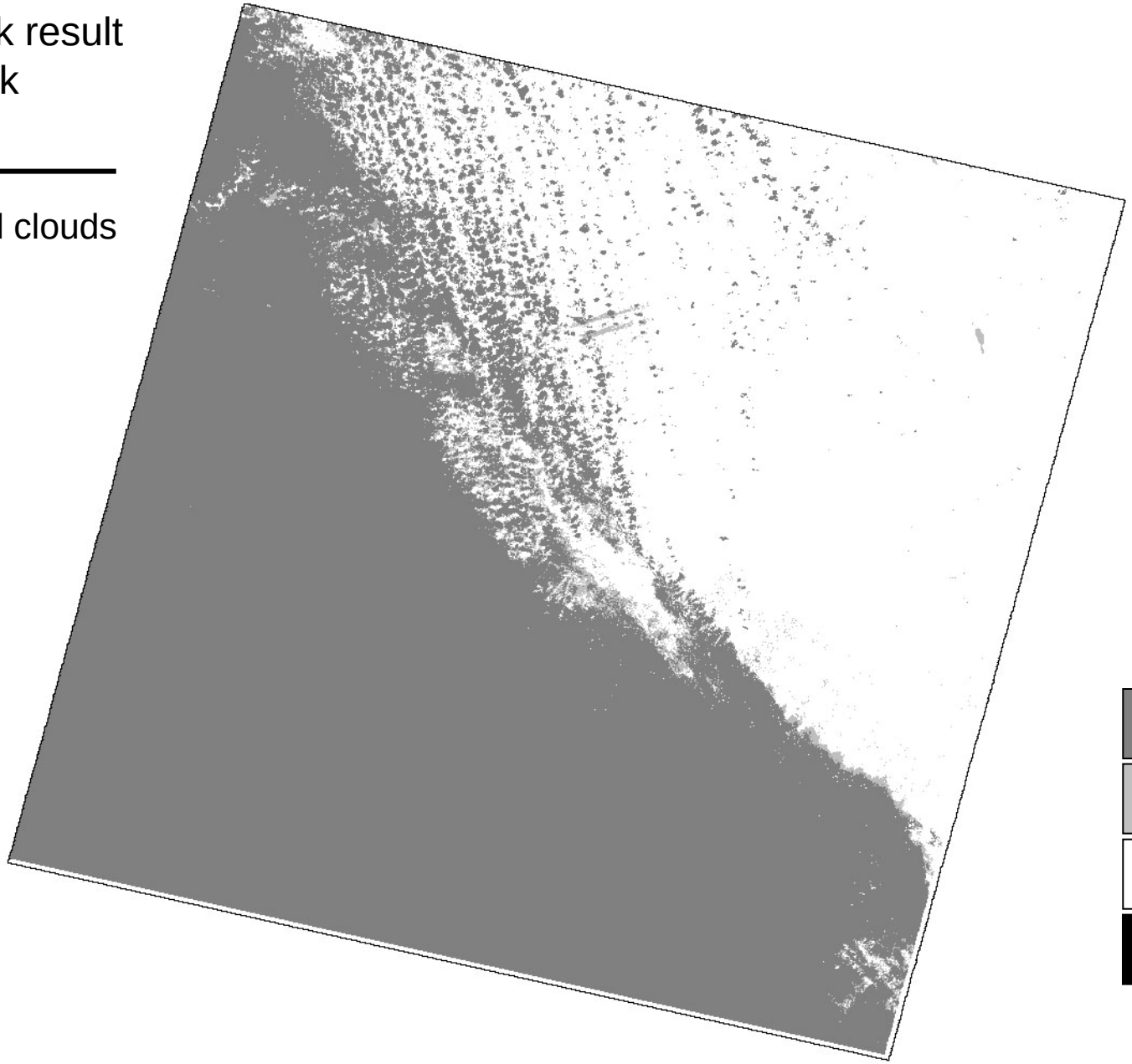
Cloud bank
over vegetation
with urban areas.

