Web-enabled Landsat data (WELD): a consistent seamless near real time MODIS-Landsat data fusion for the terrestrial user community

a NASA Making Earth System data records for Use in Research Environments (MEASURES) funded project

David Roy & Junchang Ju

david.roy@sdstate.edu junchang.ju@sdstate.edu

Geographic Information Science Center of Excellence

South Dakota State University (SDSU), USA

Landsat Science Team Meeting

U.S. Forest Service, NRRC Building E - Fort Collins, Colorado, January 6-8, 2009

What the Landsat user community ultimately wants …

- Free & derived Landsat data products
- Systematic, consistent, community endorsed data processing
  - calibration, geolocation
  - radiometric normalization / BRDF correction, atmospheric correction
  - cloud-screened, snow-screened, SLC-off gap filling
  - (without this bio/geophysical products cannot be generated reliably e.g., LAI)
- Composited large-area mosaics
  - updated at the pixel level
  - using all the Landsat data, not just select acquisitions
- Processed shortly after acquisition “near real time”
- *Similar to the NASA MODIS land products but at high spatial resolution*
- *Above is what this 5 year NASA funded project is seeking to achieve, building on our 10 year MODIS Land product development, processing (and reprocessing) experience.*

This presentation is a 9 month progress report

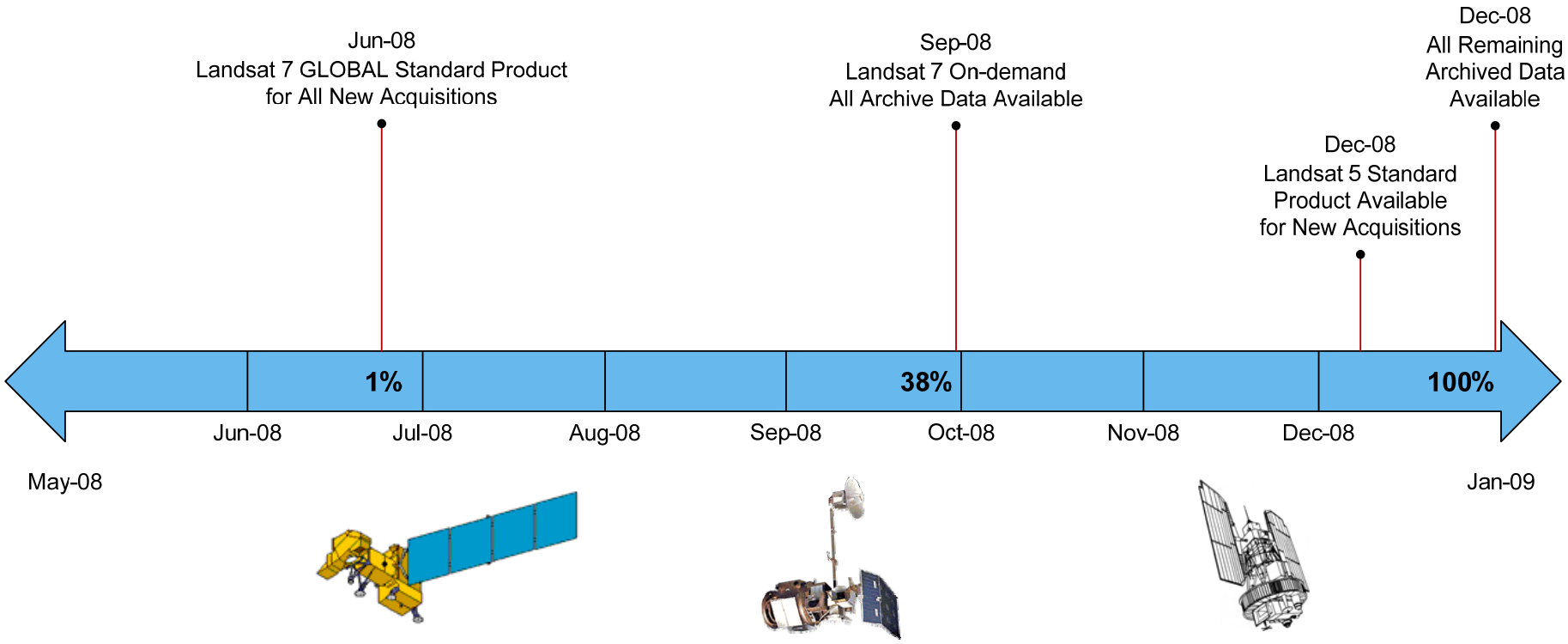
- Project funding started in Spring 2008
- Seek to generate 30m mosaic temporally composited products
  - CONUS (conterminous USA) & Alaska
  - 7 year period
  - Make freely available to the user community
- Seek to develop a production system that will be ported to EROS in year 3
- Improve the consistency & quality of ETM+ data processing through a fusion with NASA MODIS land products.
- In the last 9 months have been putting processing elements together to protoype system, emphasis on CONUS data

Proposed Weld Products (may change after community feedback)
ETM+ composited mosaics for all CONUS & Alaska, 7 years

Monthly Composite	Seasonal (3 month) Composite	Annual Composite
Surface reflectance: 6 ETM+ 30m reflective λ bands	*As monthly but no land cover characterization*	*As monthly*
Brightness temperature: 2 ETM+ resampled 30m thermal λ bands		
NDVI		
Band saturation		
Cloud masks: ACCA & classificiation tree		
Day of year selected		
Number of observations considered		
QA & algorithm processing path		
Land cover characterization: % Tree, % Herbaceous, % H_20 % Bare Ground, %Snow / Ice		

Products will be accessible via the internet and updated at the pixel level as the ETM+ data are acquired and processed

US Landsat Web-Enabled Data Availability Schedule



- Only LTAP acquisitions with cloud cover <= 40% due to resource constraints
- WELD project has a dedicated ftp link to the Landsat archive

Kristi Kline, "Free Data Lady", USGS EROS

SDSU WELD CONUS ETM+ L1T ARCHIVE
(as of January 1st 2008)

- Automated FTP-Cron job every Friday night, pulls L1T data from EROS to SDSU via a dedicated EROS-SDSU gigabit network

