



LDCM Cloud Cover Assessment Update

October 27th, 2009

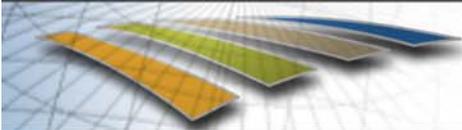
Pat Scaramuzza, SGT
U.S. Geological Survey Center for Earth Resources Observation and Science
Sioux Falls, SD

Goals for LDCM CCA

LDCM

Our Goals:

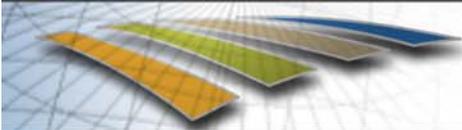
- **Mitigation**
 - This is a risk mitigation activity due to the formal absence of thermal data on LDCM.
- **ACCA-like accuracy.**
 - Our goal is not to improve on L7 ACCA, but to have a CCA method that works as well as ACCA but without a thermal band and with a full resolution cloud mask.
- **Level 1 mask only.**
 - Designed as input to LTAP and as an aid to typical users.
 - Compositing requires Level 2 CCA masks with 99% accuracy. This is outside our scope.
- **Empirical approach.**
 - We have large quantities of data with which to generate empirical models.
 - Statistics is a science, just not a well-respected one.
- **Algorithm should be as sensor-neutral as possible.**
 - Works with TOA Reflectance.
 - Trained and tested on L7 and Hyperion data. Will test on OLI simulated data.
 - Should work on any sensor with Landsat-like bands.
- **For CCA use only.**
 - AT (Artificial Thermal) band is not intended as a replacement for Band 6 for any other purposes.

 CCA Algorithms under consideration

- **ACCA phase 1**
 - Only portion of L7 ACCA algorithm that creates a pixel mask.
 - Studied for comparison purposes.

- **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 Cirrus and gD02 algorithms.
 - Cirrus algorithm will use the OLI cirrus band, and is under development.
 - gD02
 - Series of threshold tests designed to maximize likelihood ratio and minimize errors.
 - Only to be used to resolve pixels marked as ambiguous by AT-ACCA.

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



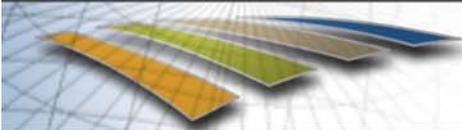
CCA Algorithms compared

LDCM

	Pixels Correct	Pixels False	Pixels Ambiguous	Misclassified Clouds	Misclassified Clear
<i>ACCAp1</i>	79.9%	4.3%	15.8%	7.8%	2.3%
See5-rE	88.5%	7.3%	4.2%	12.1%	4.5%
Expanded AT-ACCA (ATgD02)	89.8%	8.5%	1.7%	12.3%	6.3%
ATgD02 on Hyperion data (non-cirrus only)	75.7%	8.5%	1.9%	30.7%	7.5%

The Hyperion results indicate that AT-ACCA works acceptably across Landsat-like platforms

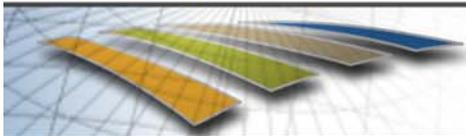
LDCM ACCA with TIRS has not yet been studied.



Hyperion Cirrus Study

LDCM

- Hyperion data was reformatted to resemble OLI bands 1-7 and 9 (Historic Landsat bands 1-5 and 7, plus Cirrus and Coastal Aerosol bands).
- OLI-like bands were created with square spectral response.
- 22 Hyperion scene dataset (23,593,157 pixels)
 - Split into 11 Training scenes, 11 Validation scenes.
(12,795,587 pixels) (10,797,570 pixels)
 - 4 scenes (2 T, 2 V) were high-altitude playa.
- Manual masks created
 - Same methodology as used for L7 masks, with the addition of a cirrus designation.
- See5 algorithm trained on random, stratified sample of Training scenes. (60,809 pixels)
 - 2000 pixels of each existing classification (thickcloud, thincloud, clear, cirrus) taken at random from each scene.



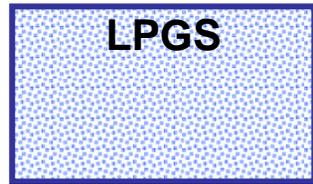
Hyperion Cirrus results so far

	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.

It may be more effective to mask out playas and use the simple test, rather than to create a global algorithm.

- (7/16/09) Proposal -



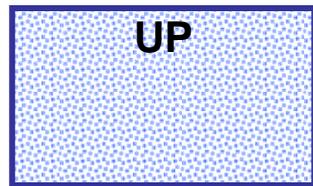
**Quality (Assessment) Band
(Pixel Level Metadata)**



L1 Product QB (16-bit)



Online Cache



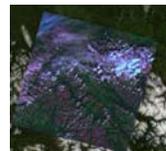
QB (8-bit)



QB Overlay (PNG)



Online Cache



FRB

Bit	Description	Bit	Description
0	Designated Fill [±]	8	Vegetation Confidence
1	Dropped Frame [±]	9	
2	Terrain Occlusion	10	Snow/Ice Confidence
3	<Artifact X*>	11	
4	Water Confidence	12	Cirrus Confidence
5		13	
6	<Type #6 Confidence*>	14	Cloud Confidence
7		15	

* - Placeholder; determined by CDR ± - Not currently required

Bit	Description	Color
0	Designated Fill [±]	Black
1	Dropped Frame [±]	Brown
2	Terrain Occlusion	Red
3	Water ^{**}	Blue
4	Vegetation ^{**}	Green
5	Snow/Ice ^{**}	Light Blue
6	Cirrus ^{**}	Yellow
7	Cloud ^{**}	White