CCA Difficulty: Easy

p16 r29 -- Manual Mask

Manual mask result
58.3% thick
+ 3.1% thin

61.4% total clouds



Thick Clouds

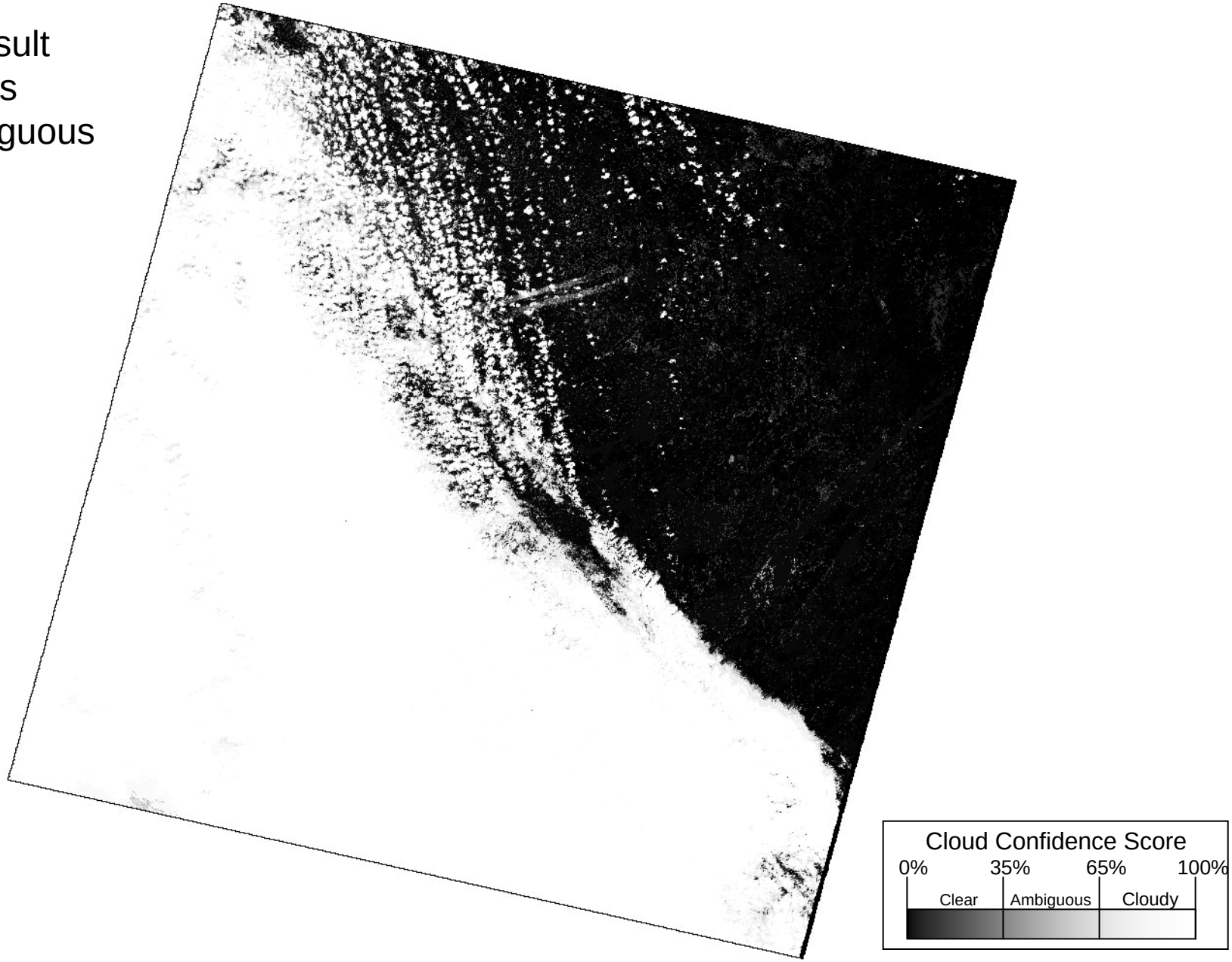
Thin Clouds

Clear

Shadows

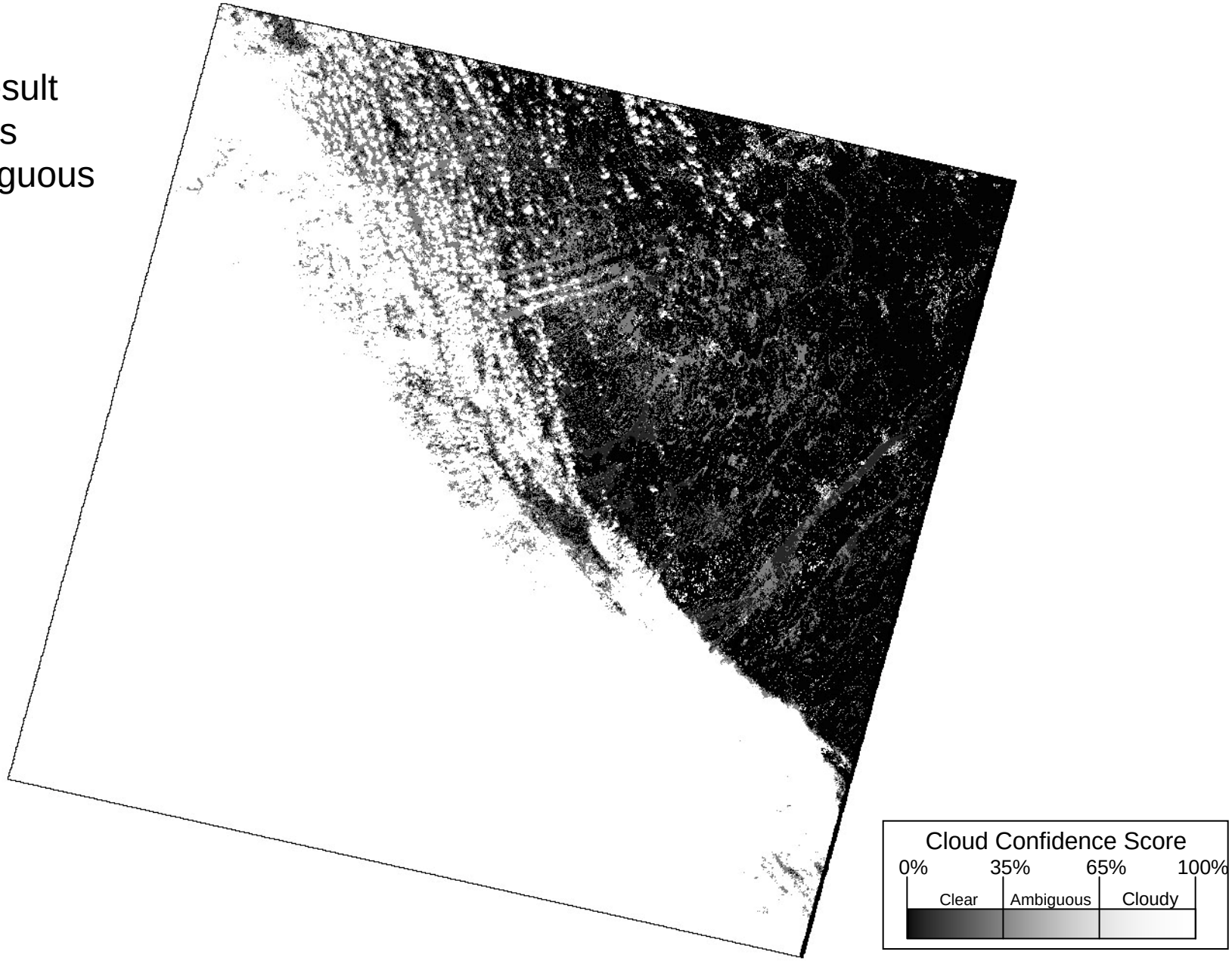
p16 r29 -- See5 rB2 confidence mask

See5 rB2 result
60.4% clouds
  1.6% ambiguous



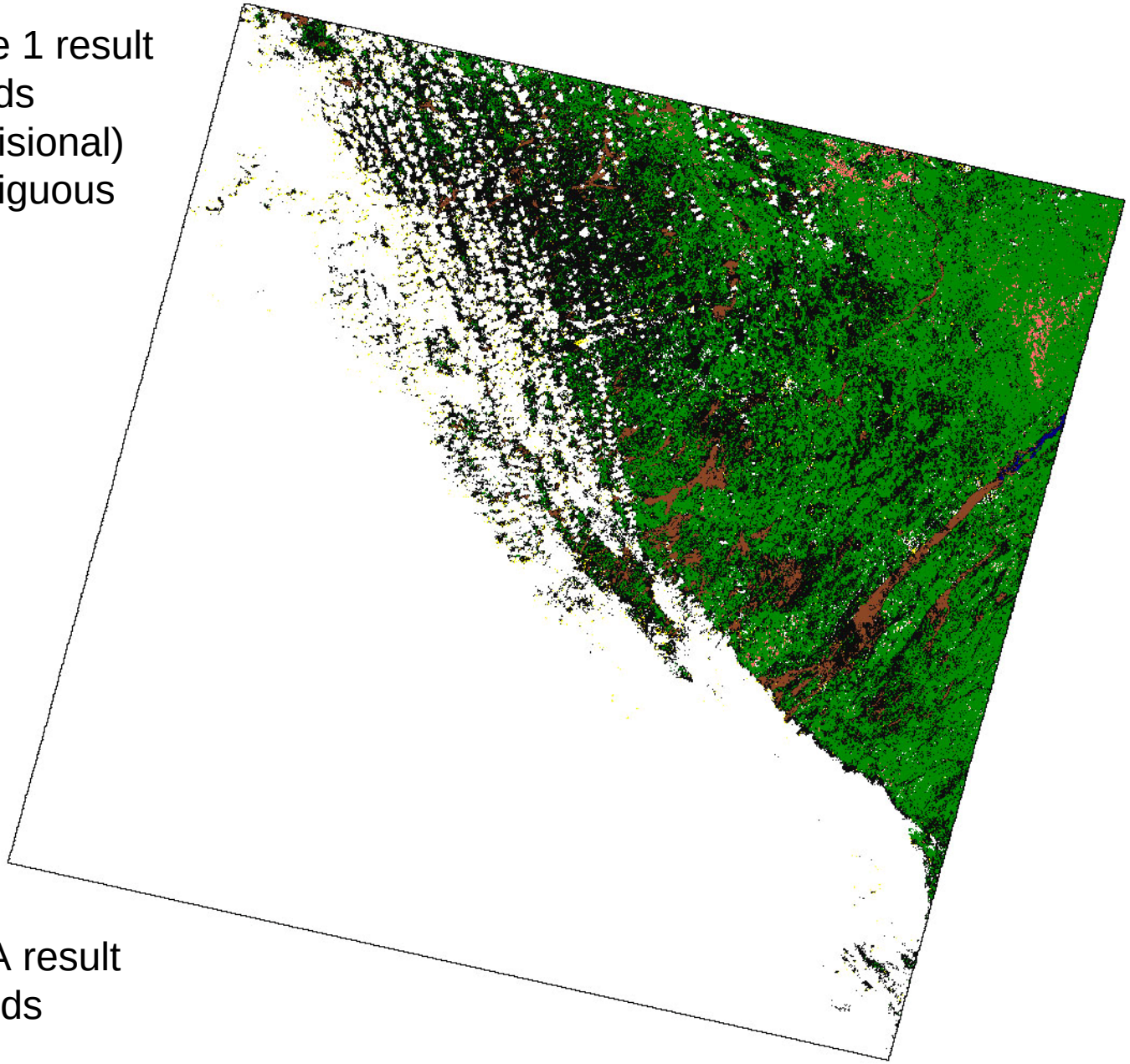
p16 r29 -- Expanded AT-ACCA confidence mask

Expanded
AT-ACCA result
62.0% clouds
10.6% ambiguous



p16 r29 -- ACCA phase 1 mask

ACCA phase 1 result
61.1% clouds
(0.1% provisional)
17.7% ambiguous

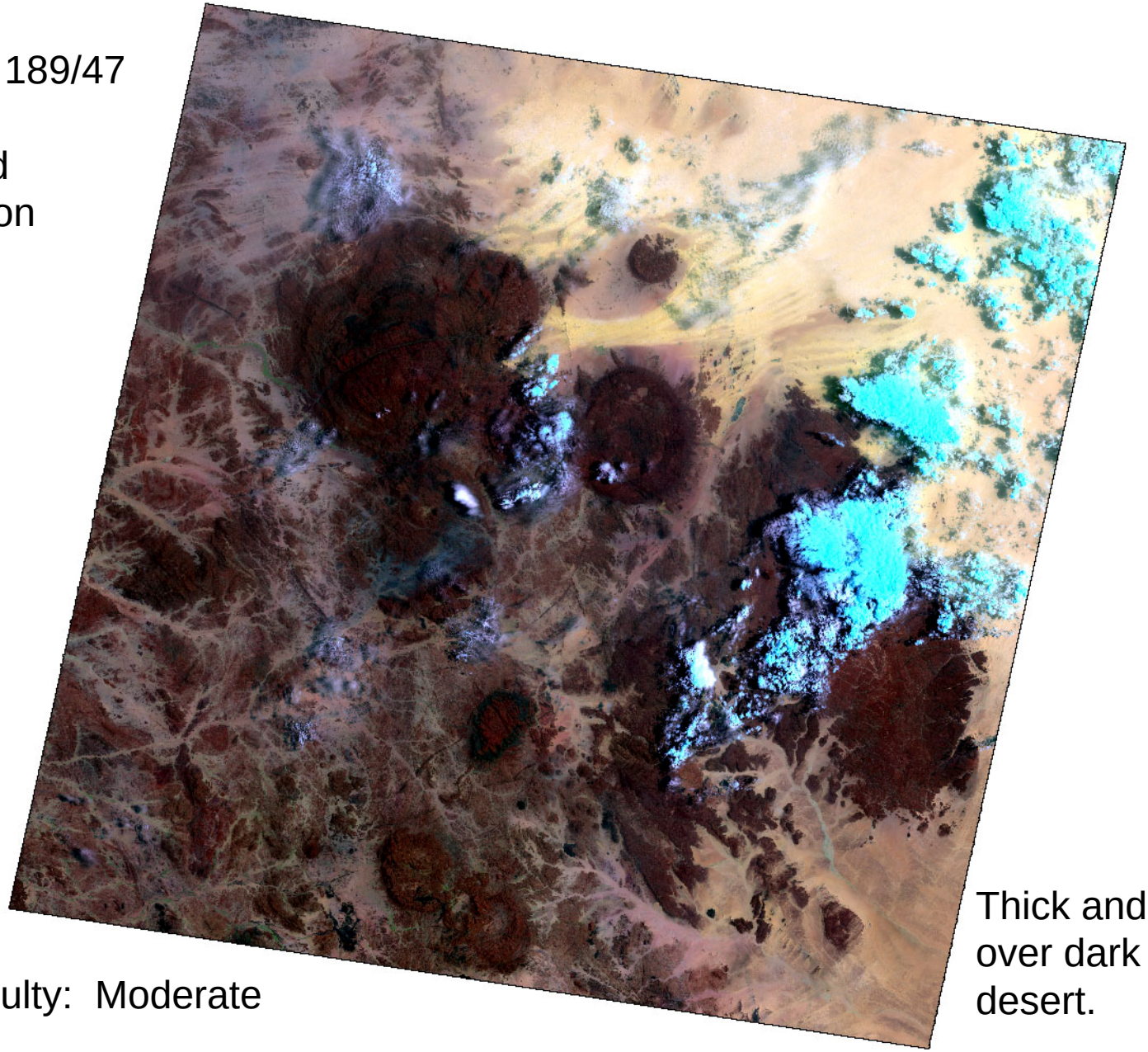


Final ACCA result
60.3% clouds

Cold Cloud
Provisional Cloud
Water
NDSI Noncloud
Snow/Ice
B6 Noncloud
B5 Noncloud
Ambiguous

p189 r47 -- Image

Path/Row 189/47
8/5/2001
7,4,2 band
combination



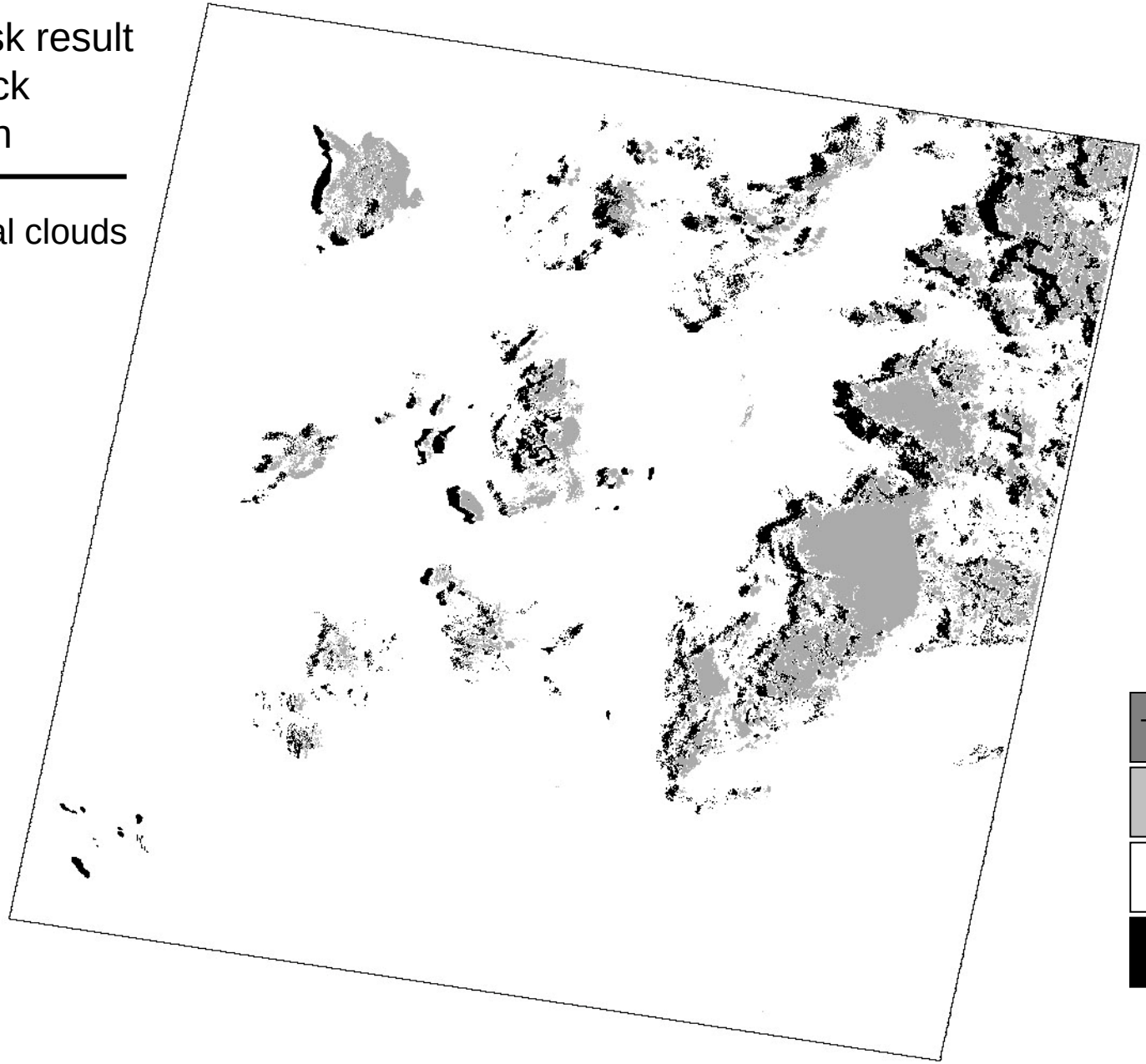
Thick and thin clouds over dark and bright desert.

