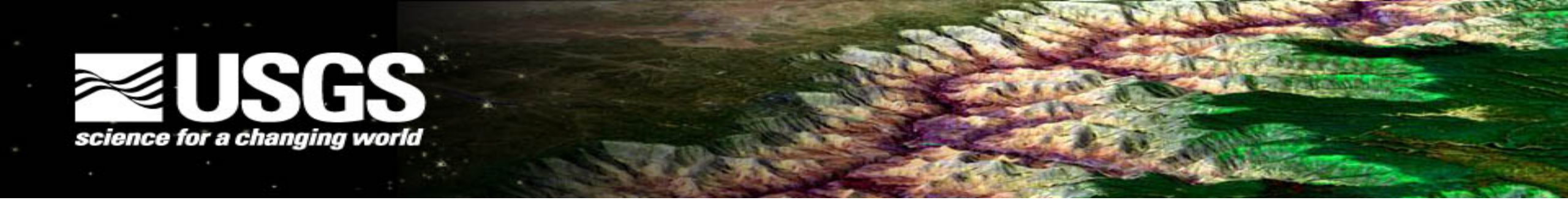


Quick overview of EROS Cloud Efforts

Pat Scaramuzza, SGT / EROS / USGS

pscar@contractor.usgs.gov

U.S. Department of the Interior

U.S. Geological Survey

Cloud Algorithms at EROS

- **ACCA**
  - Originally developed at NASA for scene-wide cloud scores in 1984.
  - EROS version creates a per-pixel cloud mask.

- **AT-ACCA**
  - Version of ACCA that does not require a thermal band.
  - Substitutes a modeled thermal band and uses a neural network to resolve ACCA ambiguous pixels.

- **See5**
  - Cloud mask developed using machine learning techniques on L7 data.
  - Was primary cloud mask algorithm on launch of Landsat 8.

- **CFMask**
  - Current default cloud mask algorithm for Landsat data.
  - Chosen because validation showed it to be more accurate than competing algorithms.