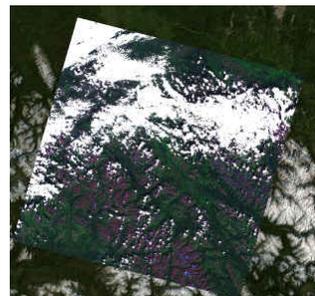
(H) ← Precedence → (L)

Confidence Levels
00 = none or unset
01 = 0-33% confidence
10 = 34-66% confidence
11 = 67-100% confidence

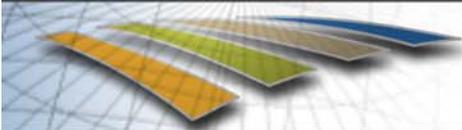
- 16-bit QB rolls off the Online Cache with the product
- 8-bit QB available with the FRB
- QB Overlay is configurable
- Precedence is for overlapping pixels

** - Set for highest Confidence Level (11)

FRB with Quality Band Overlay example



Preliminary – Final at CDR

Ancillary Data possibilities

Scene Level quality metrics in Metadata

- In:
 - Number of Dropped Frames
 - Failed CRCs
 - Inoperable Detectors
 - Scene-wide Cloud score
 - Geodetic Accuracy.
- Other possibilities:
 - Striping metric
 - SNR
 - Average NDVI.

Science Team input wanted. What other scene-level metrics would be useful?

End of planned slides
(after this are supporting slides)

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**

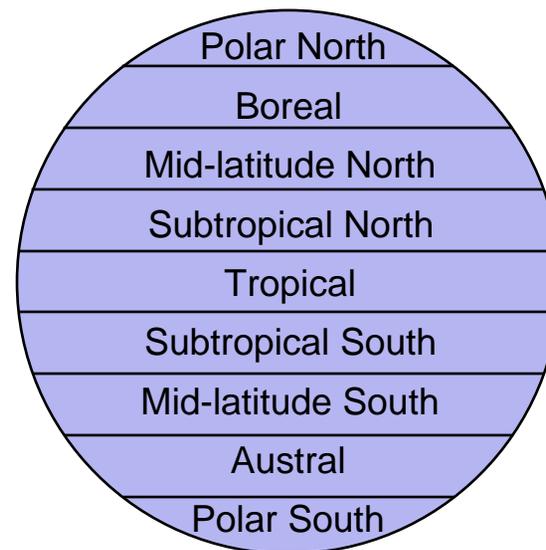
- 5 scenes were excluded due to visible gain change artifacts.
 - One scene had a gain change but no visible artifacts; it was retained.
- Manual masks were created by three operators at EDC.
- Cloud masks should be treated as having ~7% error.
 - Based on 11 masks duplicated by all three operators.

- **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 pscar@usgs.gov for instructions.