CCA Difficulty: Moderate

p189 r47 -- Manual Mask

Manual mask result
6.3% thick
+ 5.7% thin

12.0% total clouds



Thick Clouds

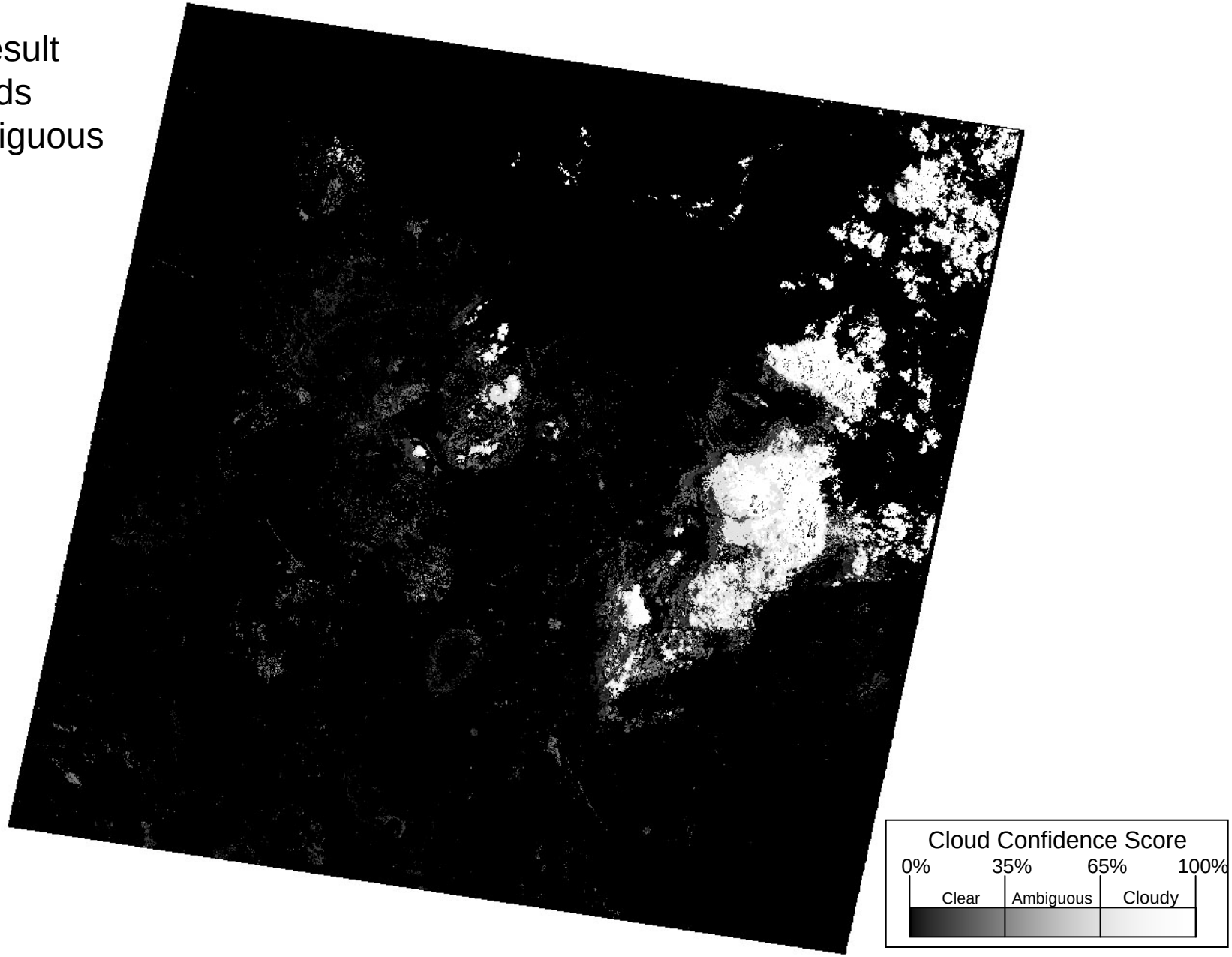
Thin Clouds

Clear

Shadows

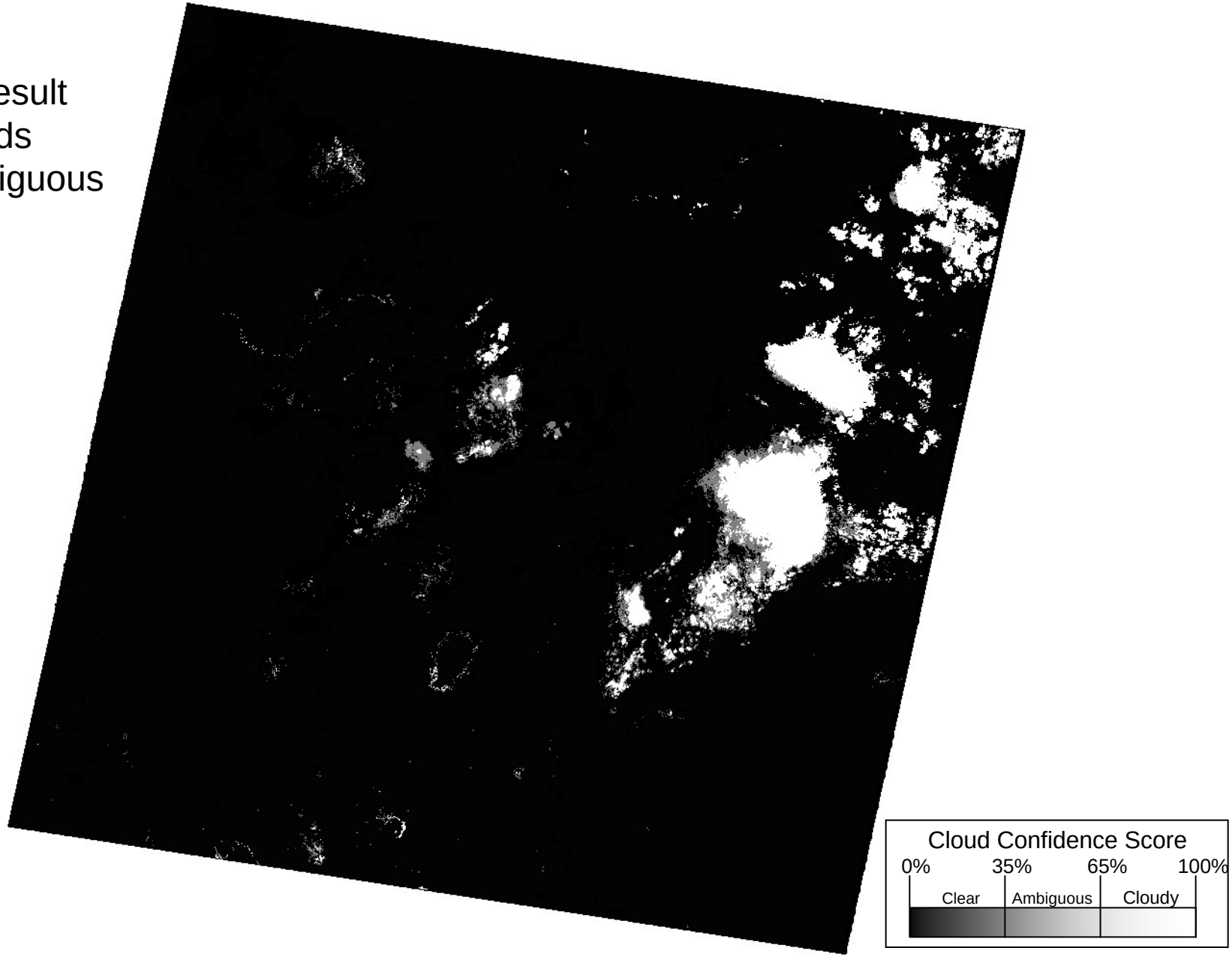
p189 r47 -- See5 rB2 confidence mask

See5 rB2 result
5.6% clouds
1.5% ambiguous



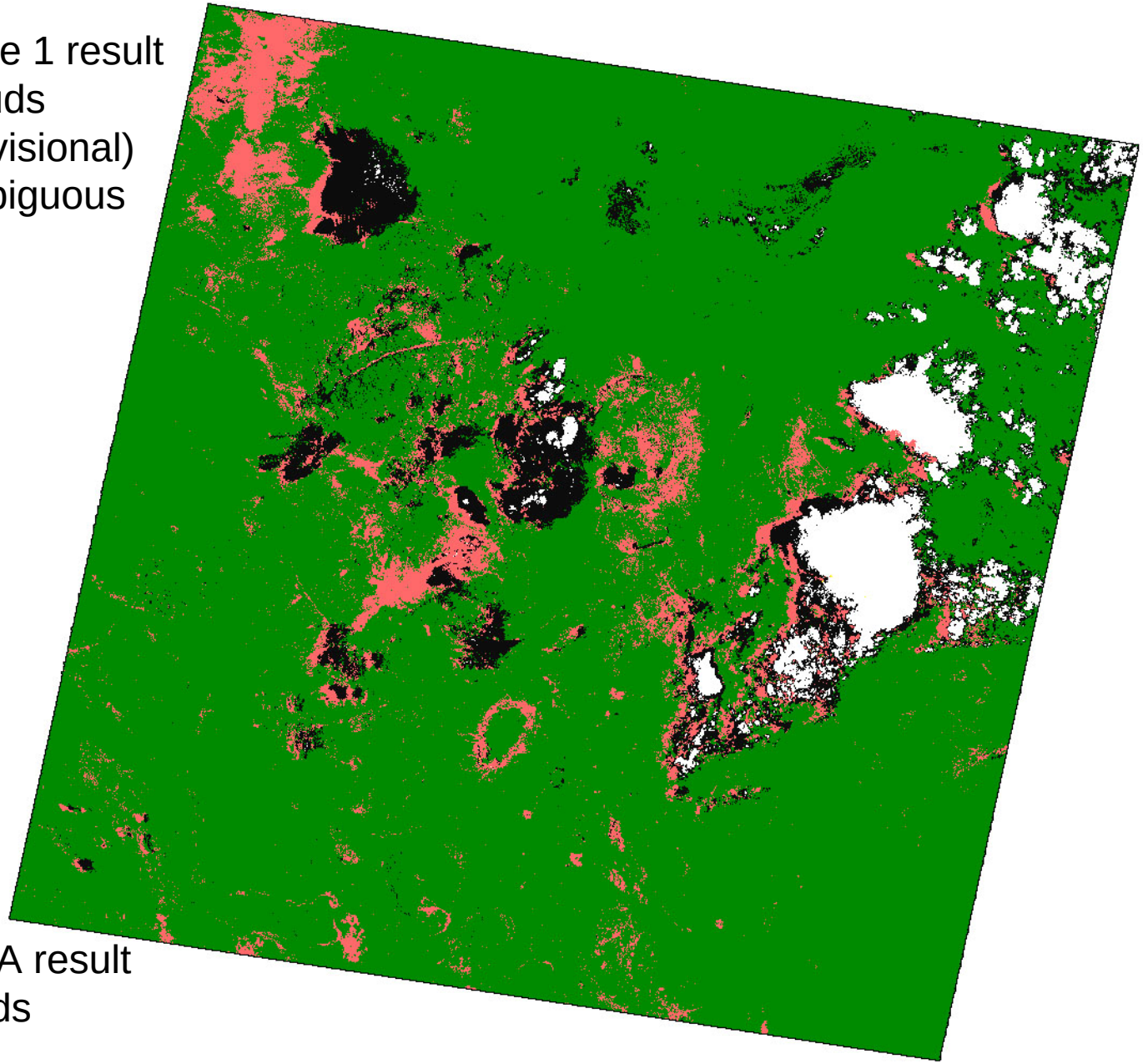
p189 r47 -- Expanded AT-ACCA confidence mask