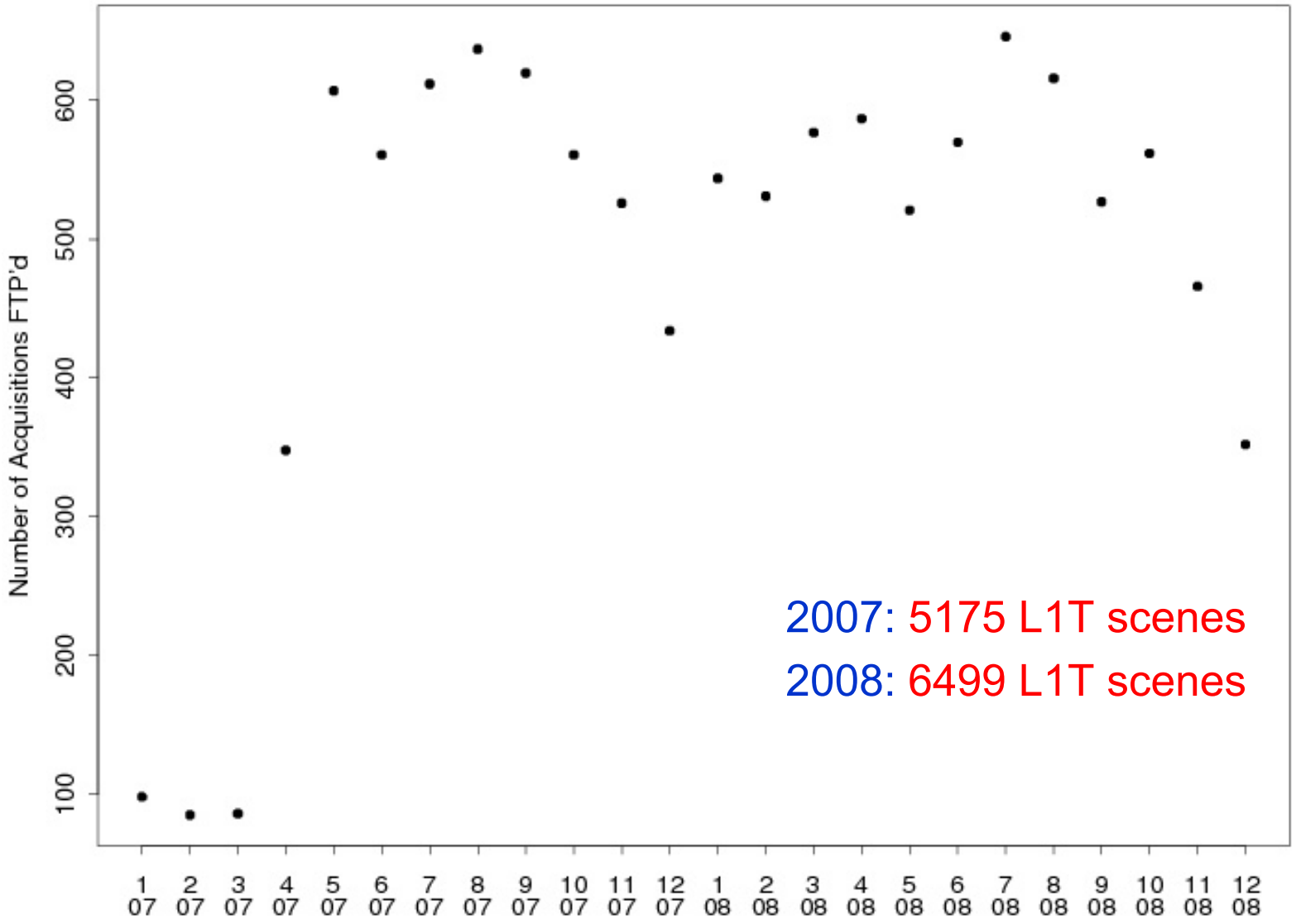
- 455 CONUS ETM+ path/rows

- 2007: 5175 L1T scenes
- 2008: 6499 L1T scenes

~ 1.2 TB / year externally compressed GeoTIFF

~ 1.1 TB / year internally compressed HDF

EROS – SDSU WELD
FTP CONUS DATA ARCHIVE



WELD ETM+ Data Processing Steps

- TOA reflectance & brightness temperature
- Cloud mask
- SLC-Off and cloud gap filling
- Reprojection: UTM to continental map proj.
- Compositing: monthly, seasonal, annual
- Atmospheric correction
- Land cover characterization
- Browse generation

WELD ETM+ Data Processing Steps

Steps that use contemporaneous MODIS Products

- TOA reflectance & brightness temperature
- Cloud mask
- SLC-Off and cloud gap filling **(Jan 08 meeting)**
- Reprojection: UTM to continental map proj.
- Compositing: monthly, seasonal, annual
- Atmospheric correction **(Eric Vermote)**
- Land cover characterization **(Matt Hansen)**
- Browse generation

Big data volume / processing requirements

- ~ 11,000,000,000 30m land pixels CONUS
- ~ 150, 000,000 1km land pixels GLOBAL

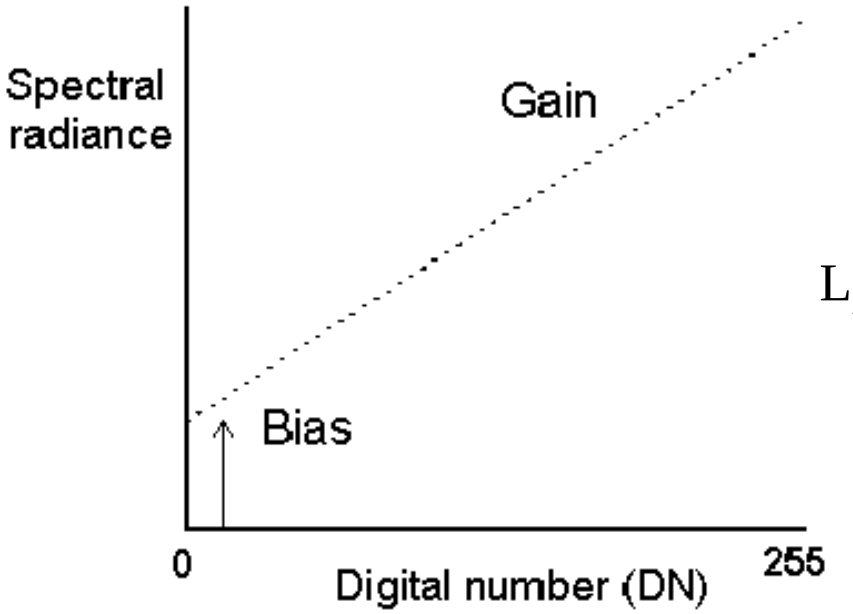
Year 1 hardware

- **Dell PowerEdge R900 server**
  - 4 Quad Core X7350 2.9GHz Xeon processors
  - 8MB Cache / processor
  - 32 GB RAM
  - RAID1 home directory
  - 4Gbits/second fiberoptic SAN
- **T-4000 SATA RAID storage**
  - 8 Fiber ports per controller
  - 5 GB of Cache
  - 48TB useable at RAID5
- Run: Linux, C
- **1 week to process 1 year of CONUS data (black steps on previous slide)**

This Presentation: Processing Steps in Black

- TOA reflectance & brightness temperature
- Cloud mask
- SLC-Off and cloud gap filling **(Jan 08 meeting)**
- Reprojection: UTM to continental map proj.
- Compositing: monthly, seasonal, annual
- Atmospheric correction **(Eric Vermote)**
- Land cover characterization **(Matt Hansen)**
- Browse generation

TOA Reflectance & Brightness Temperature generated for every non-fill pixel



L_λ = Bias + Gain x Digital Number

$$\rho_\lambda = \frac{d^2 \pi L}{E_0 \cos(\theta_s)}$$ *allow* $\rho_\lambda < 0$ *and* $\rho_\lambda > 1$ bands 1,2,3,4,5,7

$$BT = \frac{k_2}{\log(1 + k_1 / L)}$$ k_2 = 1282.71 K, k_1 = 666.09 W/(m^2.sr.μm) bands 61, 62

Cloud Masking

- ACCA algorithm (code provided by GSFC group)

- Supervised classification tree algorithm, developed recognizing that
  - ACCA was not intended to provide per-pixel cloud mask
  - supervised classification ‘aint hard science but it works

Cloud classification tree: Training data

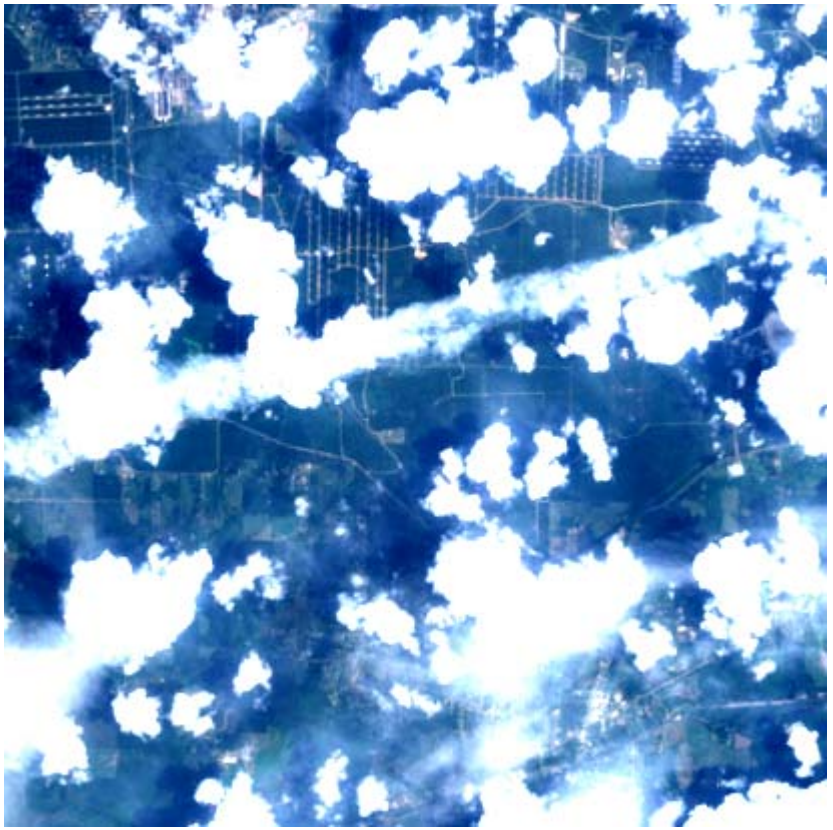
- Landsat ETM+ cloud training scenes provided by Pat Scaramuzza (EROS LDCM cloud task)

- 88 northern hemisphere training scenes
  - polar (19)
  - boreal (22)
  - mid-latitude (24)
  - sub-tropical zones (23)