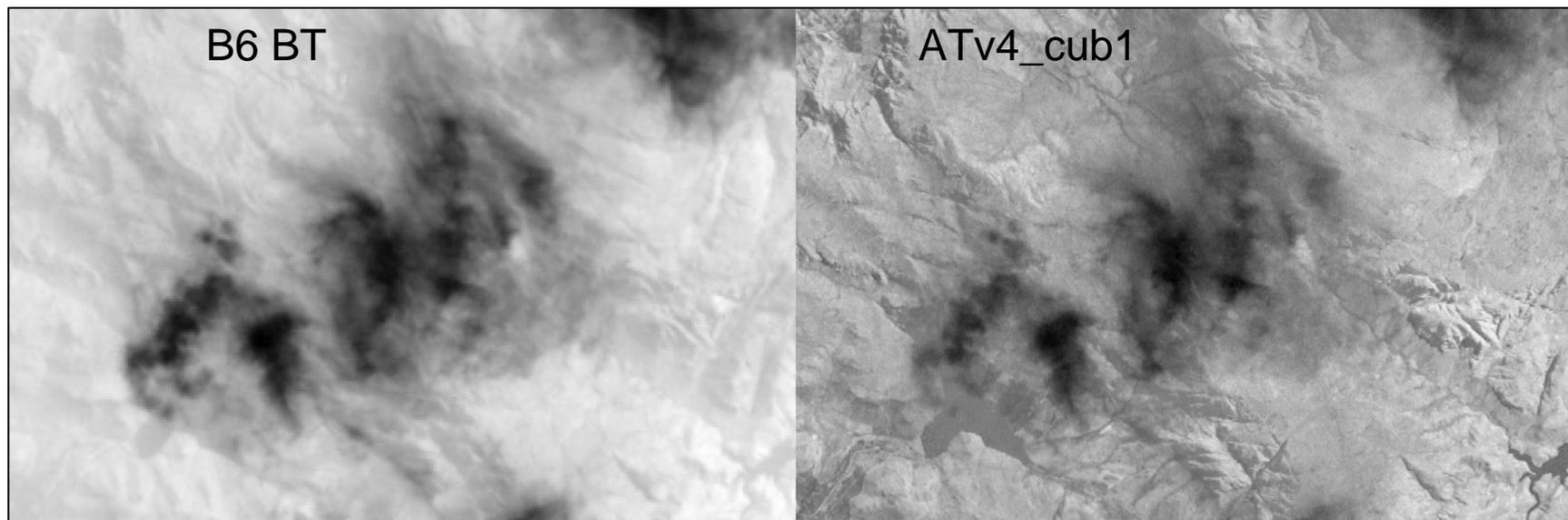


AT (Artificial Thermal) Band

LDCM

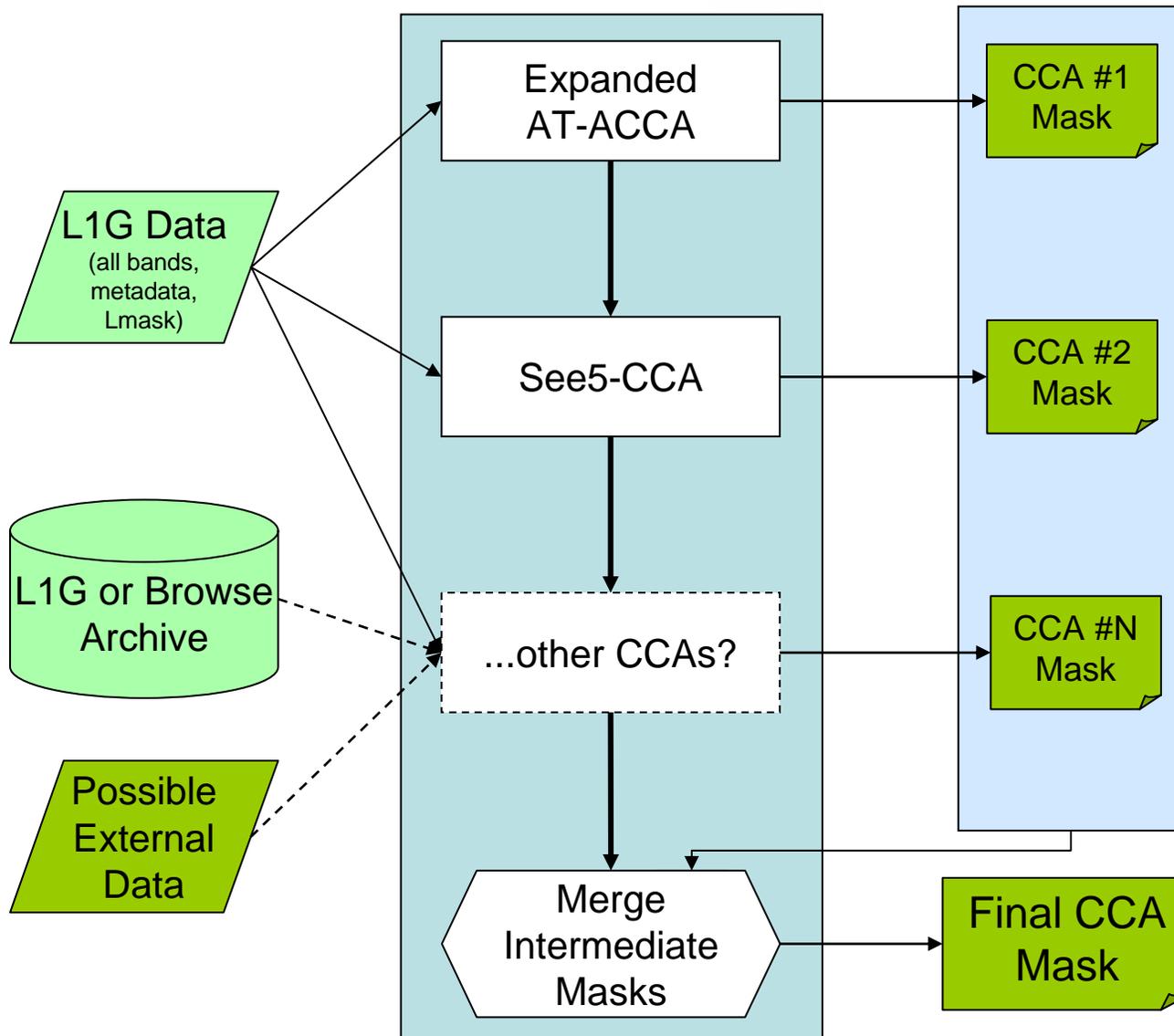
Statistical model of thermal band (B6) derived from reflective bands (B1-5 & 7)

- Model created by Rulequest Cubist software (will not be necessary in the operational system.)
- Various models examined and selected for speed and continuity (single rule).
- Model ATv4_cub1:
 - Trained on 1.1 million points selected at random from all 212 Irish scenes (~5,200 points per scene.)
 - RMS Average Difference from actual Band 6 brightness temperature over entire Irish data set (3.9×10^9 pixels) = **8.8 ° K.**



Proposed System

LDCM

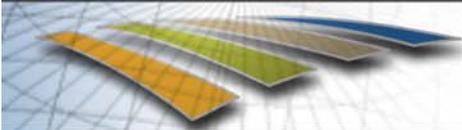


AT-ACCA and See-5 CCA are the only algorithms being implemented at this time.