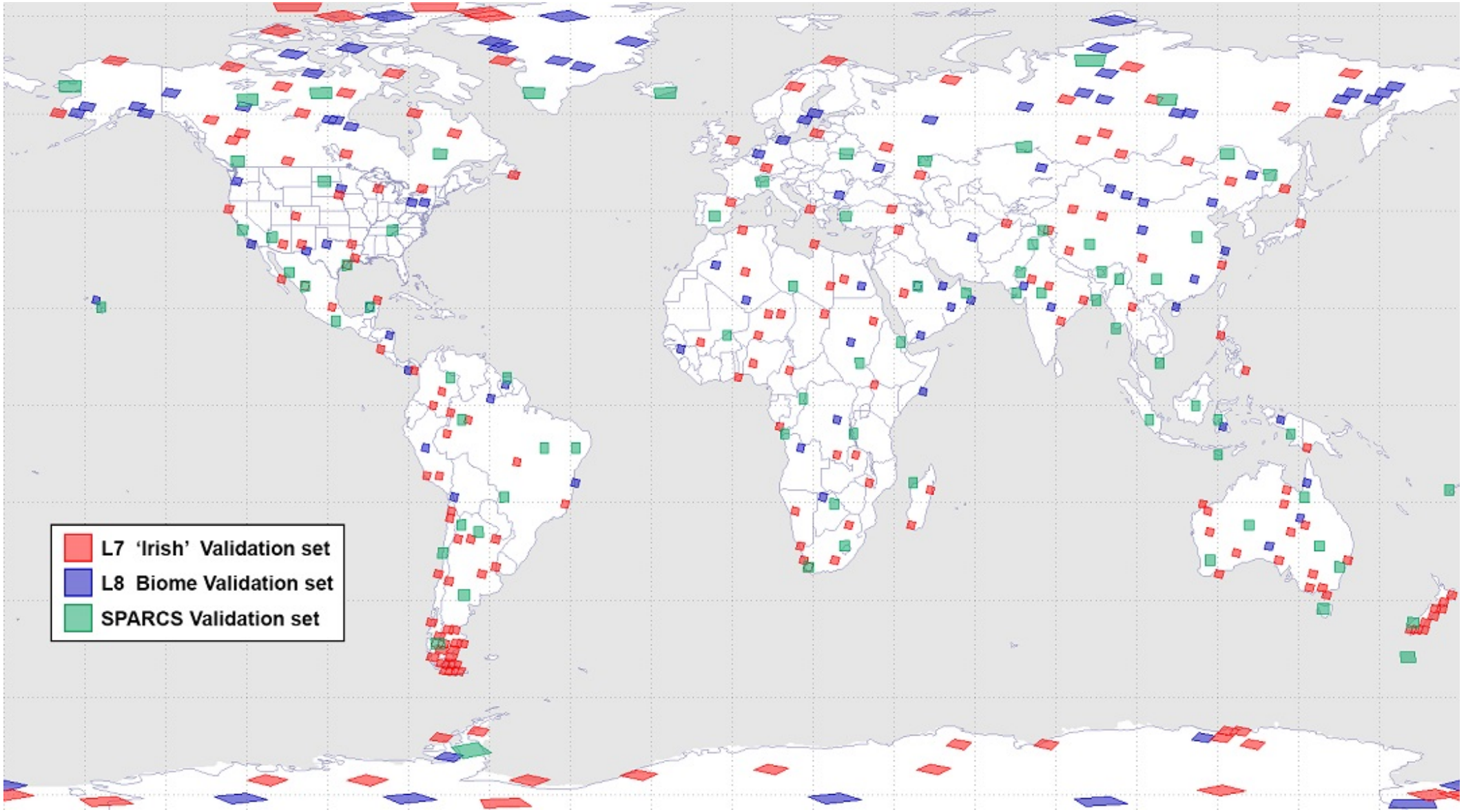
Cloud Validation Datasets at EROS

- **'Irish' dataset** http://dx.doi.org/10.5066/F7KH0KDG
  - 207 Landsat 7 scenes with manually generated cloud masks.
  - 104 scenes reserved for algorithm training and not used for validation.
  - 9 latitude zones x 21 examples per zone, plus seasonal sets over one scene per zone.
    - 5 scenes discarded from original 212 scene set due to L7 ETM+ gain change artifacts.

- **'Biocuration' dataset** http://dx.doi.org/10.5066/F79Z930Q
  - 96 Landsat 8 Scenes with manually generated cloud masks.
  - 8 Biomes x 3 cloud conditions x 4 examples.
  - Based on International Geosphere Biosphere Program (IGBP) classifications.

- **'SPARCS' dataset** http://dx.doi.org/10.5066/F7FQ9TQN
  - 80 Landsat 8 image chips
  - Created and maintained by University of Oregon

Distribution of Cloud Validation datasets



Validation results (as of 2015)

		Total Correct	Total False	Total Ambig.	Misclassified Clouds (Omission)	Misclassified Clear (Comission)
Biocuration	CFMask	89.49%	9.24%	1.28%	3.86%	11.59%
	See5	82.05%	15.84%	2.12%	27.33%	5.38%
	ACCA	84.52%	7.58%	7.90%	4.73%	8.18%
	ATACCA	88.81%	10.51%	0.68%	5.87%	11.24%
Irish	CFMask	87.88%	10.39%	1.73%	2.29%	15.01%
	See5	88.89%	8.29%	2.81%	3.92%	6.06%
	ACCA	83.31%	5.41%	11.28%	8.08%	3.96%
	ATACCA	89.80%	8.50%	1.70%	12.30%	6.30%
SPARCS	CFMask	92.54%	5.66%	1.80%	19.19%	2.41%
	See5	89.44%	7.34%	3.22%	32.21%	1.55%
	ACCA	80.37%	4.21%	15.42%	19.51%	0.54%
	ATACCA	89.92%	8.83%	1.25%	30.87%	3.53%

CCA Strategies

These results illustrate the three main strategies for cloud detection:

	Total Correct	Total False	Total Ambig.	Misclassified Clouds (Omission)	Misclassified Clear (Comission)
CFMask	89.49%	9.24%	1.28%	3.86%	11.59%
See5	82.05%	15.84%	2.12%	27.33%	5.38%
ACCA	84.52%	7.58%	7.90%	4.73%	8.18%

CFMask is *aggressive*, and tends to classify more pixels as cloud.
See5 is *conservative*, and tends to classify more pixels as clear.
ACCA is *indecisive*, and tends to classify more pixels as ambiguous/undetermined.

The response from the remote sensing community heavily favors an aggressive algorithm, which excludes most cloudy data (but may toss away a few good data points).