- 0.5% random sample extrated from each scene

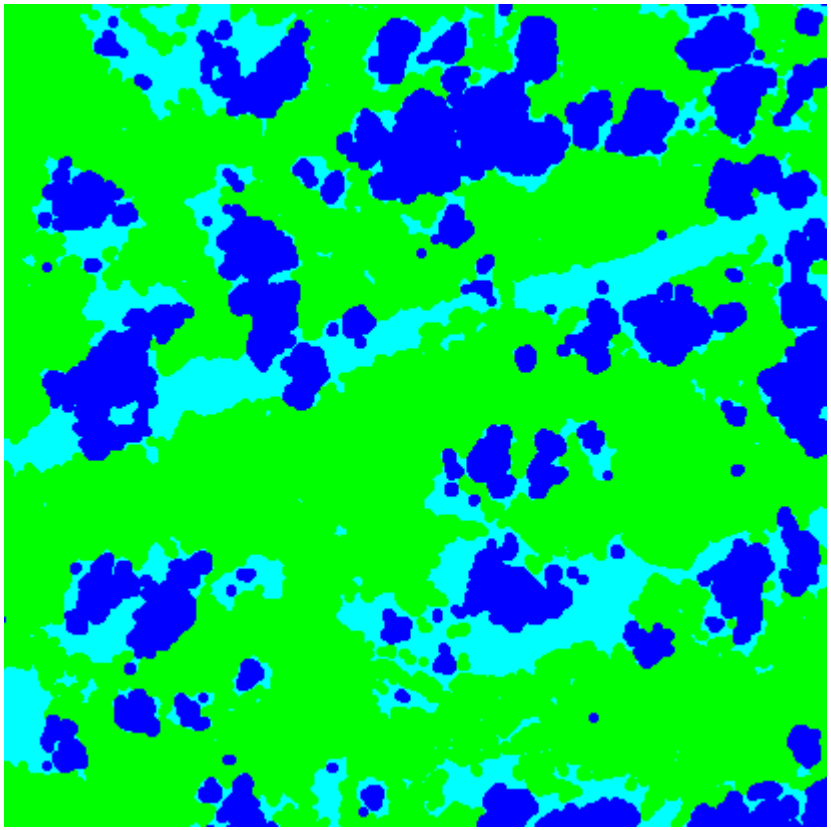
- ~18 million training pixels

Example Cloud Training Data
(400x400 30m pixels)



True Color TOA Reflectance

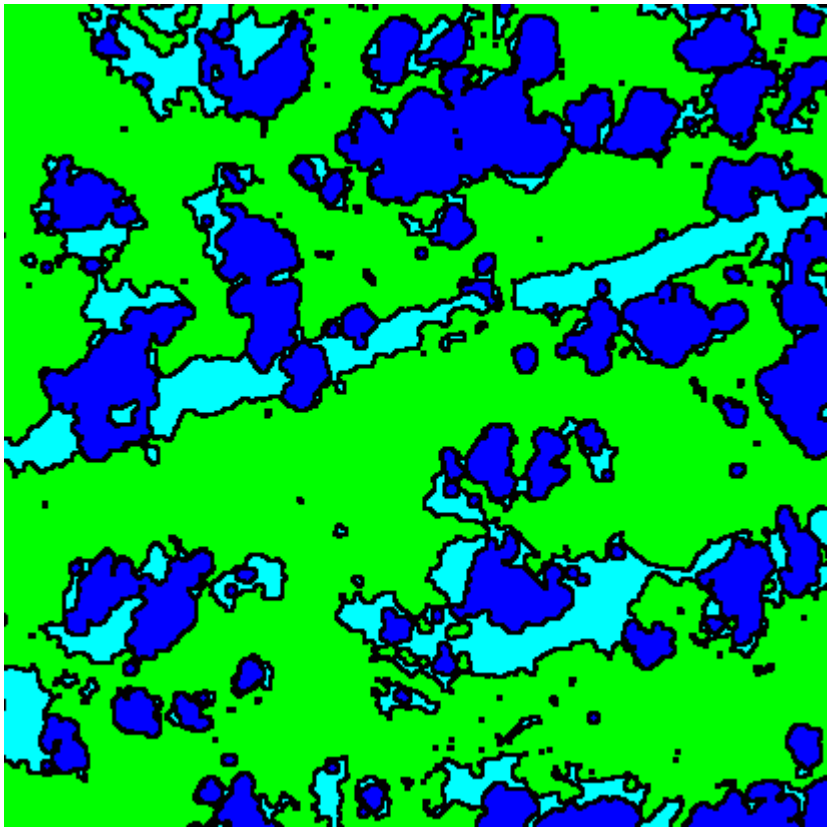
Example Cloud Training Data
(400x400 30m pixels)



Cloud Interpretation

Dark Blue = Thick cloud
Light Blue = Thin Cloud
Green = Not cloudy

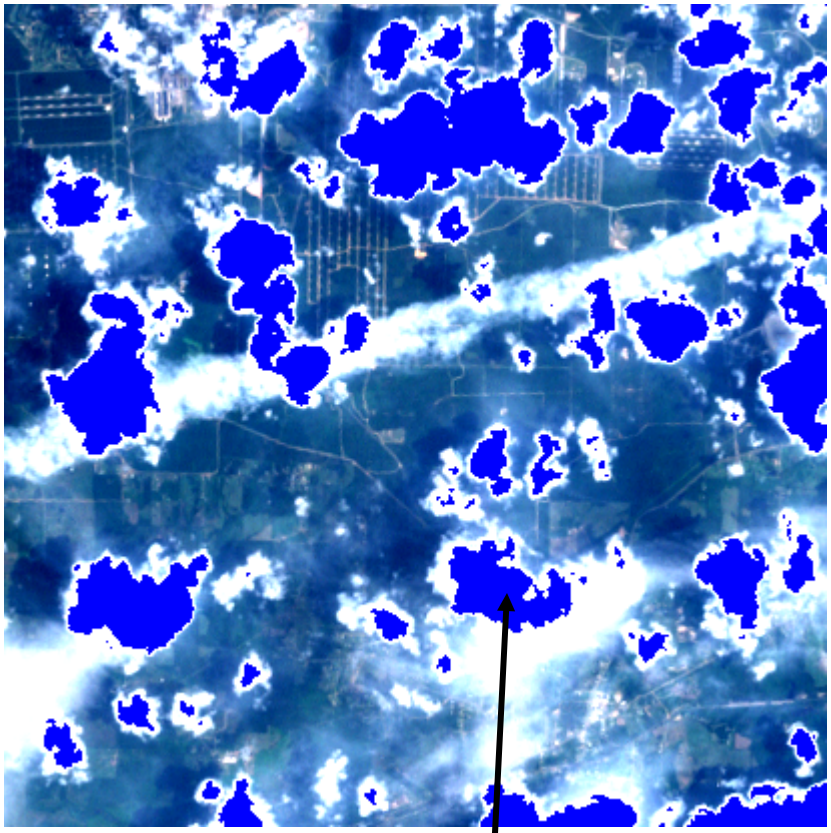
Example Cloud Training Data
(400x400 30m pixels)



Black = morphologically eroded buffer pixels not sampled
Dark Blue = Thick cloud
Light Blue = Thin Cloud
Green = Not cloudy

Need to handle Saturation

(400x400 30m pixels)



Cloud saturated reflectance
(DN=255)

2 Cloud Classification Trees

- Saturated Tree Variables
  - all TOA reflectance bands except shortest λ blue band
  - brightness temperature b61