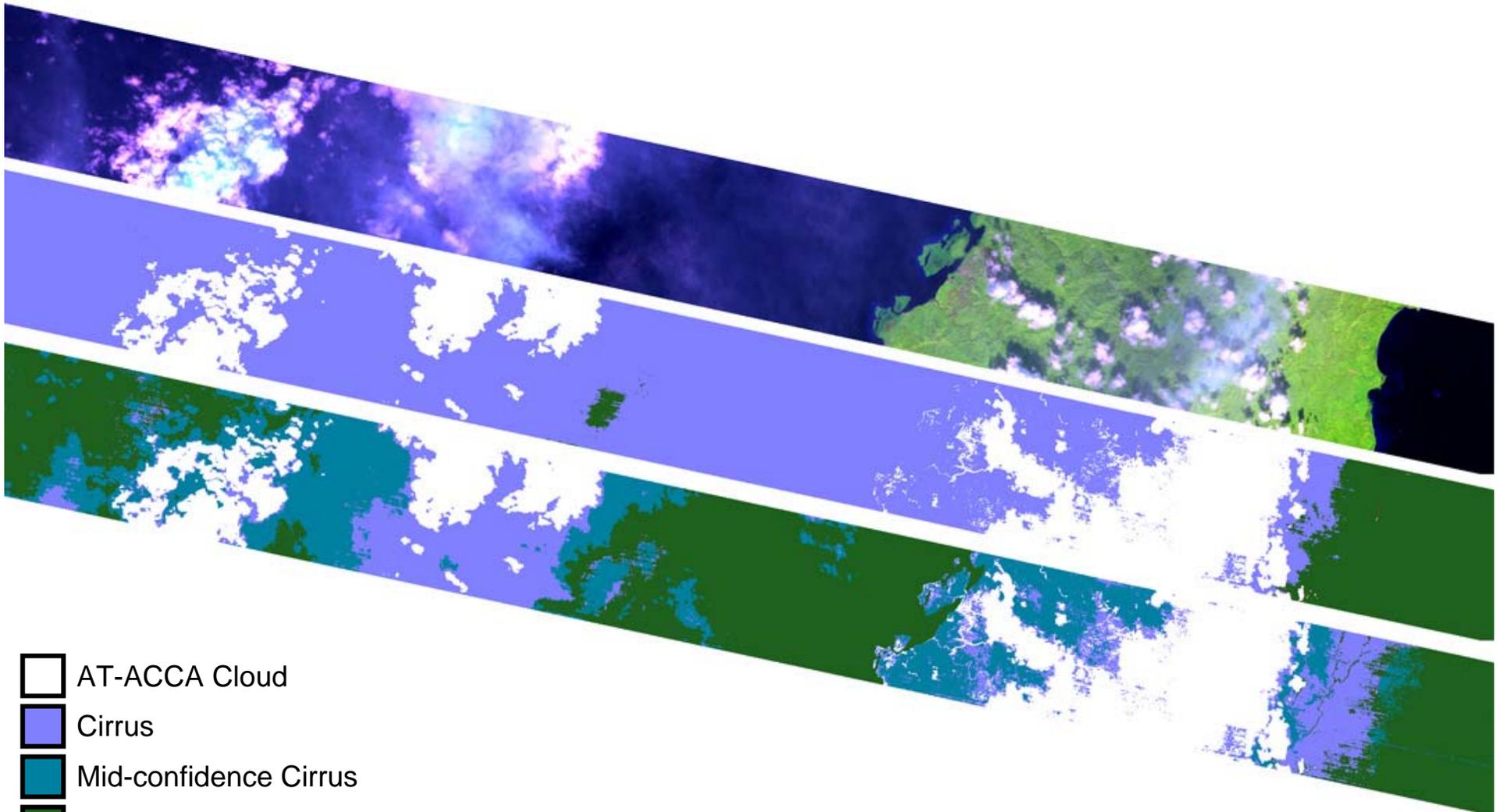
(Cirrus and gD02 are being implemented as part of Expanded AT-ACCA.)

Temporal CCA may be added after OIVP.

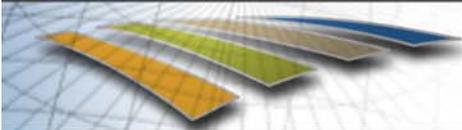
Other CCAs can still be added, within project constraints.