Expanded
AT-ACCA result
  4.6% clouds
  1.6% ambiguous



p189 r47 -- ACCA phase 1 mask

ACCA phase 1 result
4.5% clouds
(0.0% provisional)
7.4% ambiguous

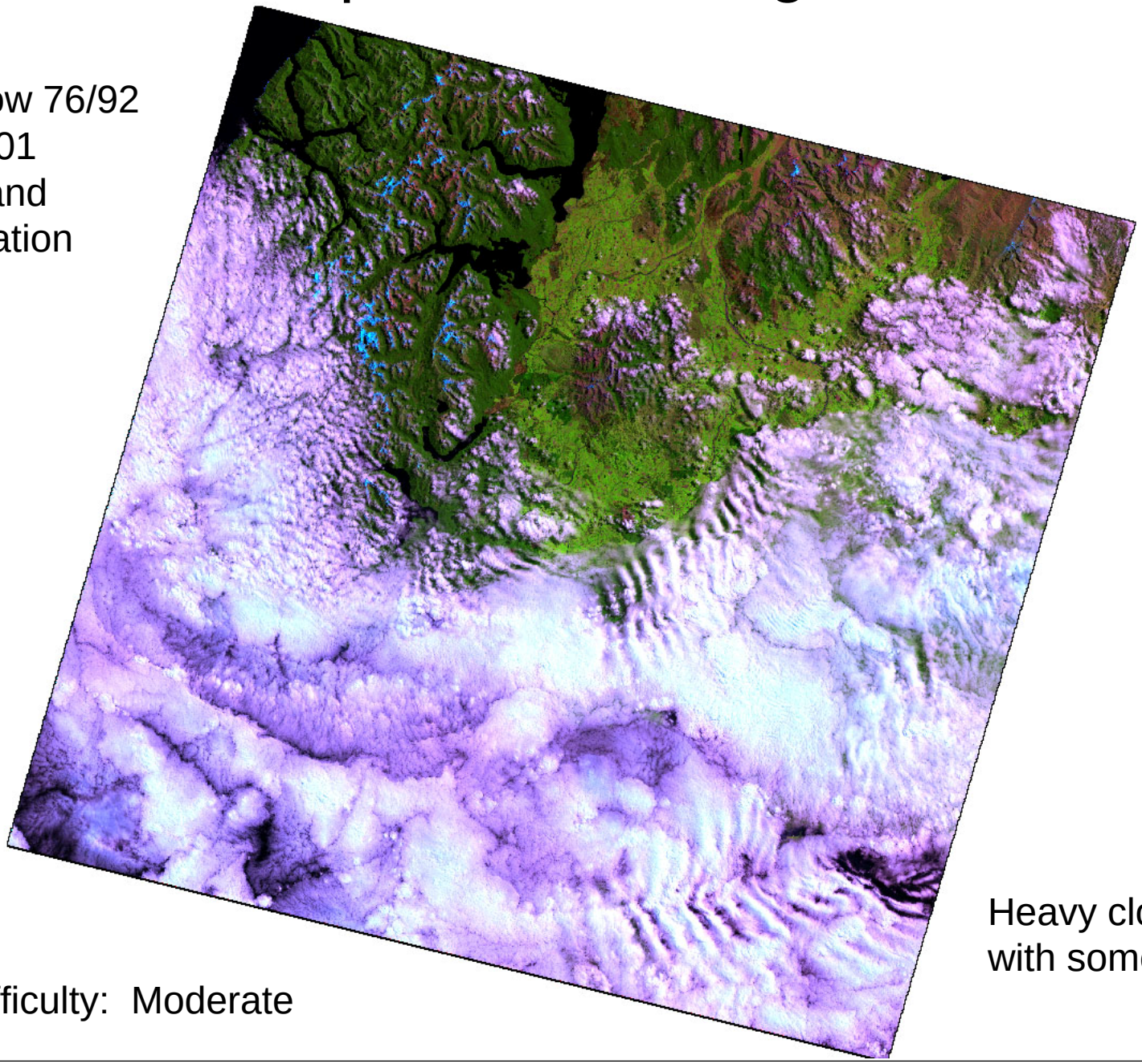


Final ACCA result
4.6% clouds

Cold Cloud
Provisional Cloud
Water
NDSI Noncloud
Snow/Ice
B6 Noncloud
B5 Noncloud
Ambiguous

p76 r92 -- Image

Path/Row 76/92
1/17/2001
5,4,3 band
combination



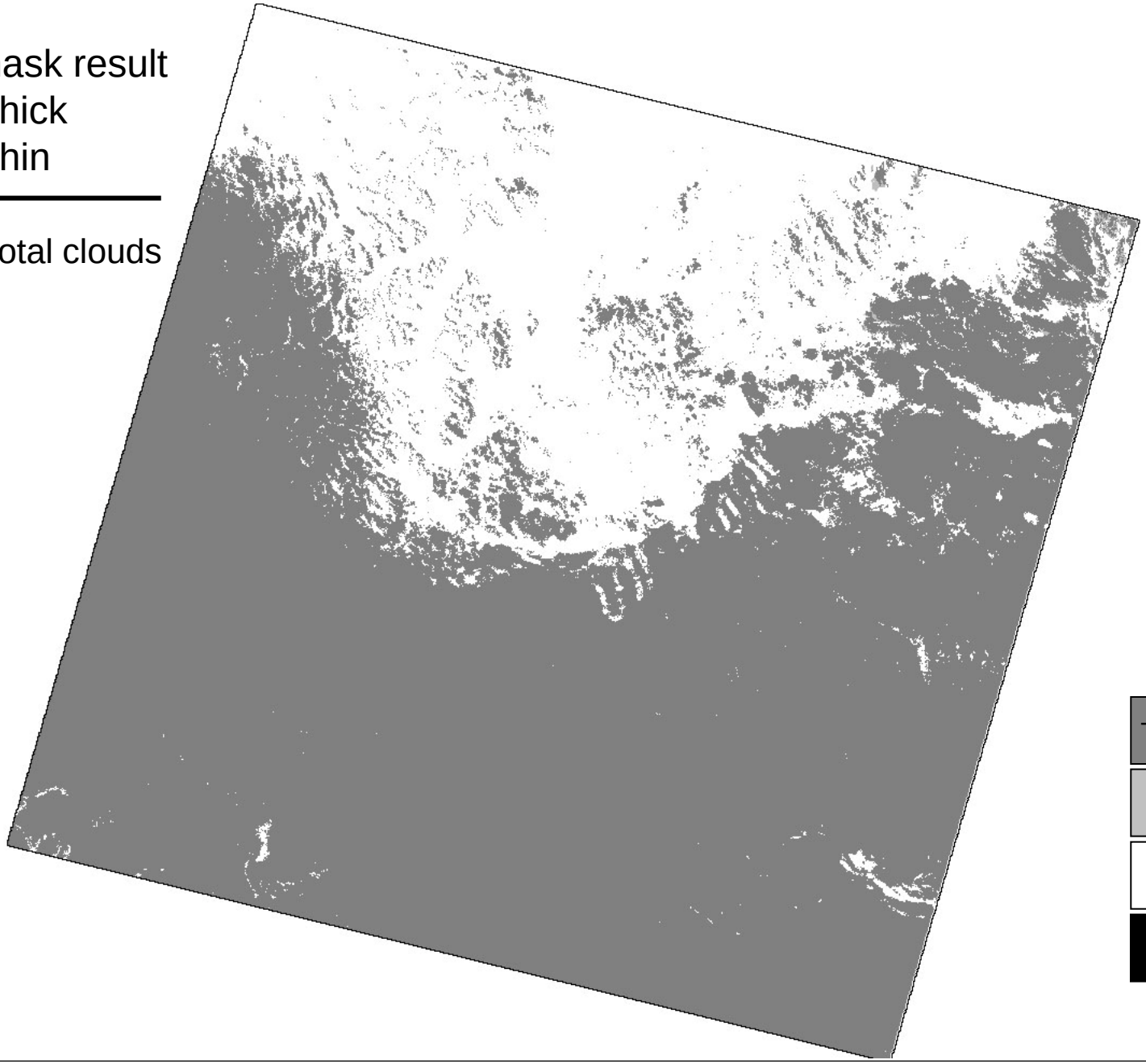
Heavy cloud cover,
with some snow.