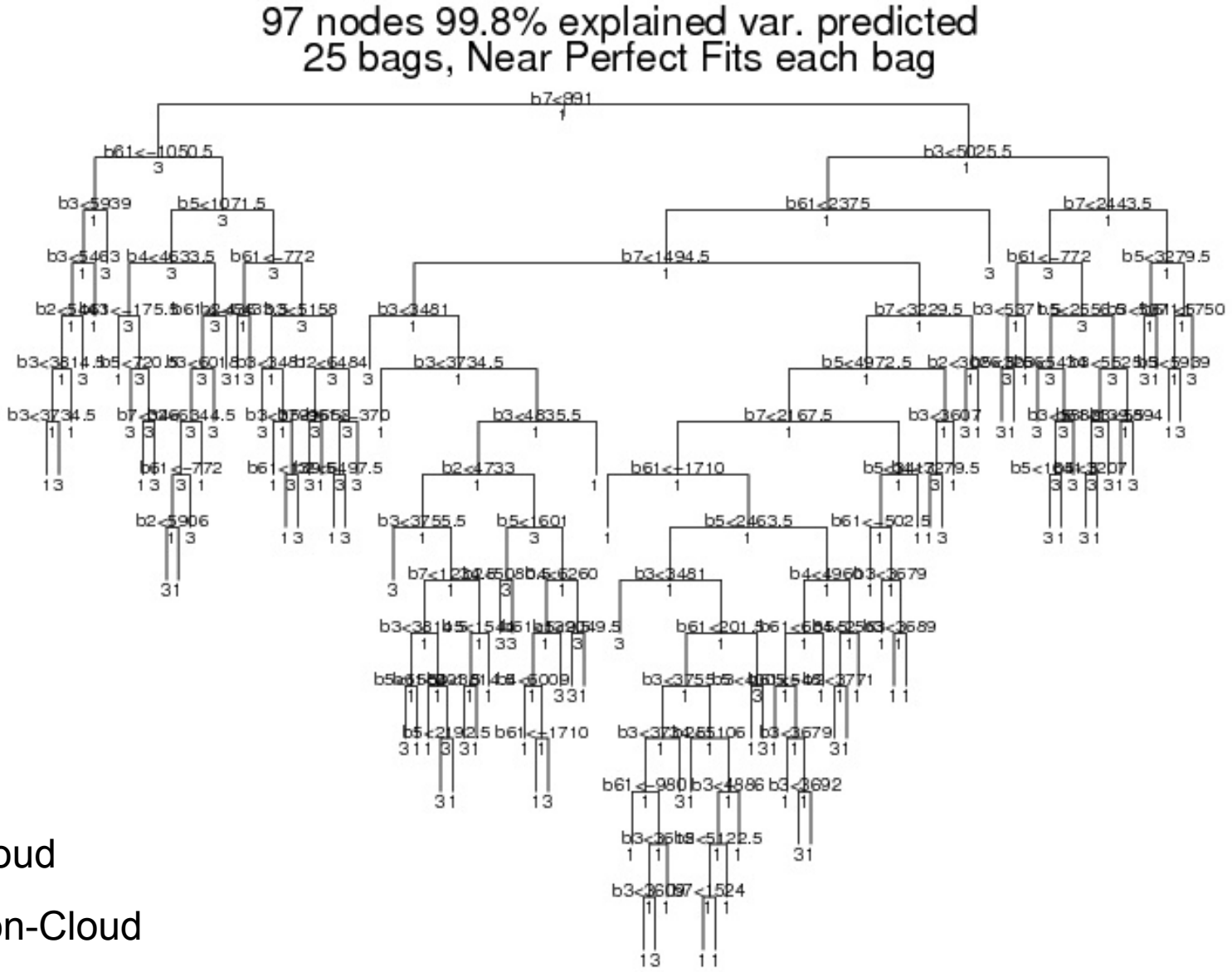
```
var <- data[, c("b2", "b3", "b4", "b5", "b61", "b7")]
```

- Unsaturated Tree Variables
  - ACCA heritage & from discussion with Pat Scaramuzza

```
ndsi <- (data[, "b2"] - data[, "b5"])/(data[, "b2"] + data[, "b5"])
nd35 <- (data[, "b3"] - data[, "b5"])/(data[, "b3"] + data[, "b5"])
nd37 <- (data[, "b3"] - data[, "b7"])/(data[, "b3"] + data[, "b7"])
nd42 <- (data[, "b4"] - data[, "b2"])/(data[, "b4"] + data[, "b2"])
nd43 <- (data[, "b4"] - data[, "b3"])/(data[, "b4"] + data[, "b3"])
nd45 <- (data[, "b4"] - data[, "b5"])/(data[, "b4"] + data[, "b5"])
nd47 <- (data[, "b4"] - data[, "b7"])/(data[, "b4"] + data[, "b7"])
nd57 <- (data[, "b5"] - data[, "b7"])/(data[, "b5"] + data[, "b7"])
var <- cbind(data[, c("b3", "b4", "b5", "b61", "b7")], ndsi, nd57, nd35, nd37, nd42, nd43, nd45, nd47)
```

- 25 bagged runs combined to generate a single parsimonious saturated and unsaturated tree

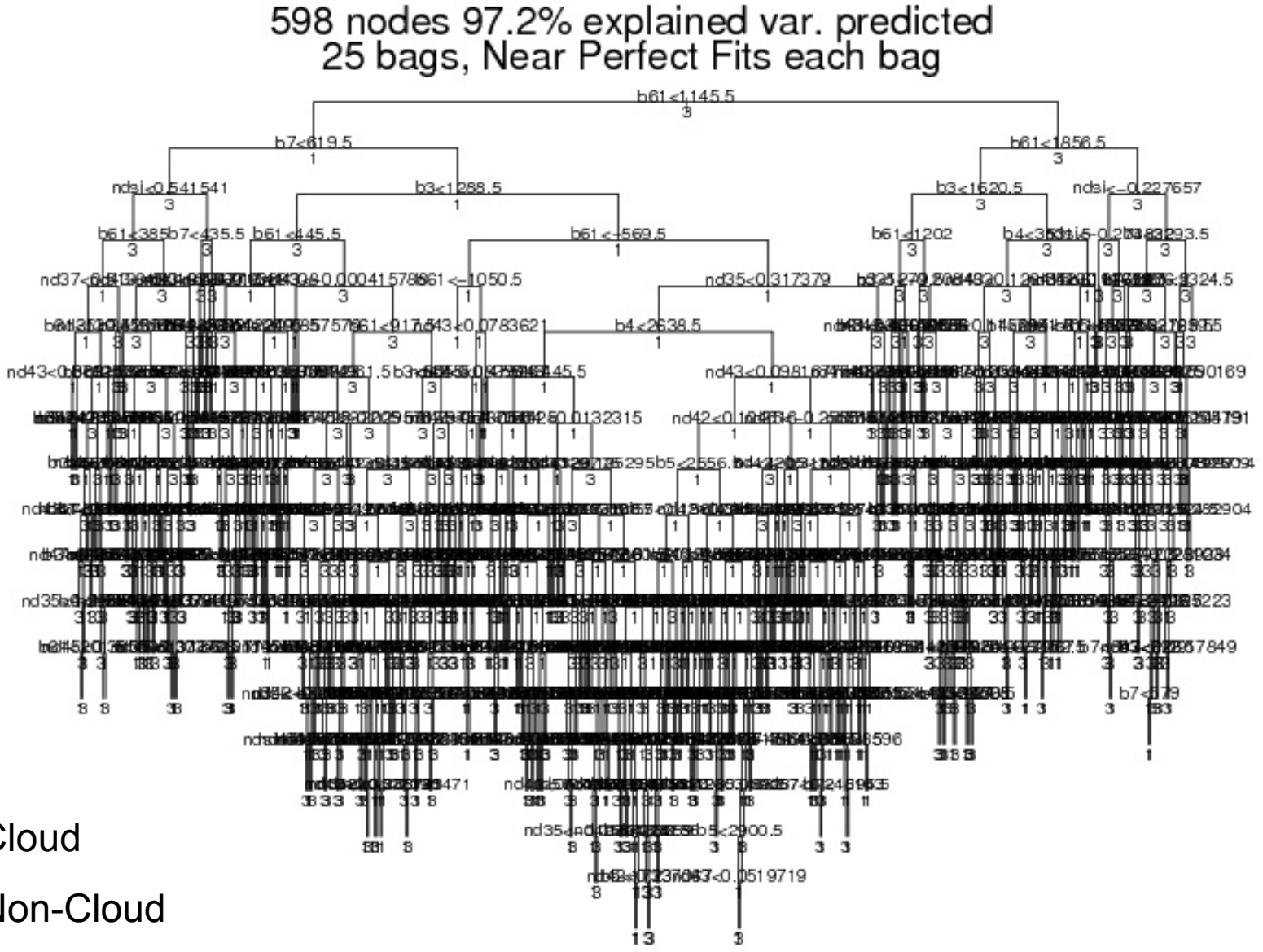
Saturated Cloud/Non-Cloud Tree
5.0 million saturated training pixels