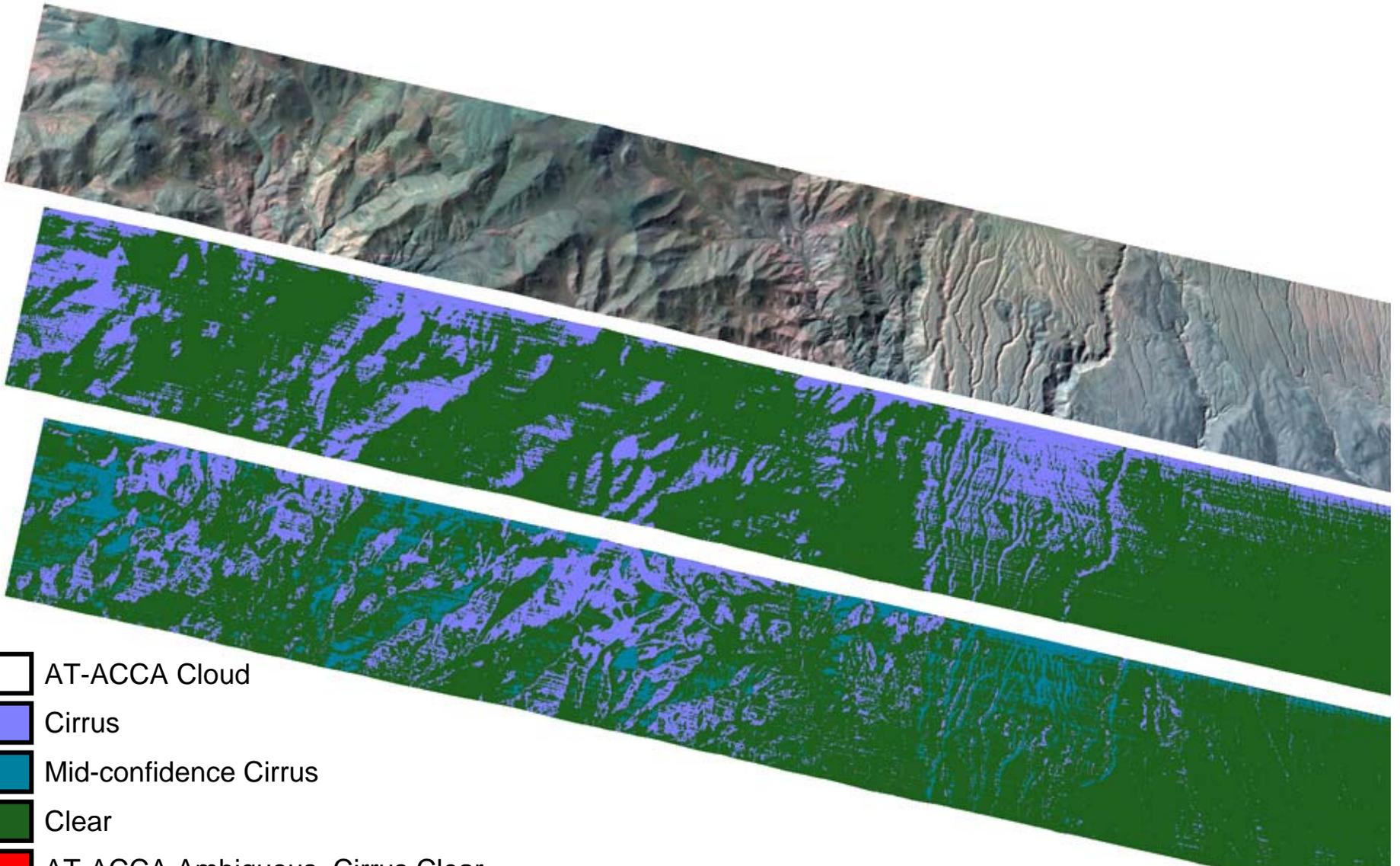
p088r065 example



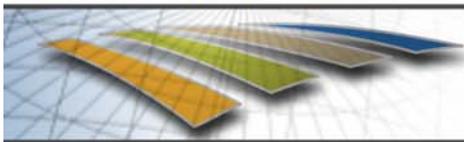
-  AT-ACCA Cloud
-  Cirrus
-  Mid-confidence Cirrus
-  Clear
-  AT-ACCA Ambiguous, Cirrus Clear



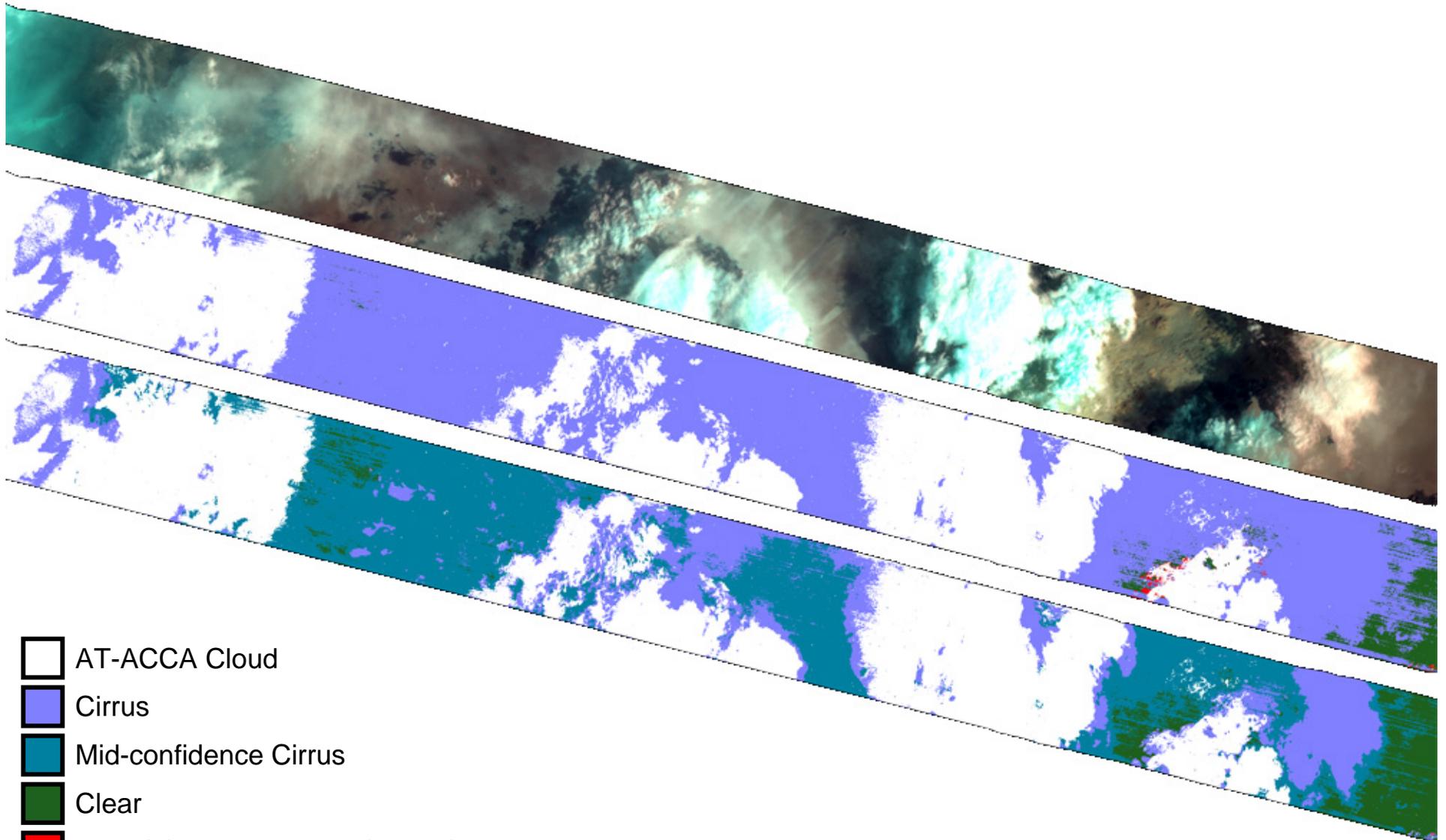
p001r074 example



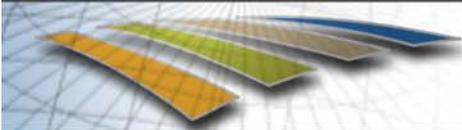
-  AT-ACCA Cloud
-  Cirrus
-  Mid-confidence Cirrus
-  Clear
-  AT-ACCA Ambiguous, Cirrus Clear