CCA Difficulty: Moderate

p76 r92 -- Manual Mask

Manual mask result
68.1% thick
+ 0.2% thin

68.3% total clouds



Thick Clouds

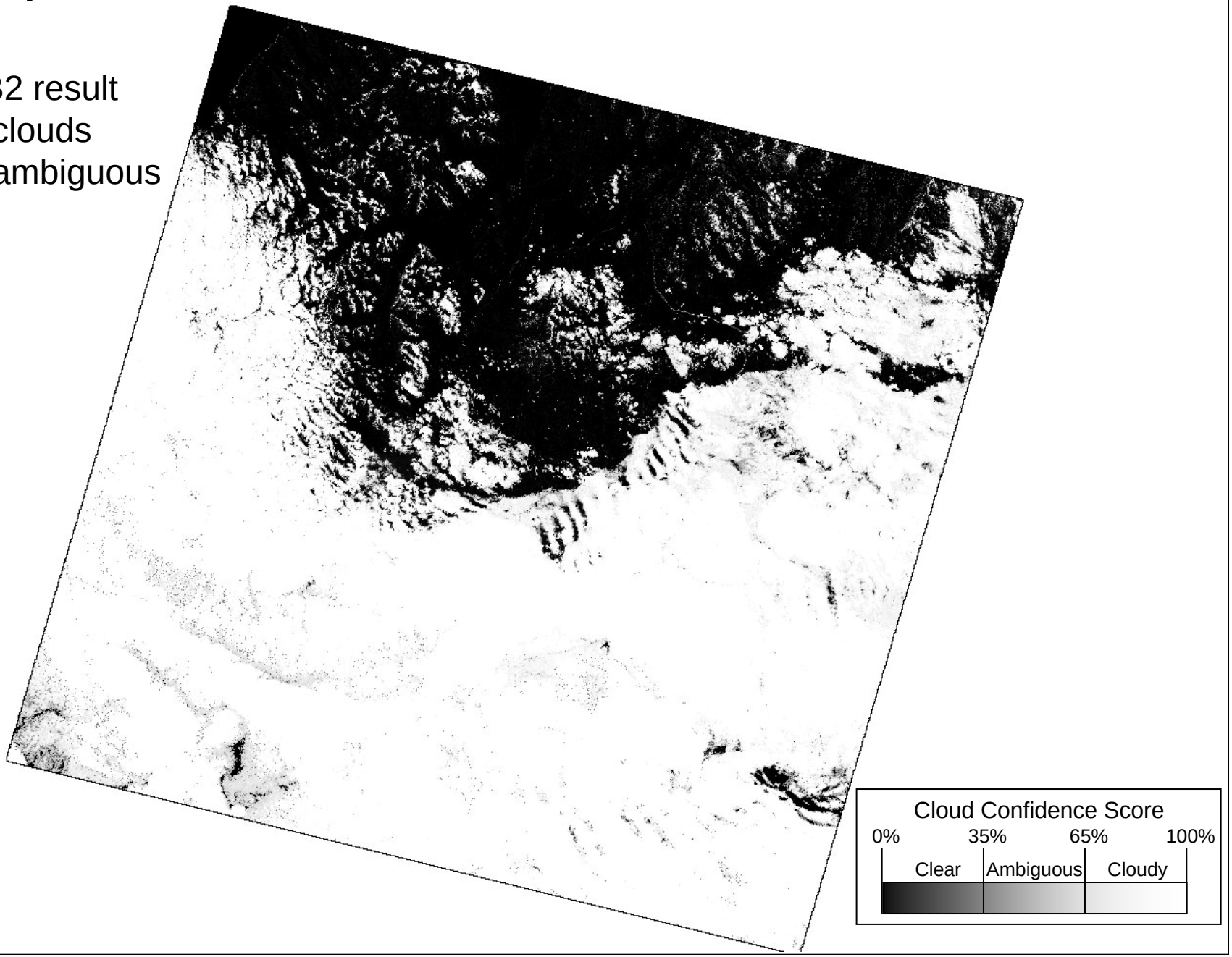
Thin Clouds

Clear

Shadows

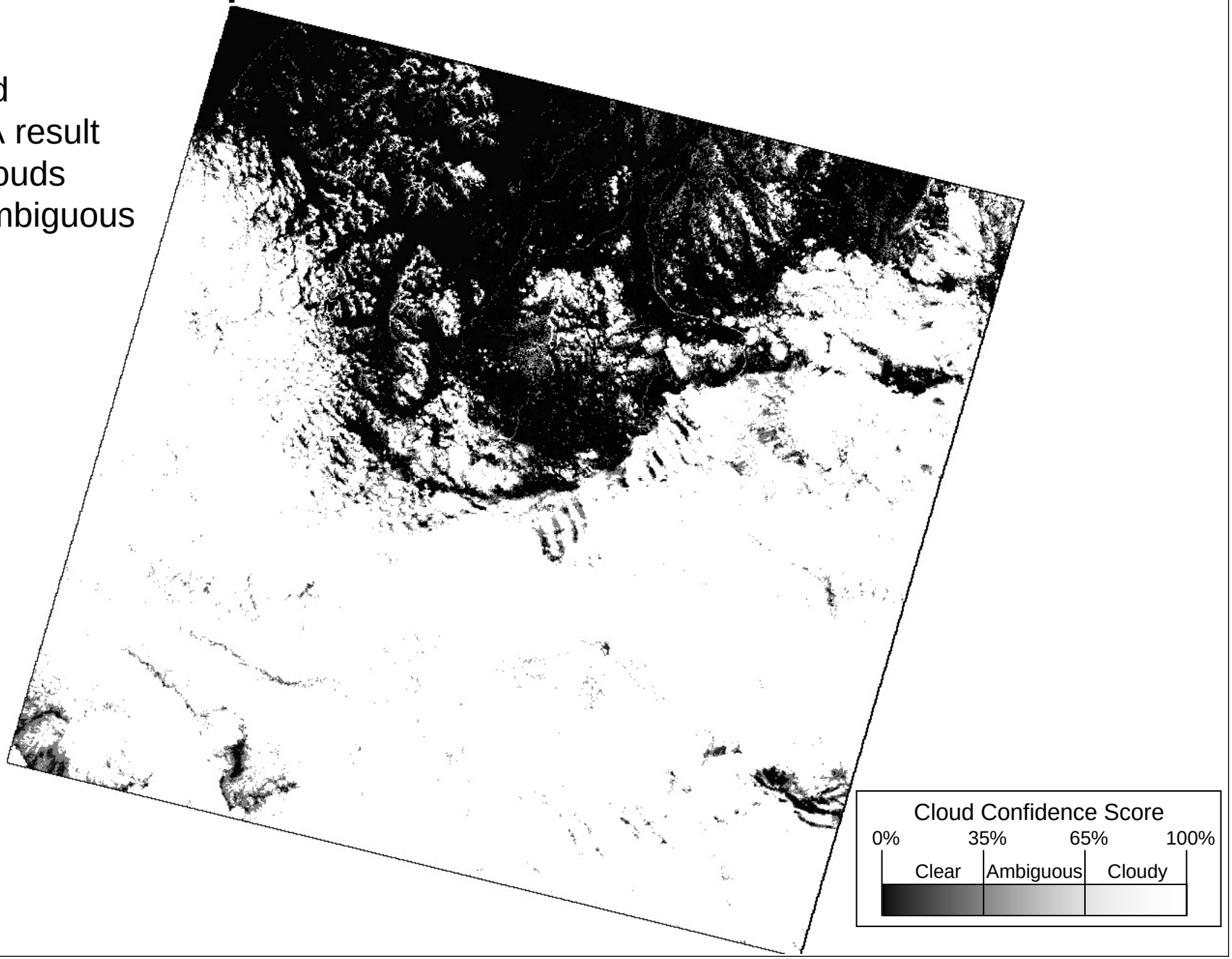
p76 r92 -- See5 rB2 confidence mask

See5 rB2 result
68.5% clouds
1.5% ambiguous



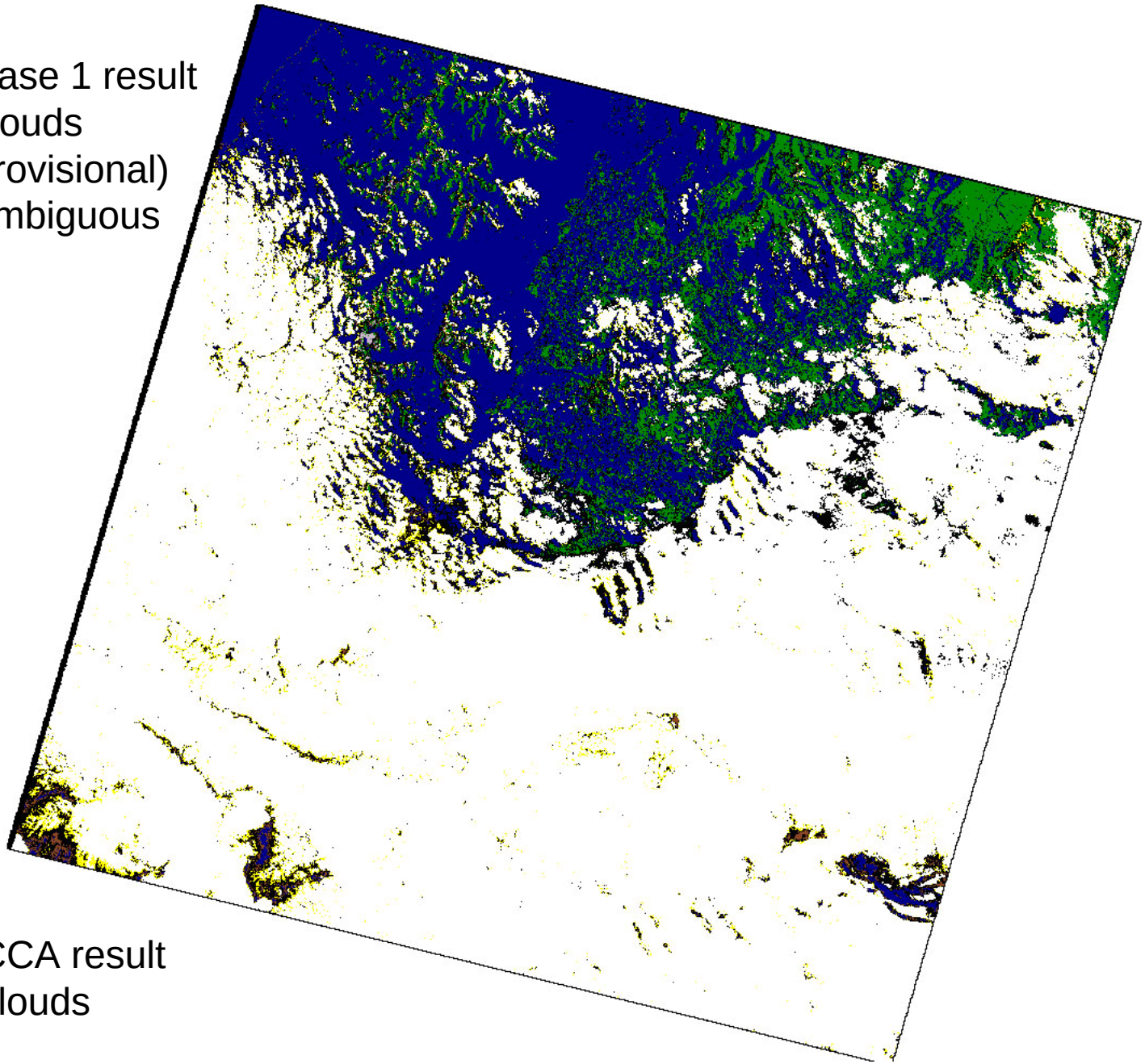
p76 r92 -- Expanded AT-ACCA confidence mask

Expanded
AT-ACCA result
68.5% clouds
5.5% ambiguous



p76 r92 -- ACCA phase 1 mask

ACCA phase 1 result
67.8% clouds
(1.6% provisional)
9.6% ambiguous

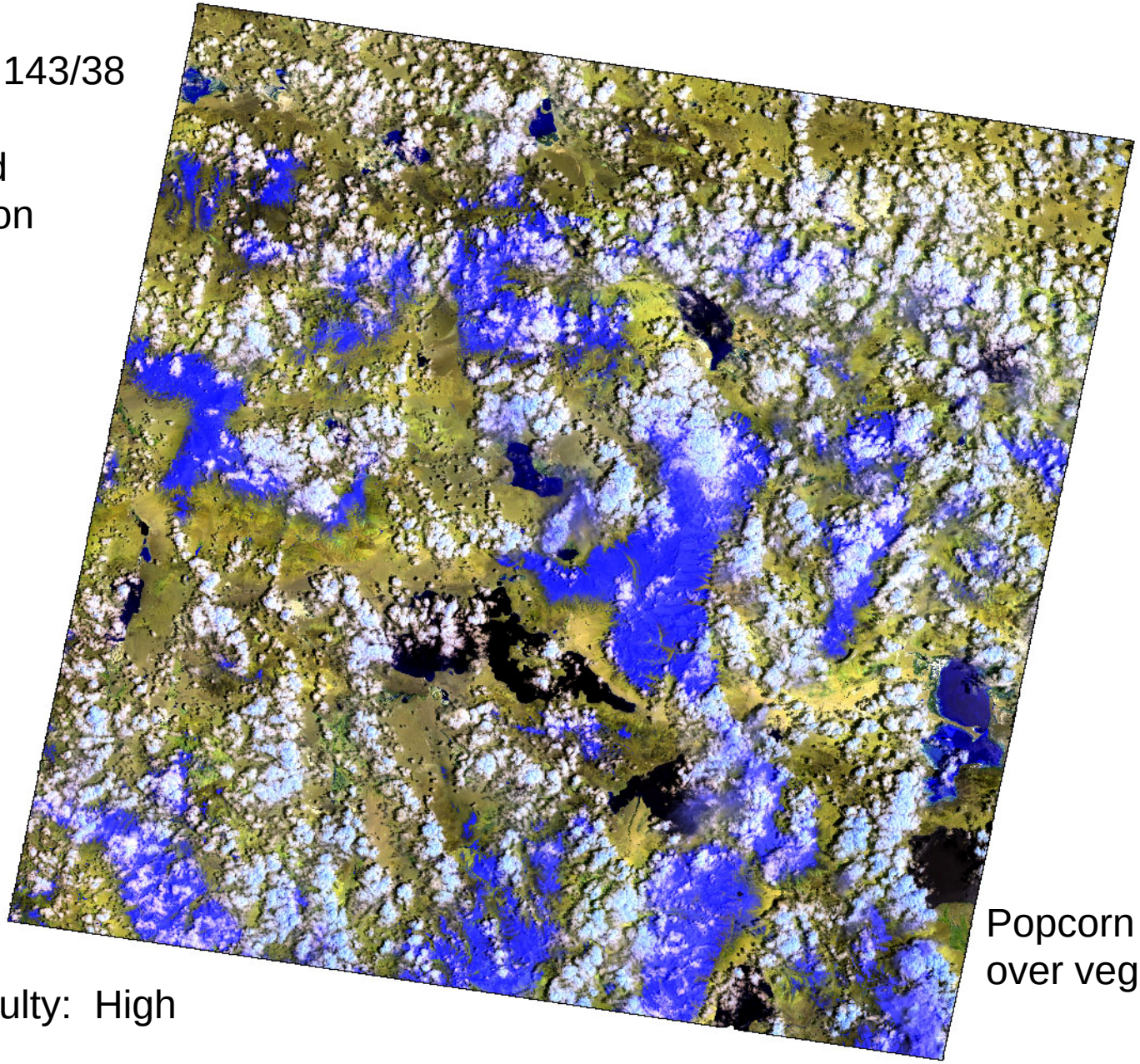


Final ACCA result
67.1% clouds

Cold Cloud
Provisional Cloud
Water
NDSI Noncloud
Snow/Ice
B6 Noncloud
B5 Noncloud
Ambiguous