1 = Cloud

3 = Non-Cloud

Unsaturated Cloud/Non-Cloud Tree
12.9 million unsaturated training pixels

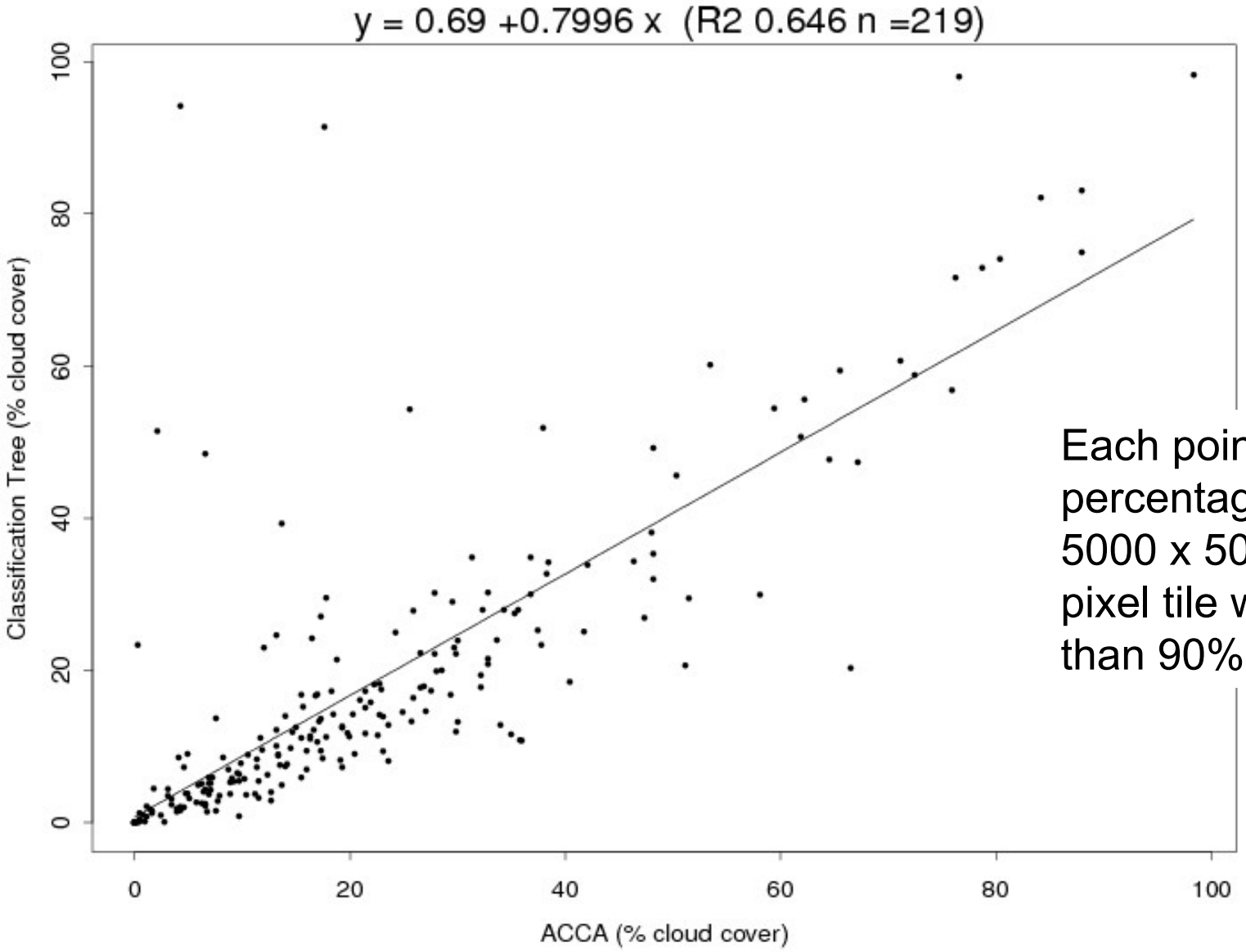


1 = Cloud

3 = Non-Cloud

CONUS cloud comparison

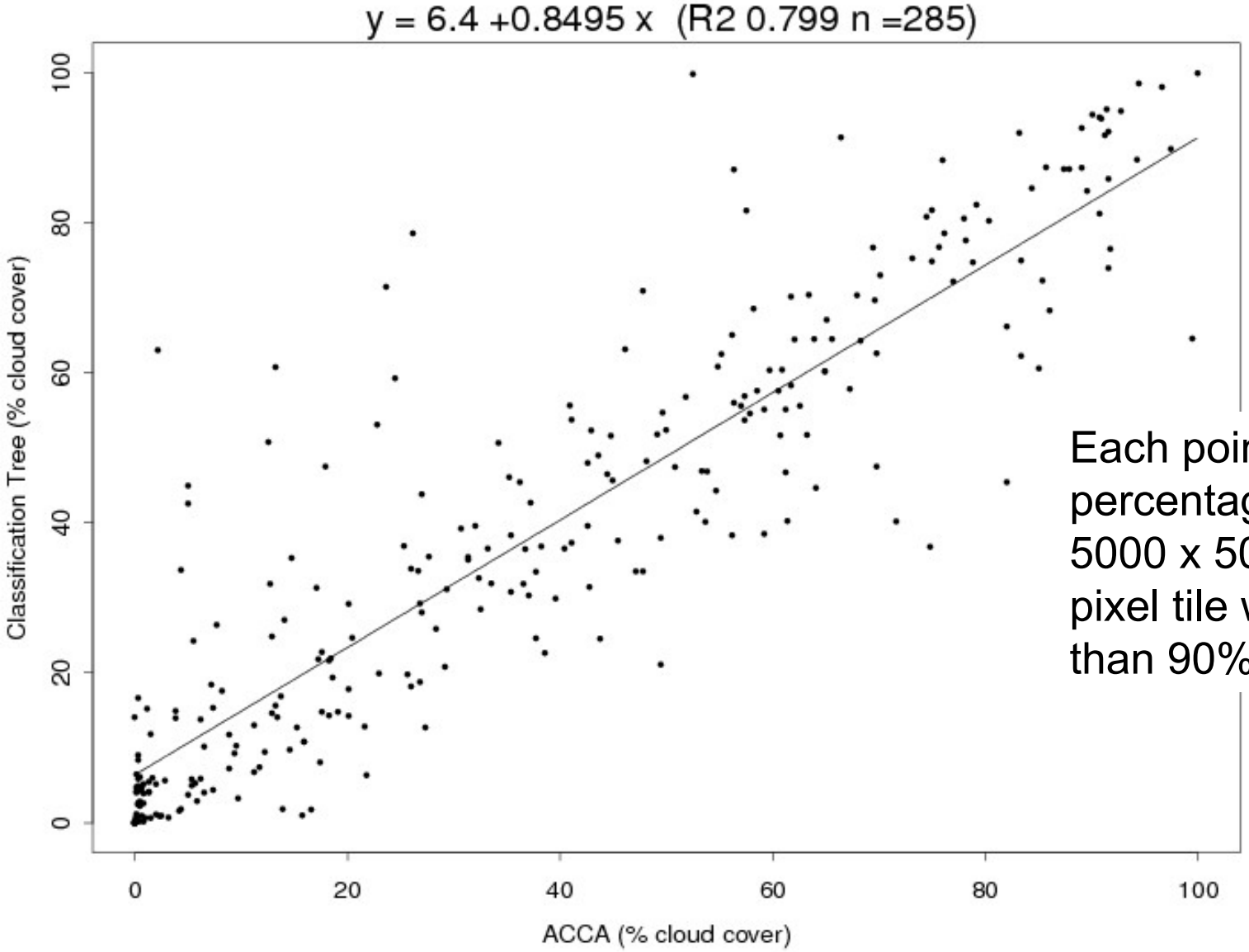
Classification Tree - ACCA, July 2008



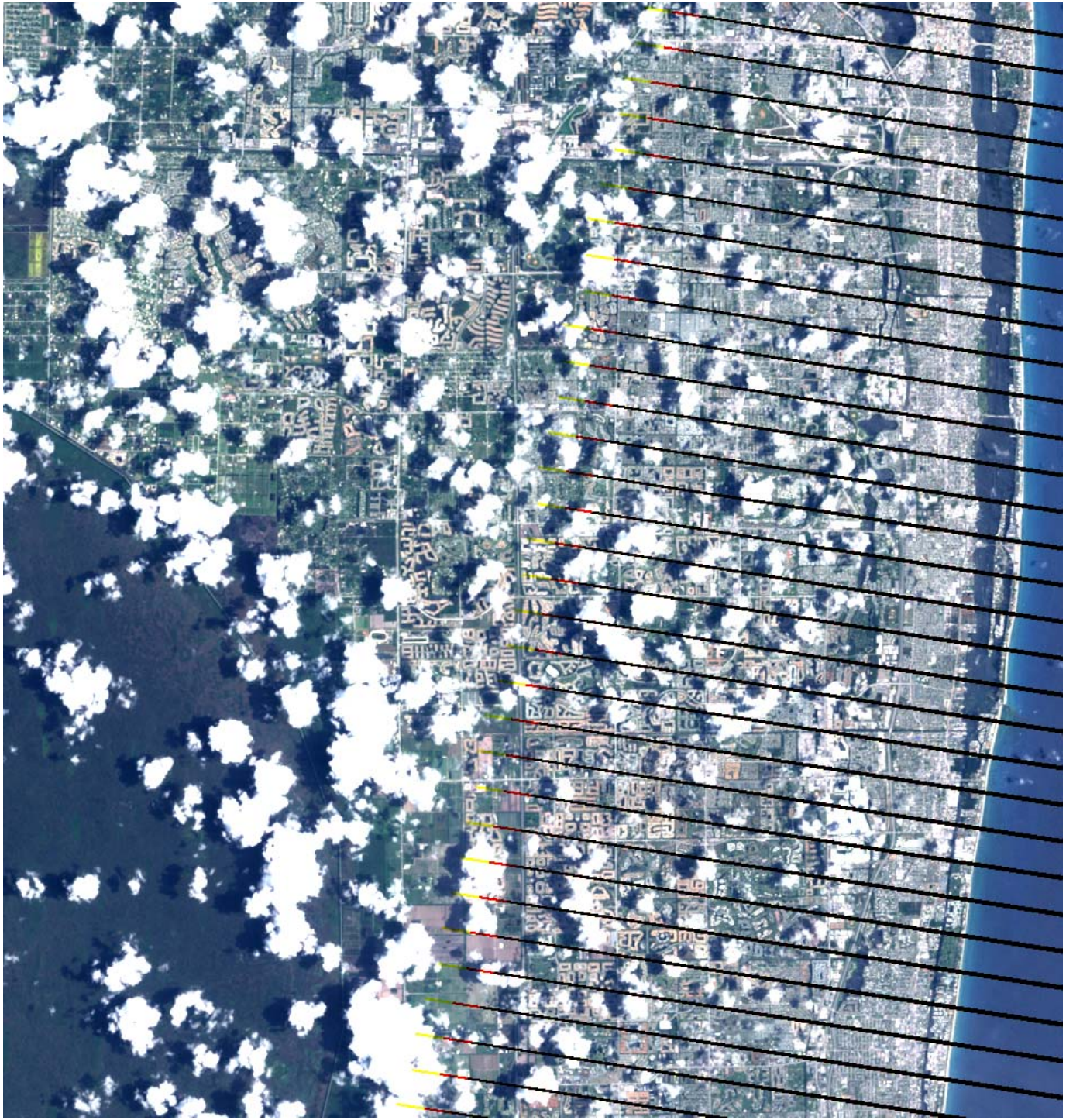
Each point = the percentage of a 5000 x 5000 30m pixel tile with less than 90% fill values

CONUS cloud comparison

Classification Tree - ACCA, Jan. 2008



Each point = the percentage of a 5000 x 5000 30m pixel tile with less than 90% fill values