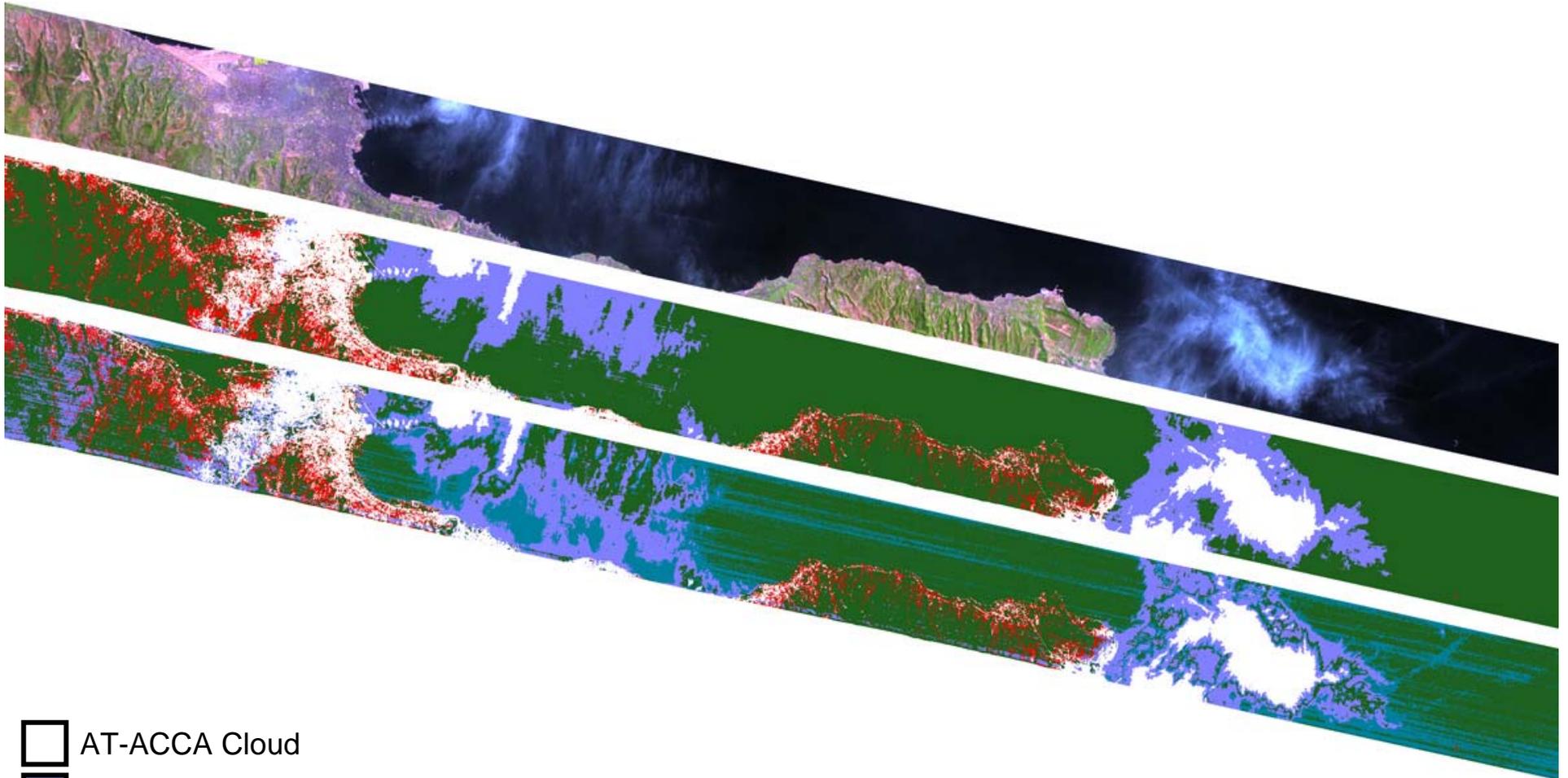
p232r077_2 example



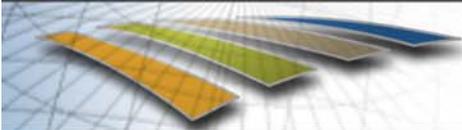
-  AT-ACCA Cloud
-  Cirrus
-  Mid-confidence Cirrus
-  Clear
-  AT-ACCA Ambiguous, Cirrus Clear