p143 r38 -- Image

Path/Row 143/38
5/31/2001
7,5,2 band
combination



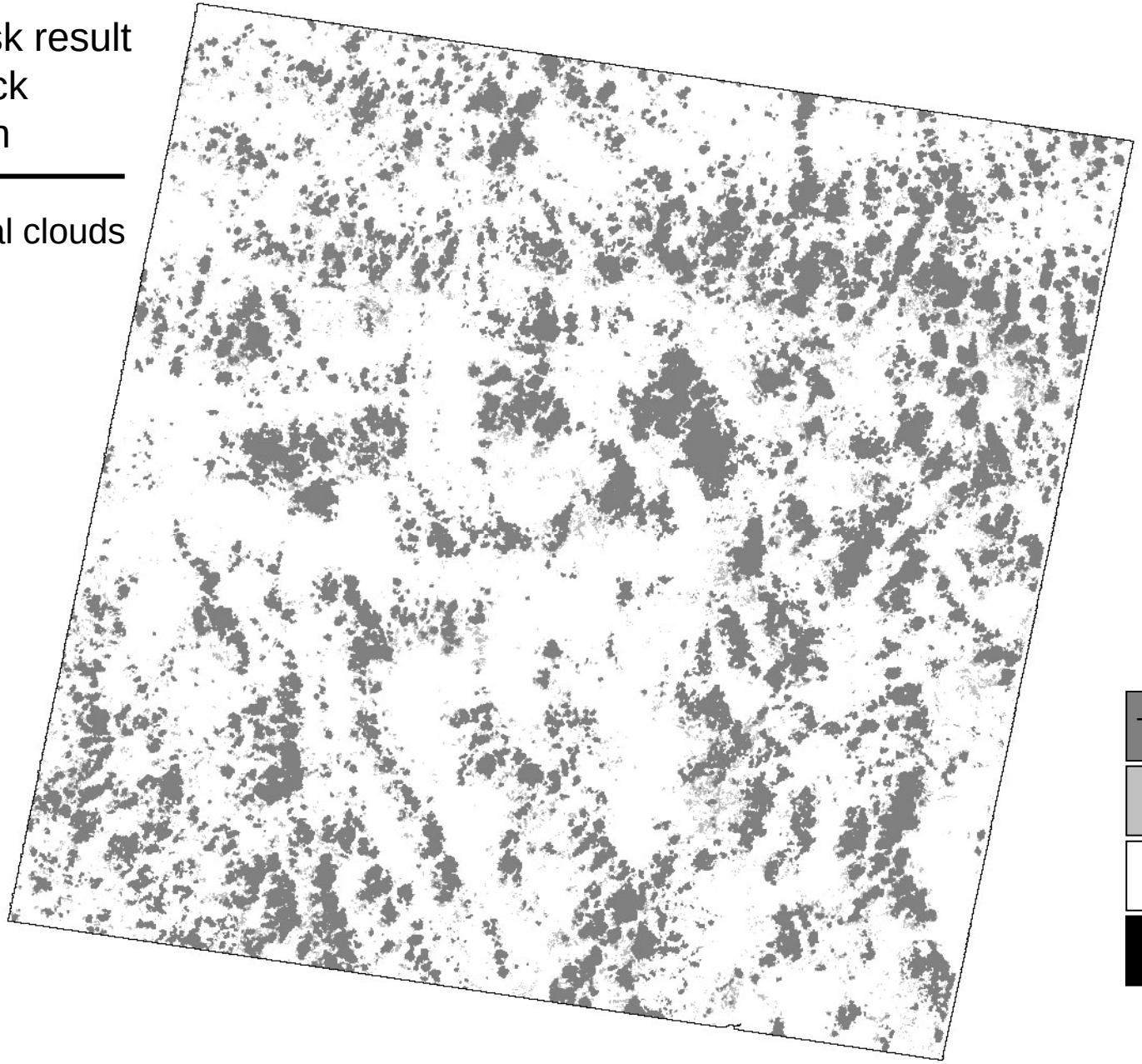
Popcorn clouds
over veg and snow.