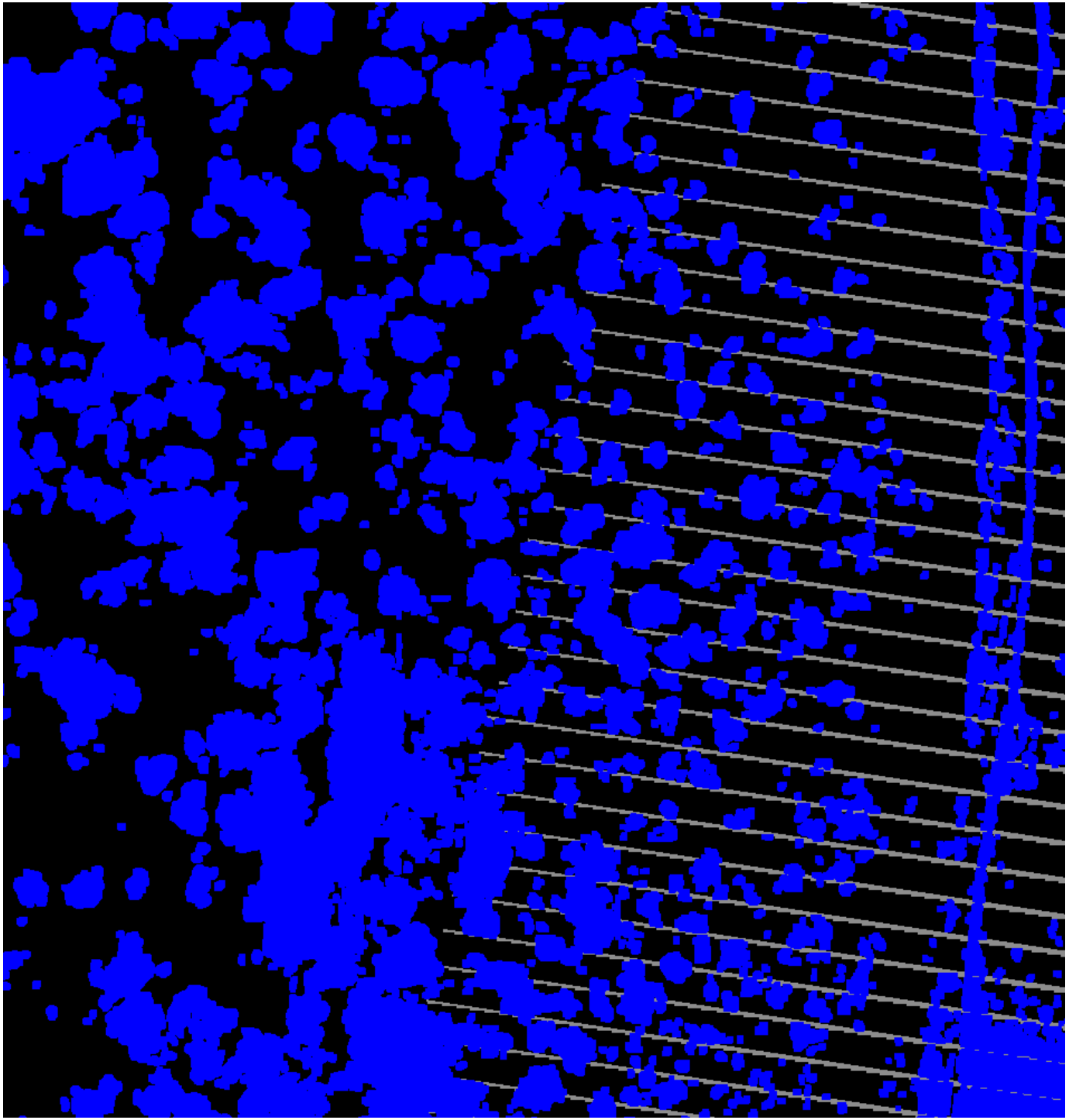


True Color
TOA
Reflectance

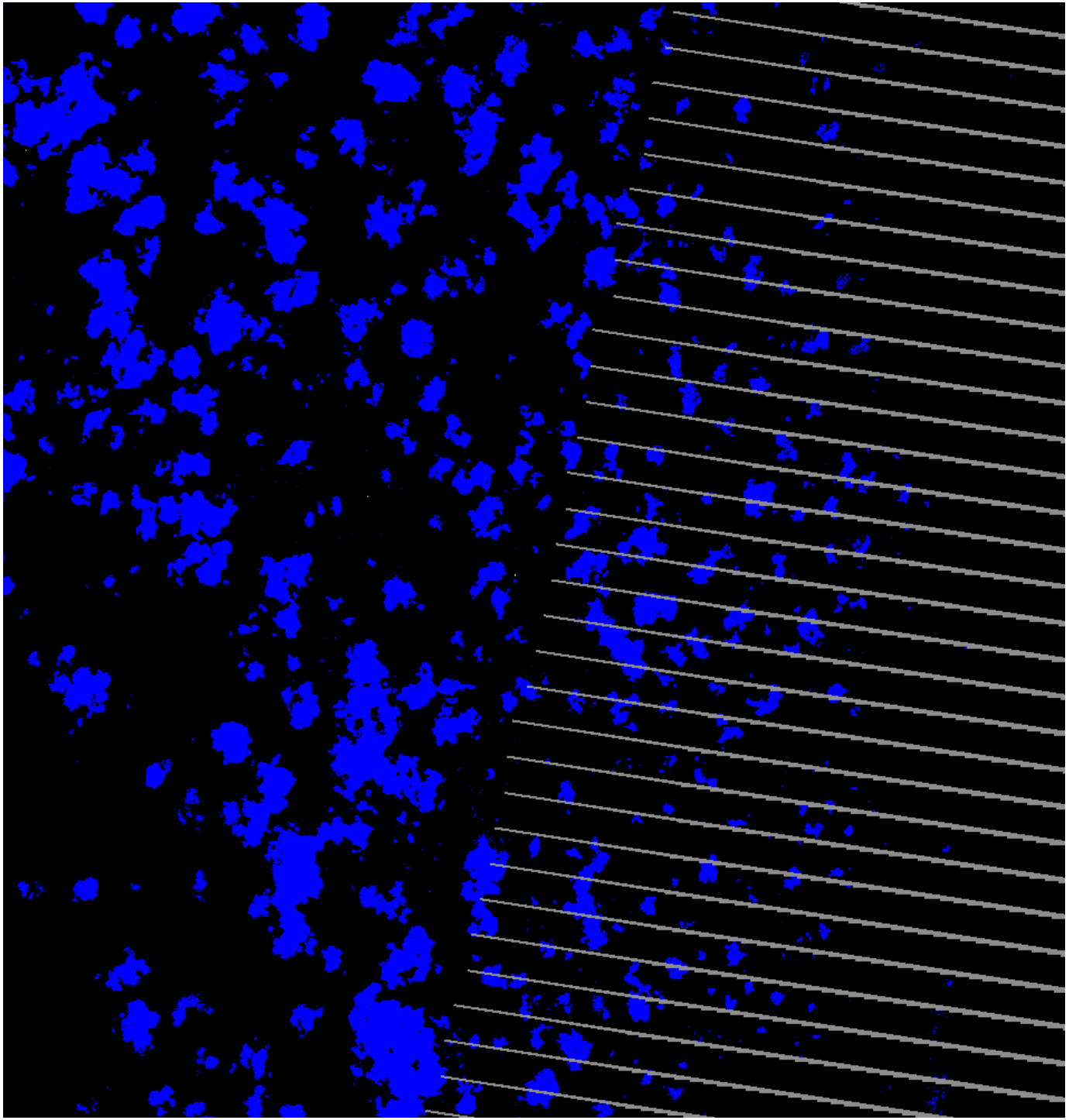
Florida
July 20 '08

Dark cool &
Bright hot
backgrounds