p174r036 example



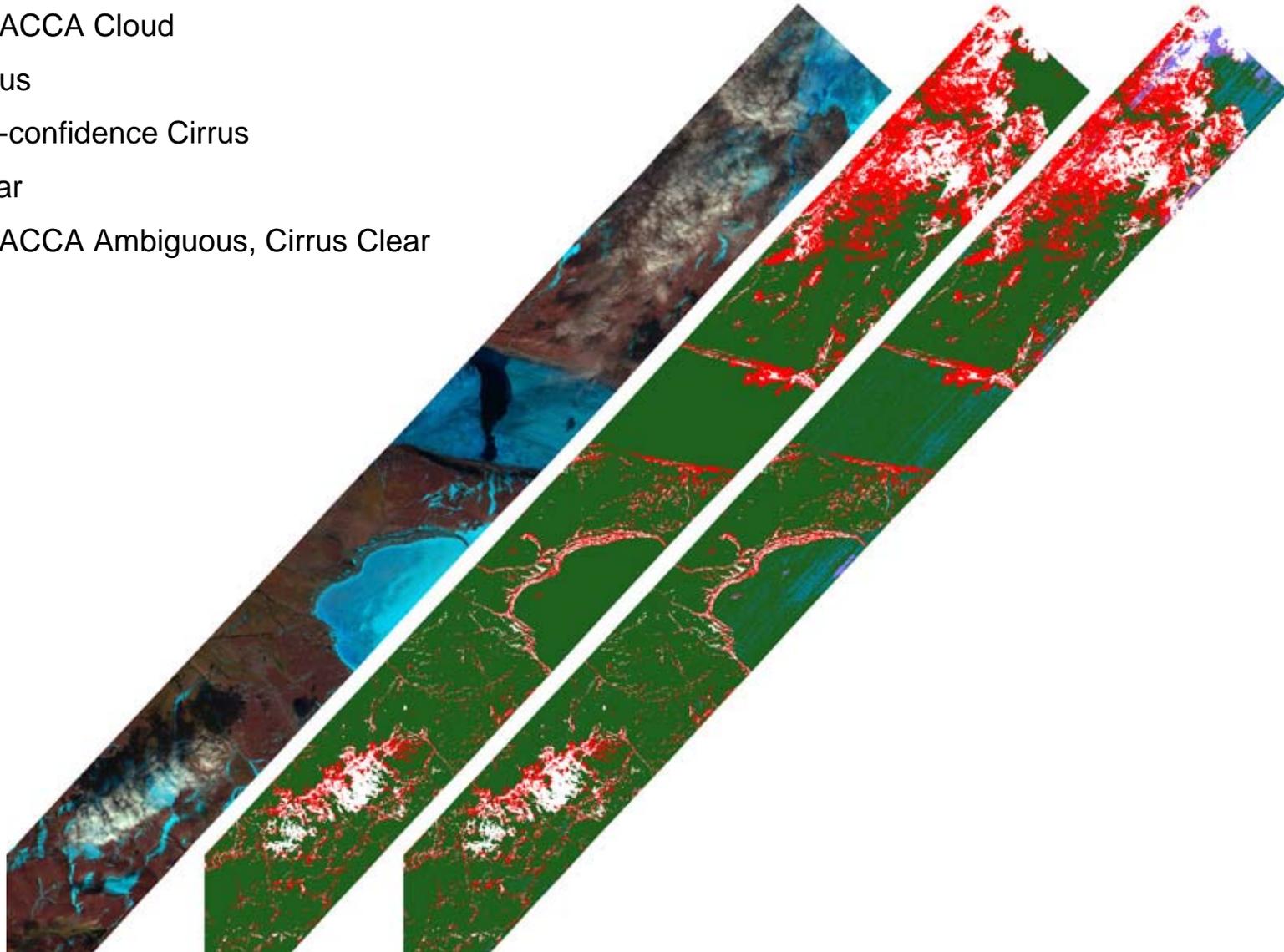
-  AT-ACCA Cloud
-  Cirrus
-  Mid-confidence Cirrus
-  Clear
-  AT-ACCA Ambiguous, Cirrus Clear