CCA Difficulty: High

p143 r38 -- Manual Mask

Manual mask result
23.6% thick
+ 6.0% thin

29.6% total clouds



Thick Clouds

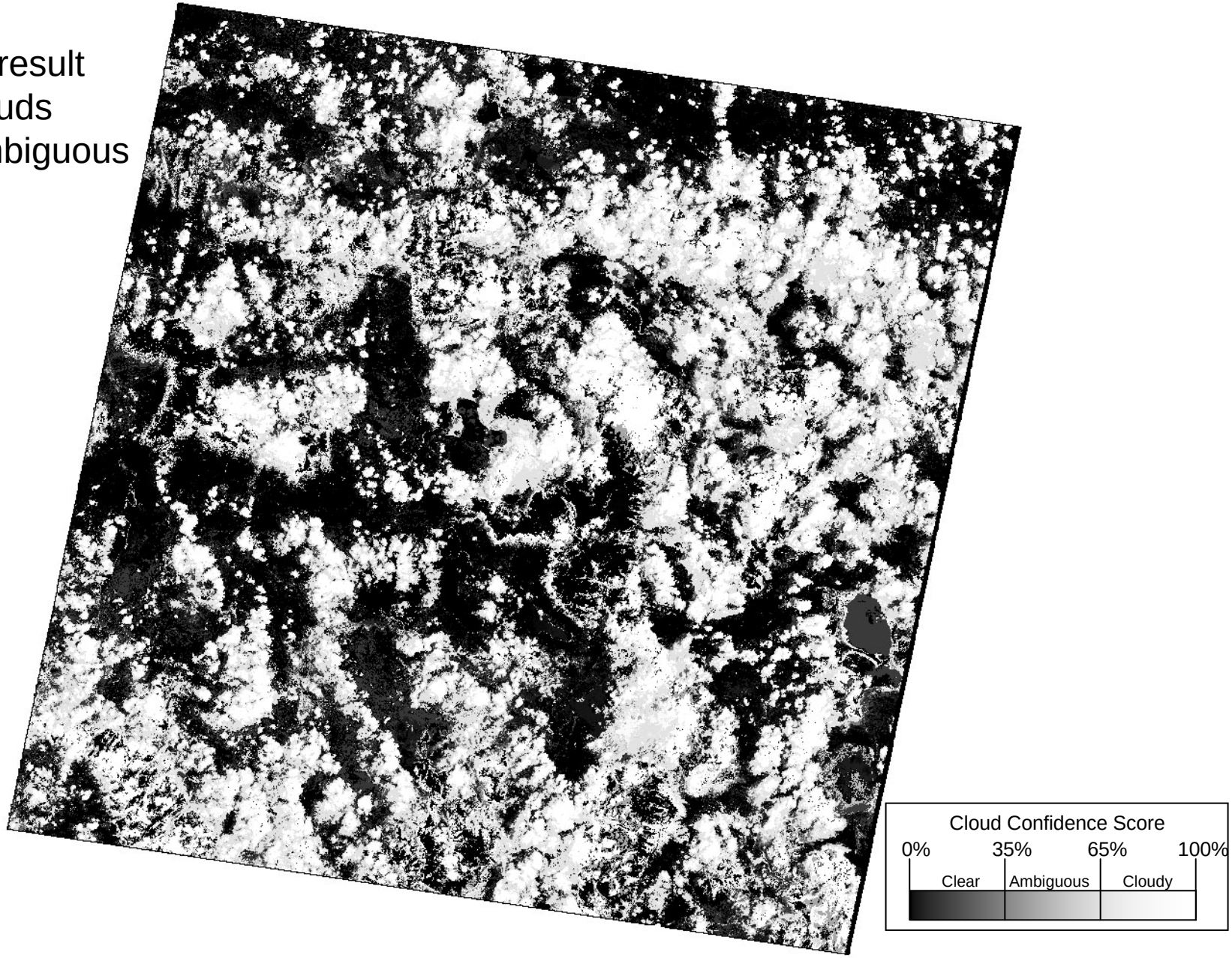
Thin Clouds

Clear

Shadows

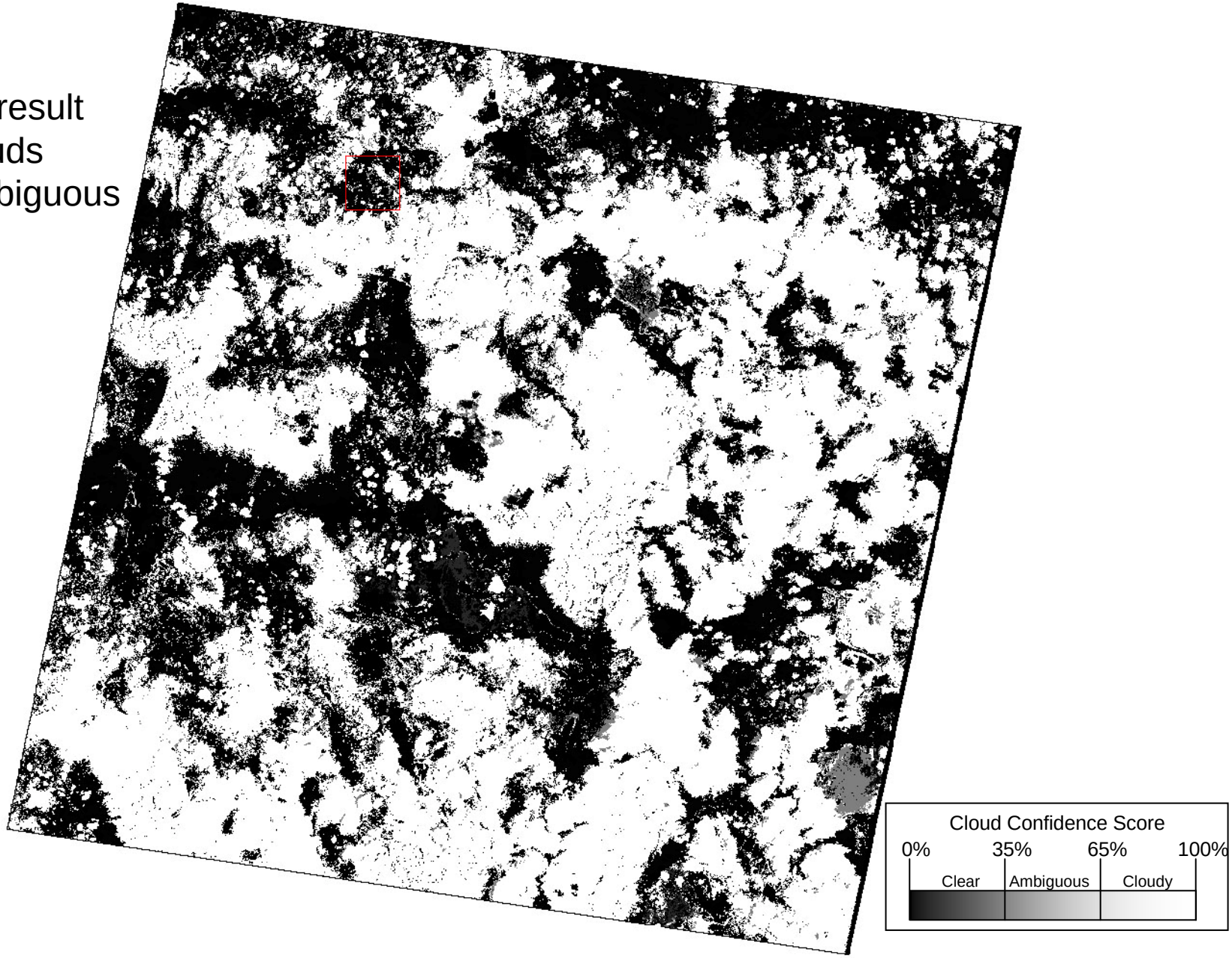
p143 r38 -- See5 rB2 confidence mask

See5 rB2 result
46.6% clouds
5.6% ambiguous



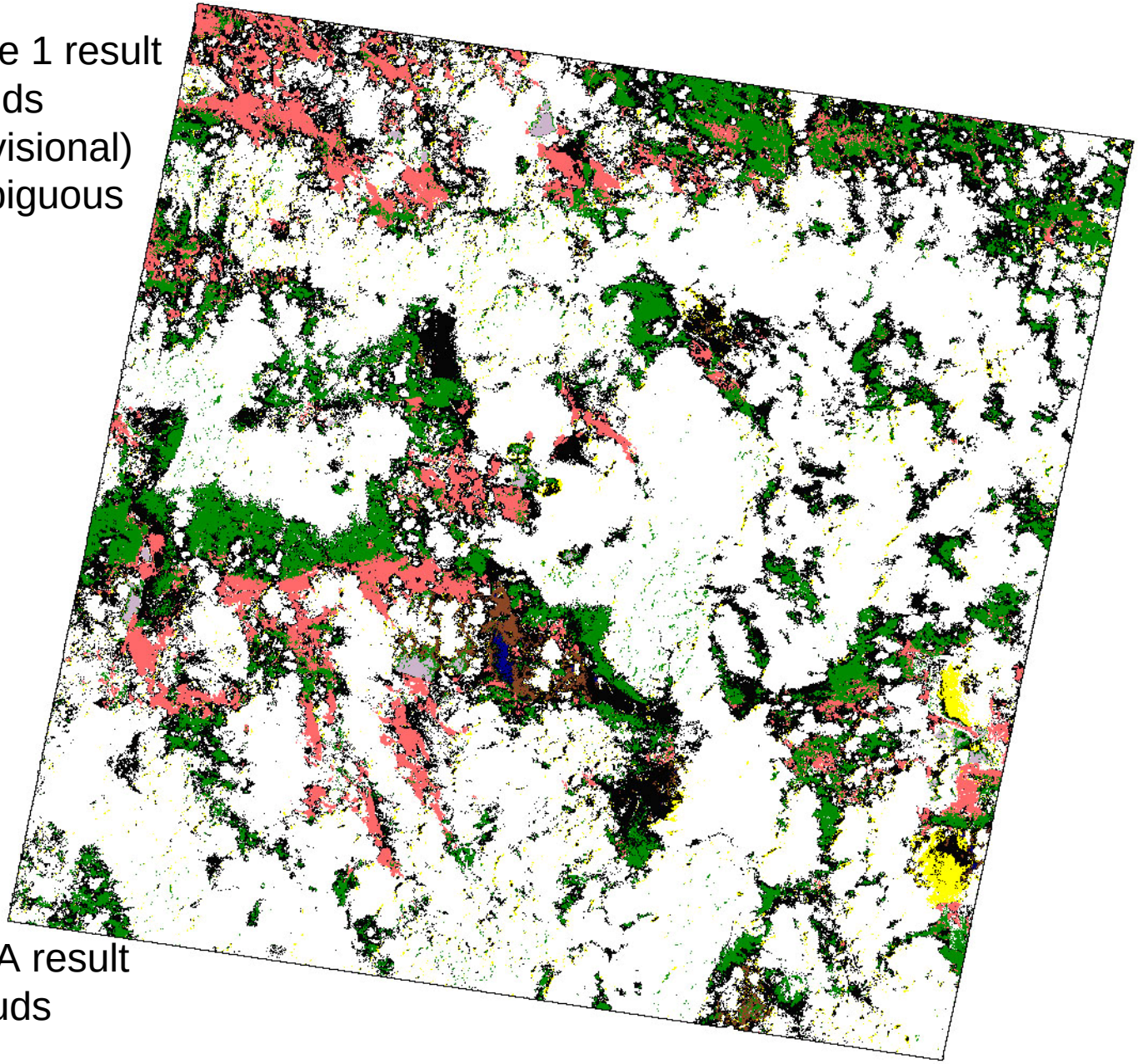
p143 r38 -- Expanded AT-ACCA confidence mask

Expanded
AT-ACCA result
58.6% clouds
  2.5% ambiguous



p143 r38 -- ACCA phase 1 mask

ACCA phase 1 result
65.1% clouds
(1.7% provisional)
19.8% ambiguous

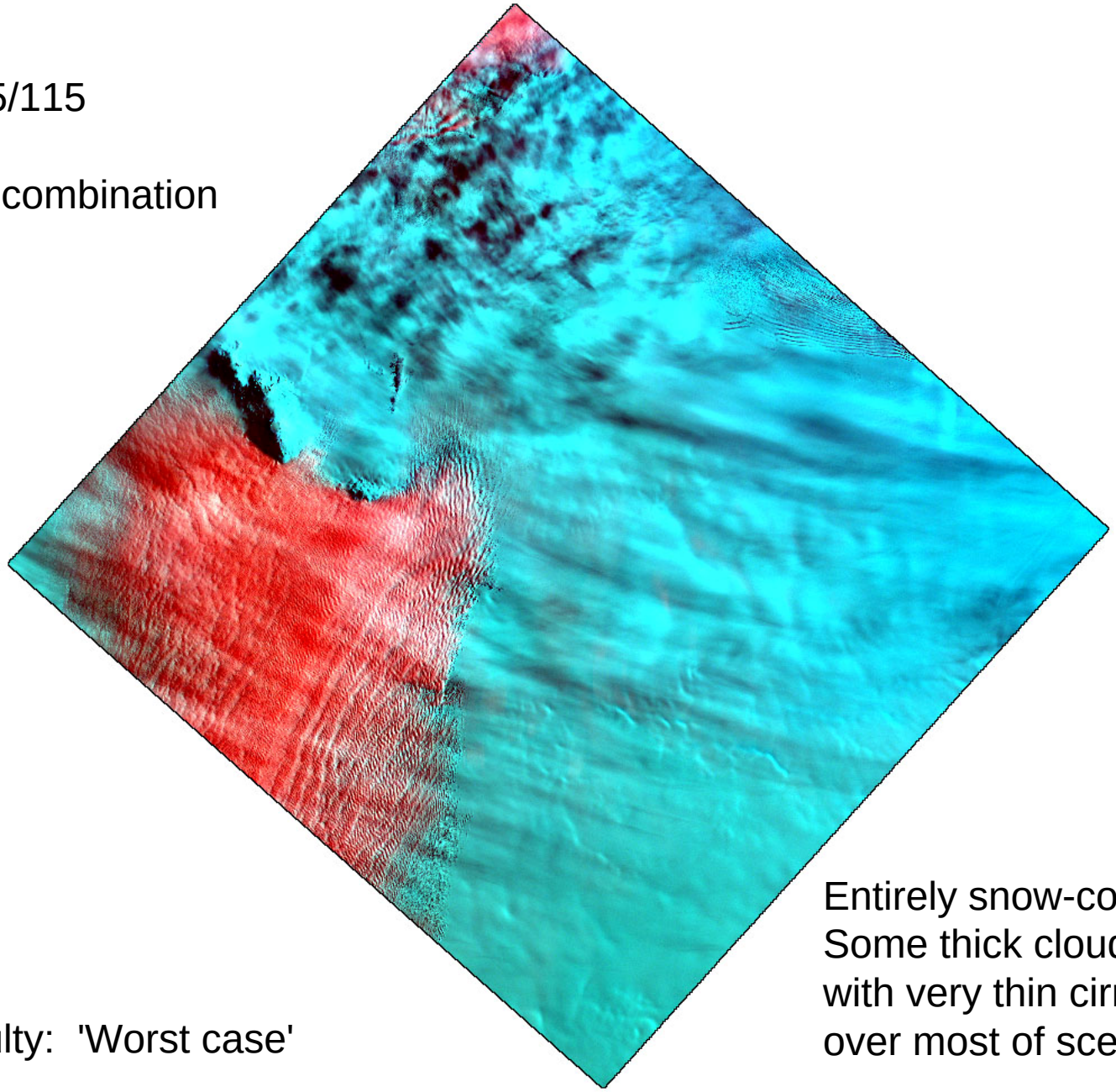


Final ACCA result
80.4% clouds

Cold Cloud
Provisional Cloud
Water
NDSI Noncloud
Snow/Ice
B6 Noncloud
B5 Noncloud
Ambiguous

p5 r115 -- Image

Path/Row 5/115
12/18/2001
5,4,1 band combination



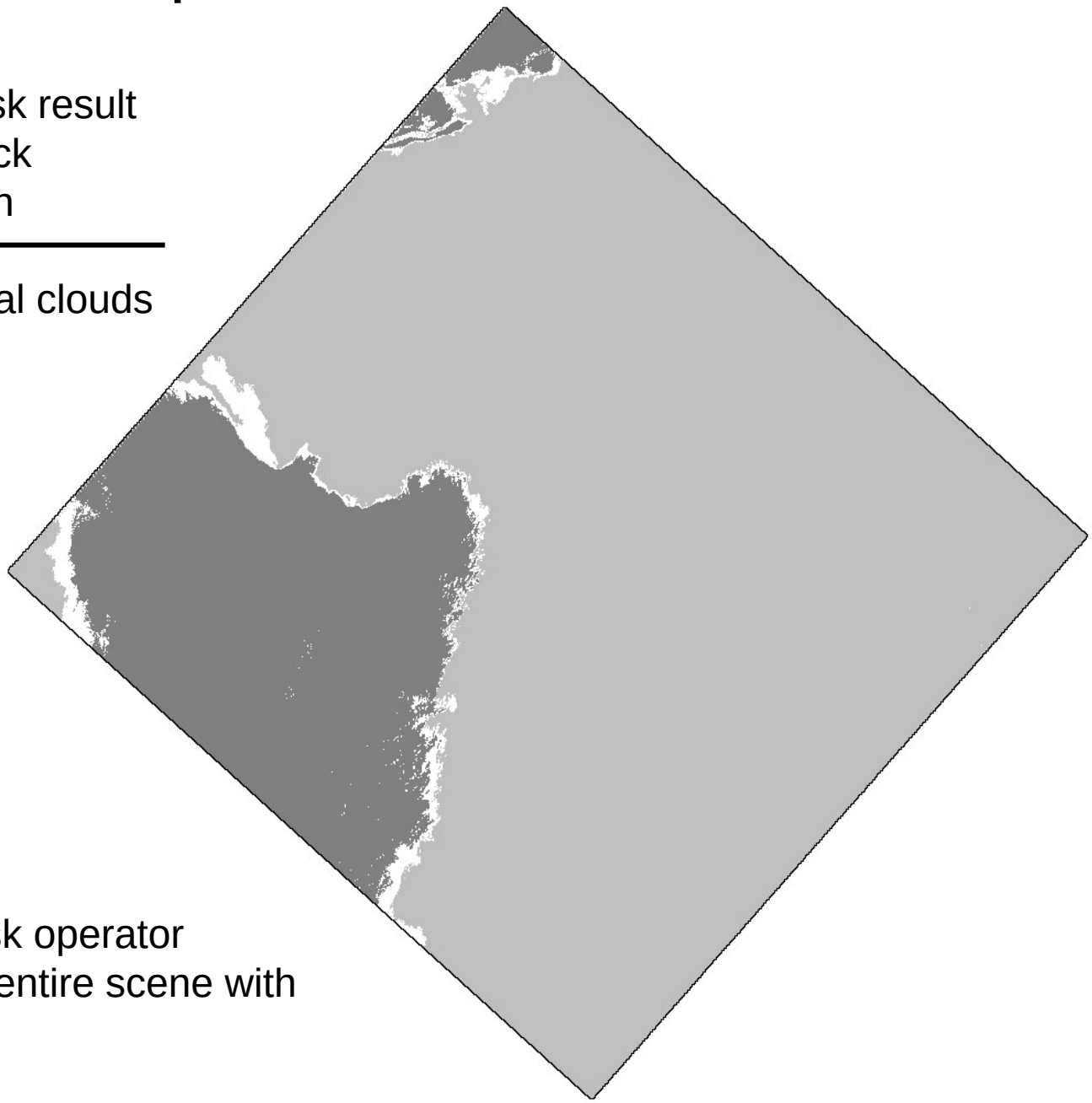
Entirely snow-covered scene.
Some thick clouds,
with very thin cirrus and fog
over most of scene.