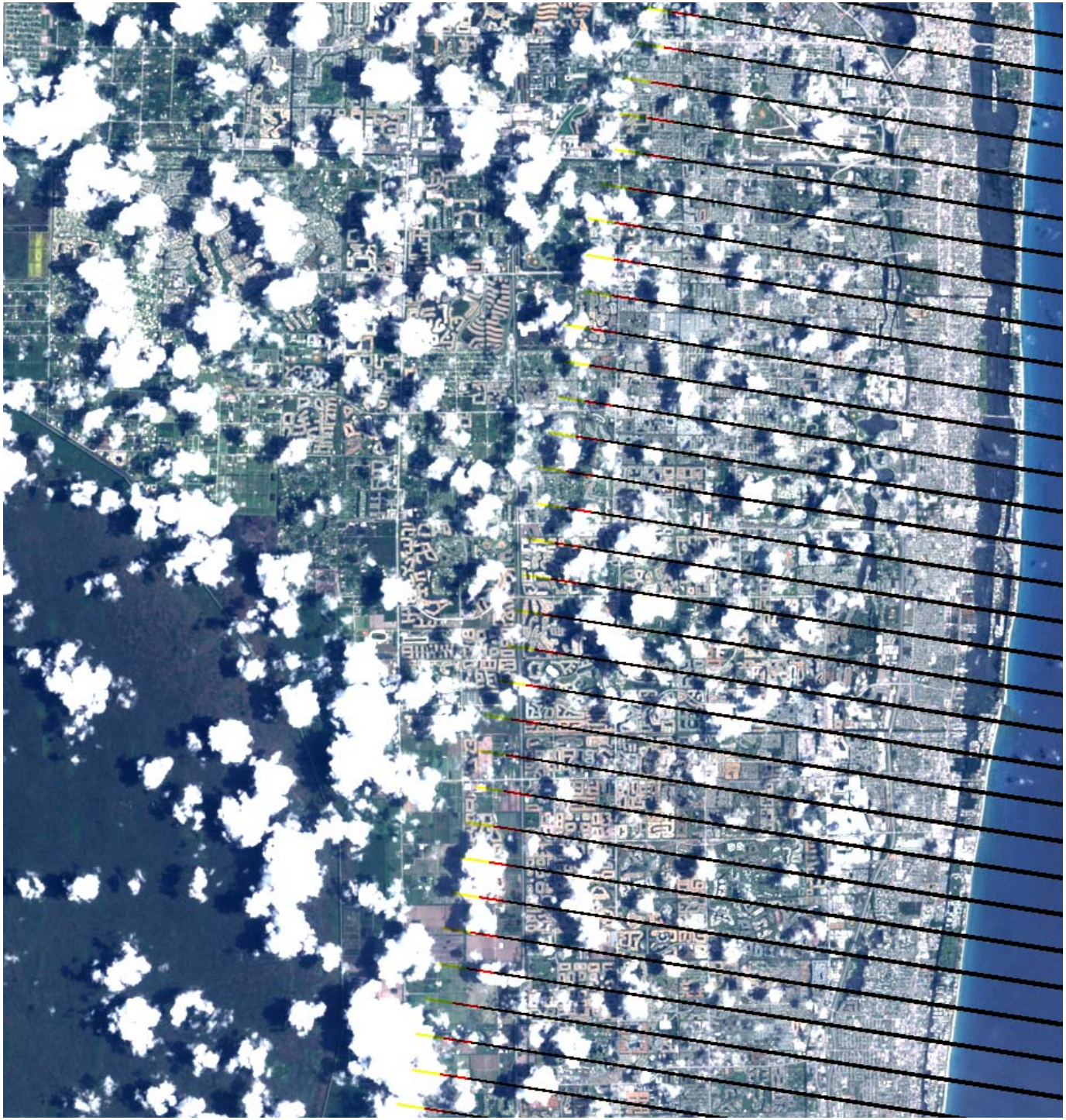
950 x 1000 30m pixels



ACCA



Classification Tree

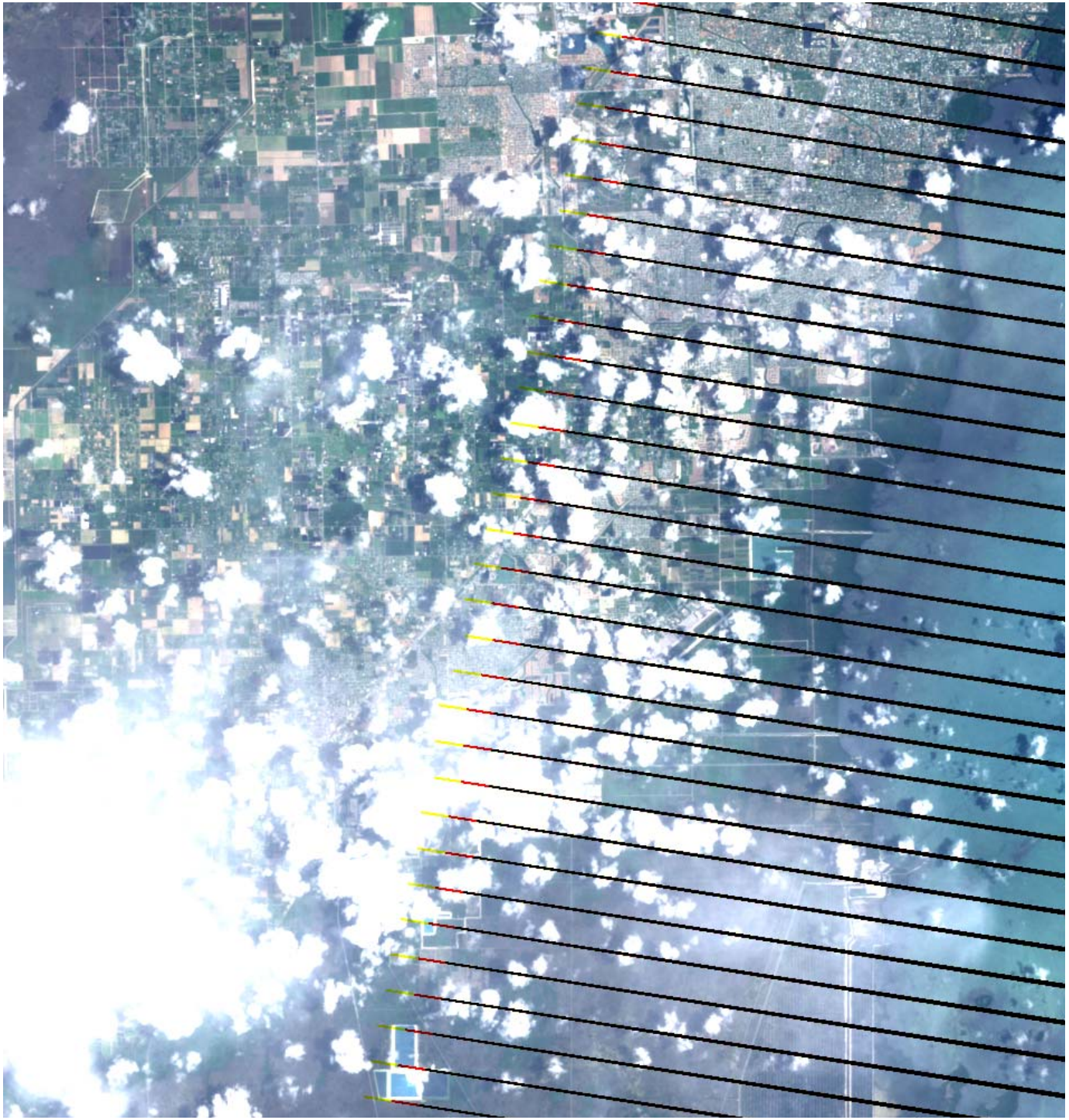


True Color
TOA
Reflectance

Florida
July 20 '08

Dark cool &
Bright hot