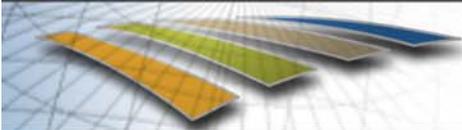


p030r005 example

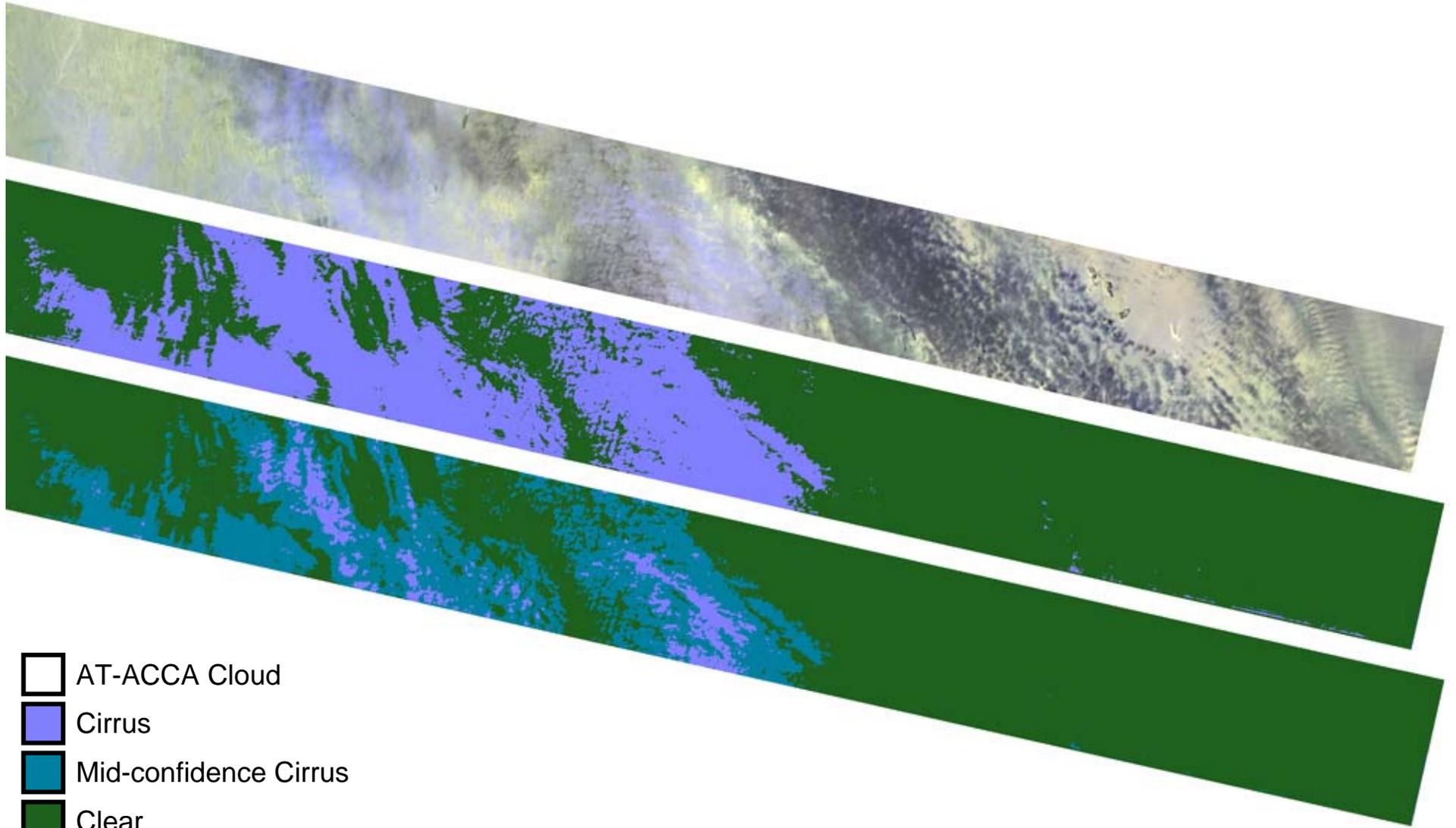
LDCM

- AT-ACCA Cloud
- Cirrus
- Mid-confidence Cirrus
- Clear
- AT-ACCA Ambiguous, Cirrus Clear

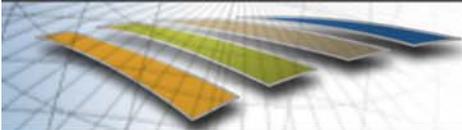




p190r045 example

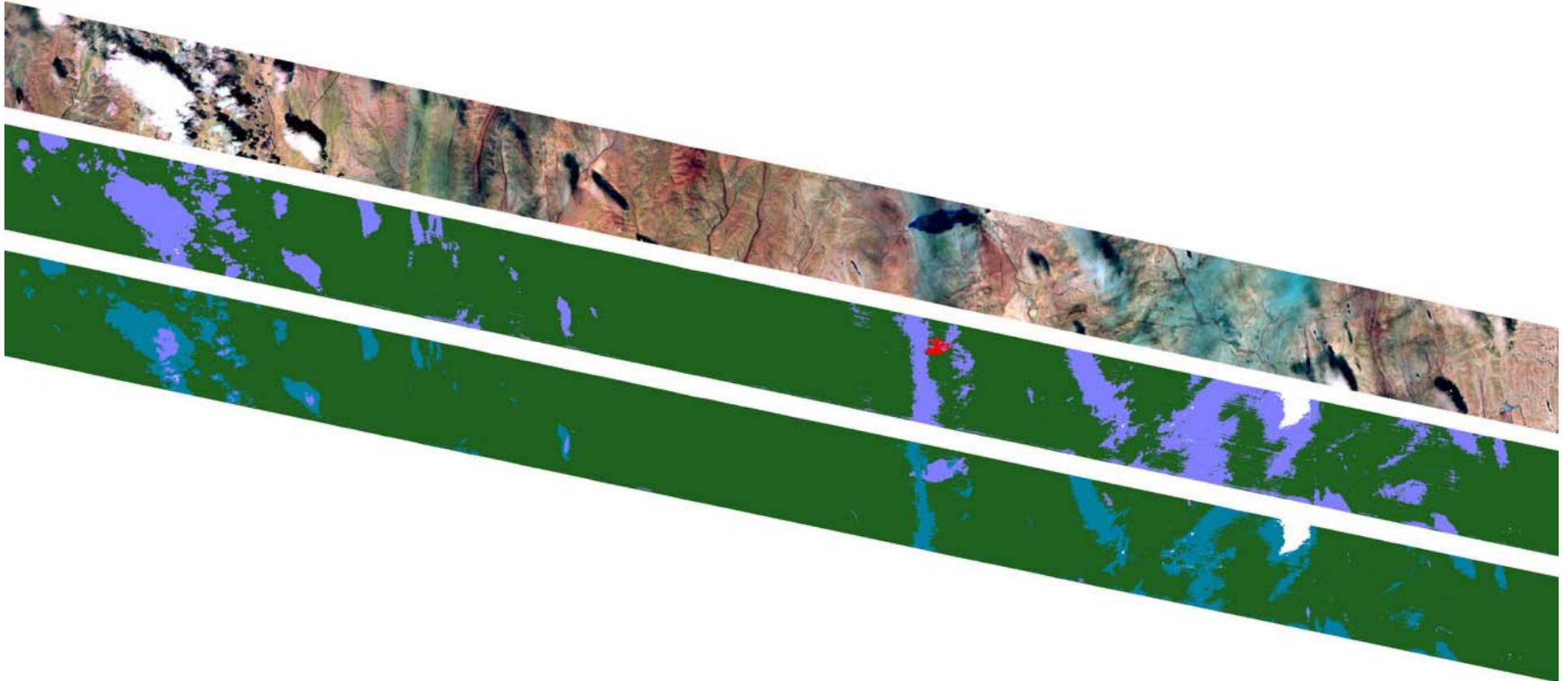


-  AT-ACCA Cloud
-  Cirrus
-  Mid-confidence Cirrus
-  Clear
-  AT-ACCA Ambiguous, Cirrus Clear



p140r036 example

LDCM



-  AT-ACCA Cloud
-  Cirrus
-  Mid-confidence Cirrus
-  Clear
-  AT-ACCA Ambiguous, Cirrus Clear