CCA Difficulty: 'Worst case'

p5 r115 -- Manual mask

Manual mask result
21.4% thick
+ 76.1% thin

97.5% total clouds

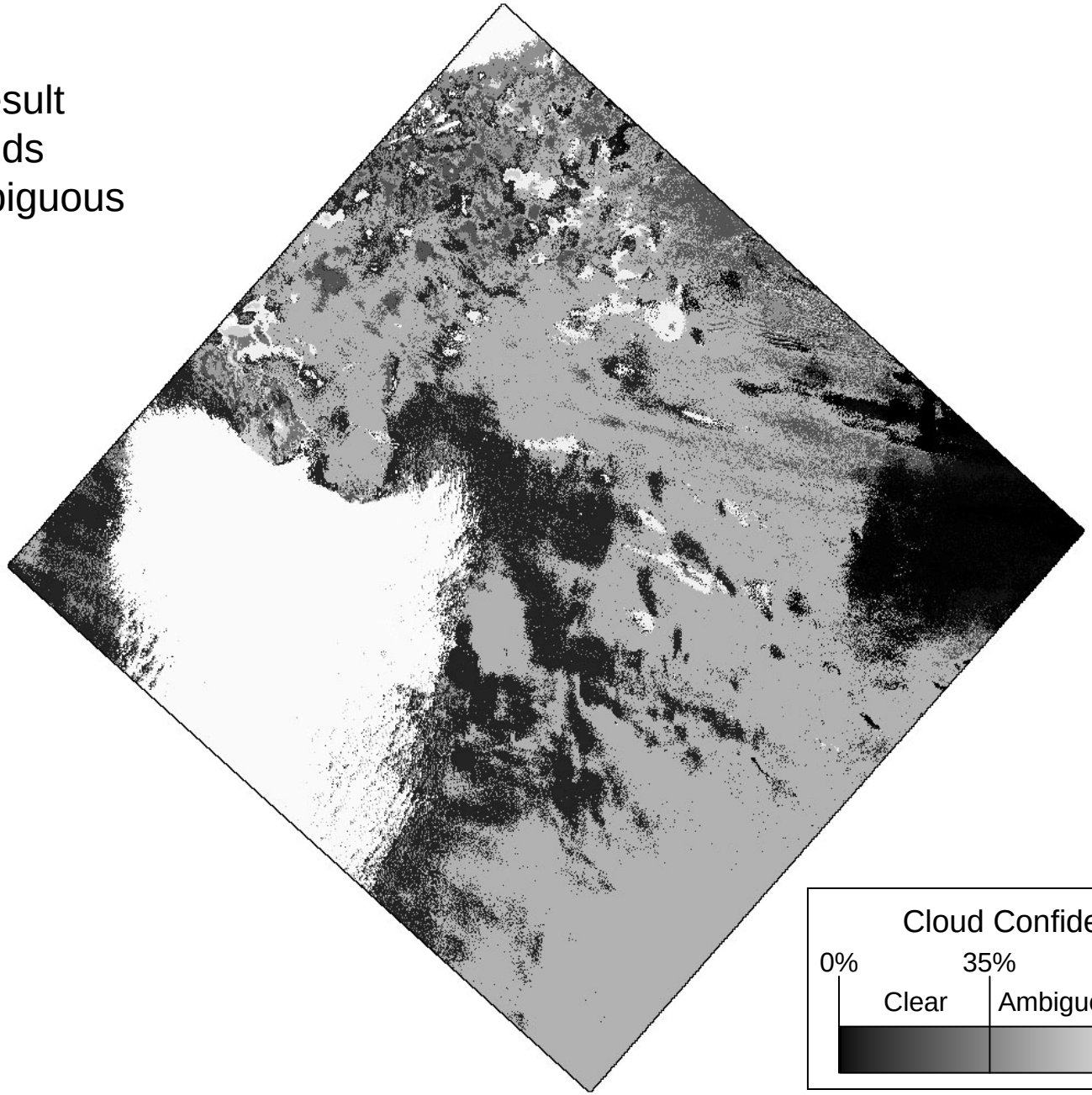


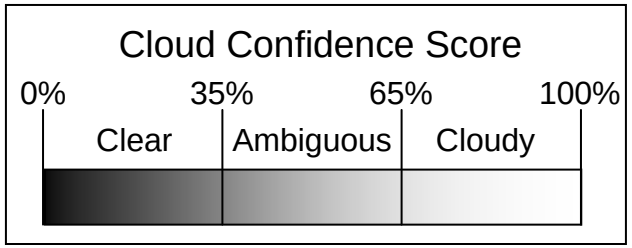
Thick Clouds
Thin Clouds
Clear
Shadows

Manual mask operator painted the entire scene with thin clouds.

p5 r115 -- See5 rB2 confidence mask

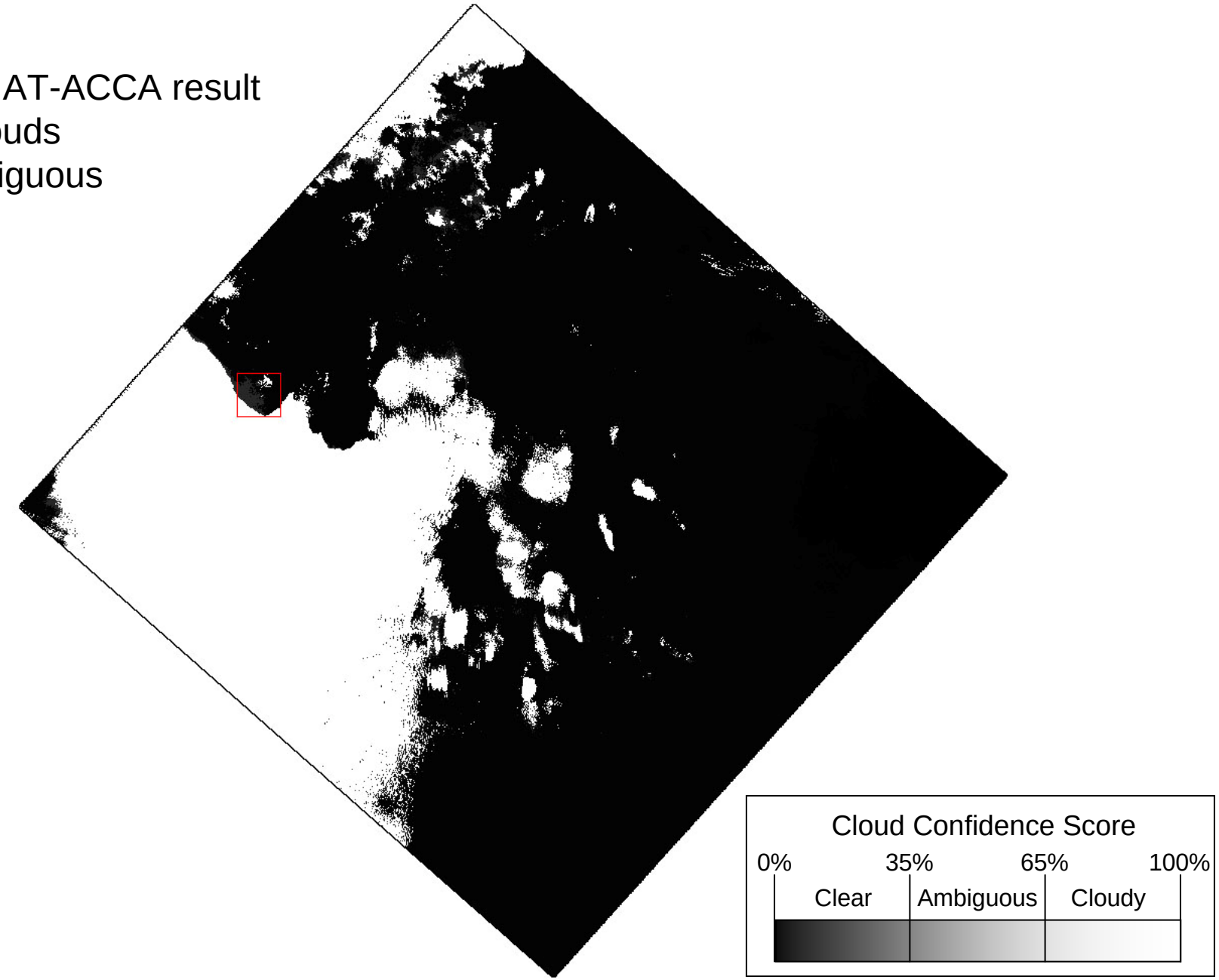
See5 rB2 result
65.3% clouds
4.4% ambiguous





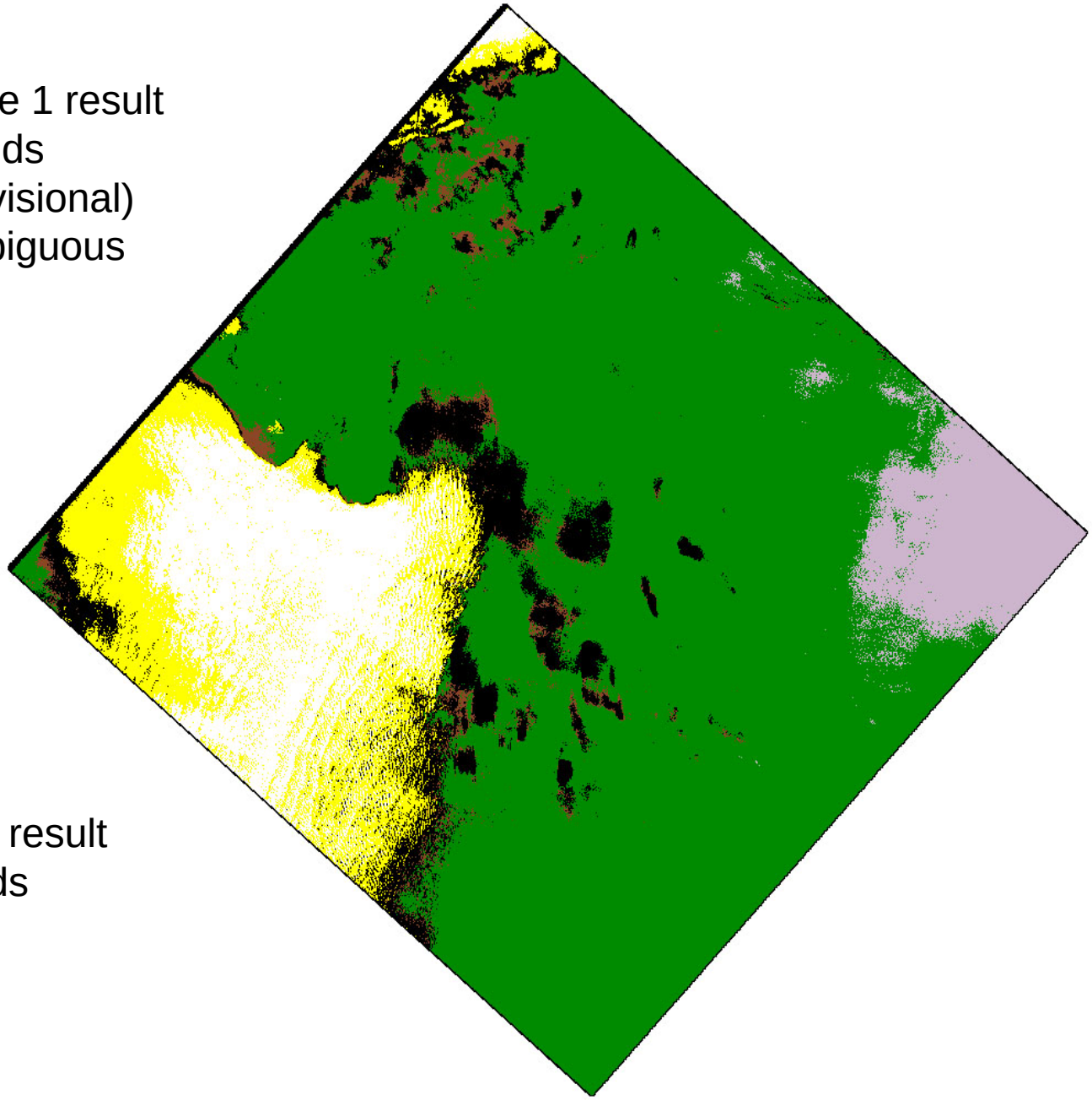
p5 r115 -- Expanded AT-ACCA confidence mask

Expanded AT-ACCA result
28.5% clouds
0% ambiguous



p5 r115 -- ACCA phase 1 mask

ACCA phase 1 result
20.7% clouds
(8.3% provisional)
7.4% ambiguous



Final ACCA result
26.5% clouds

Cold Cloud
Provisional Cloud
Water
NDSI Noncloud
Snow/Ice
B6 Noncloud
B5 Noncloud
Ambiguous