backgrounds

950 x 1000 30m pixels

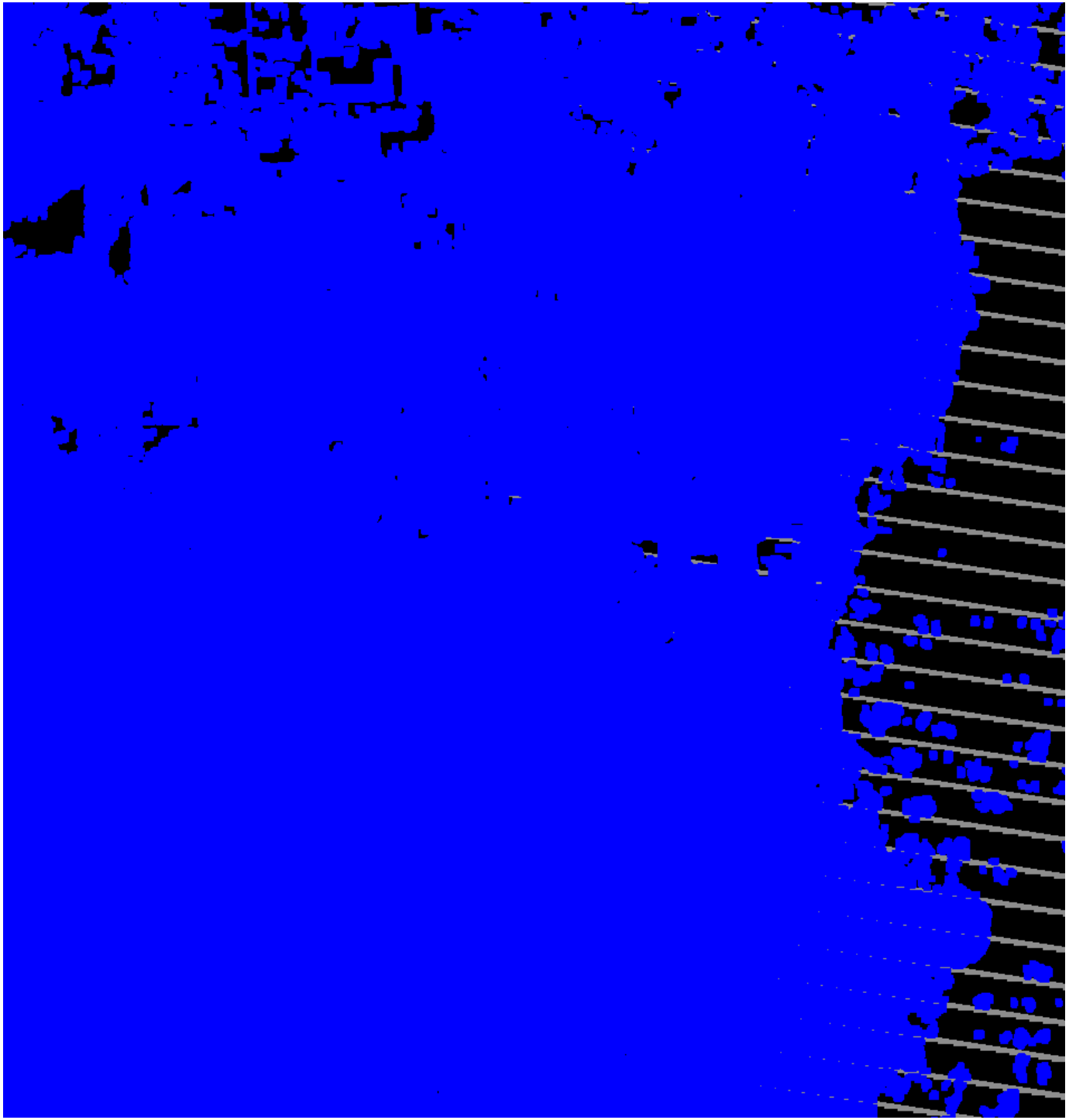


True Color
TOA
Reflectance

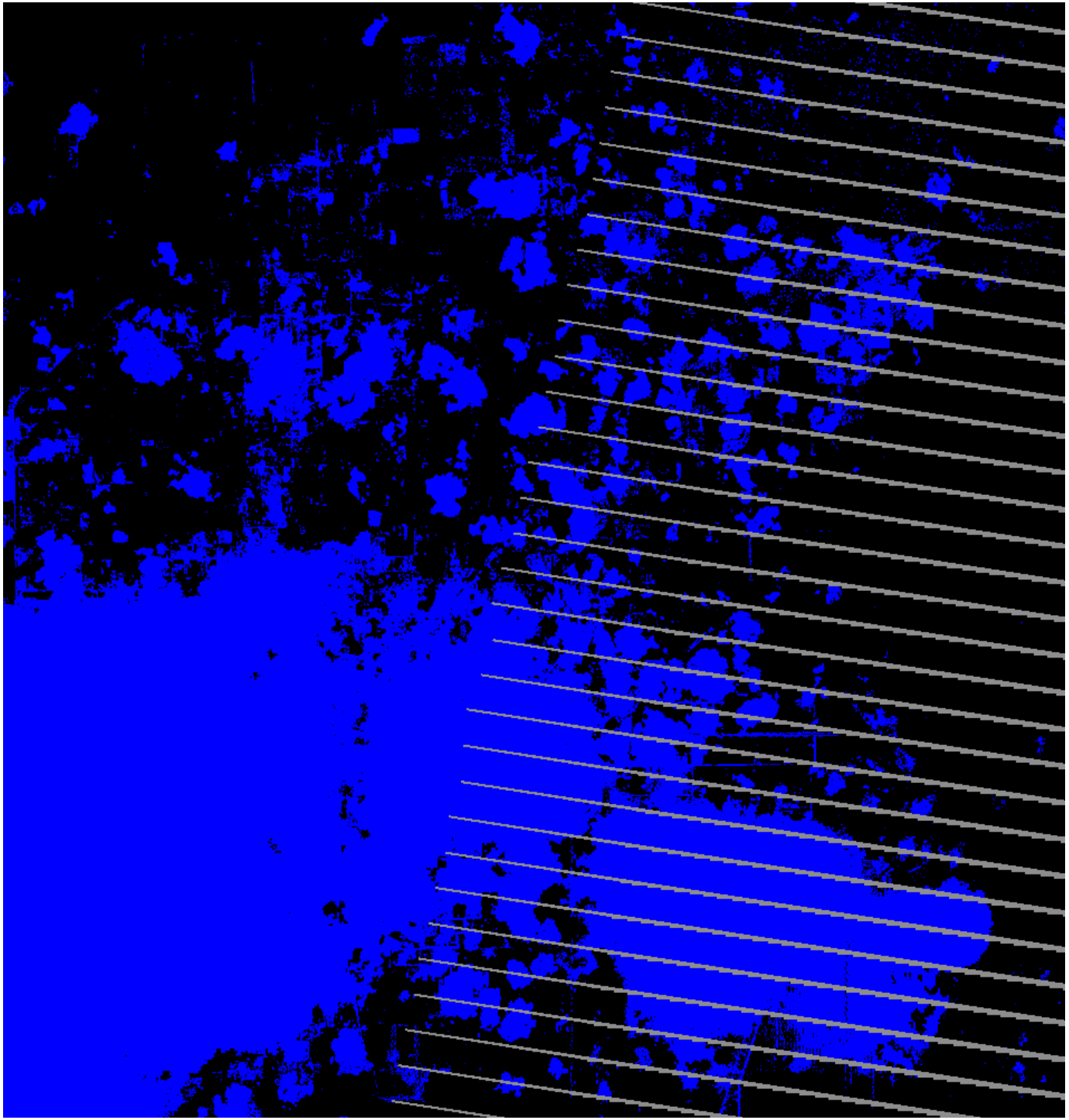
Florida
July 20 '08

Diffuse
cloud

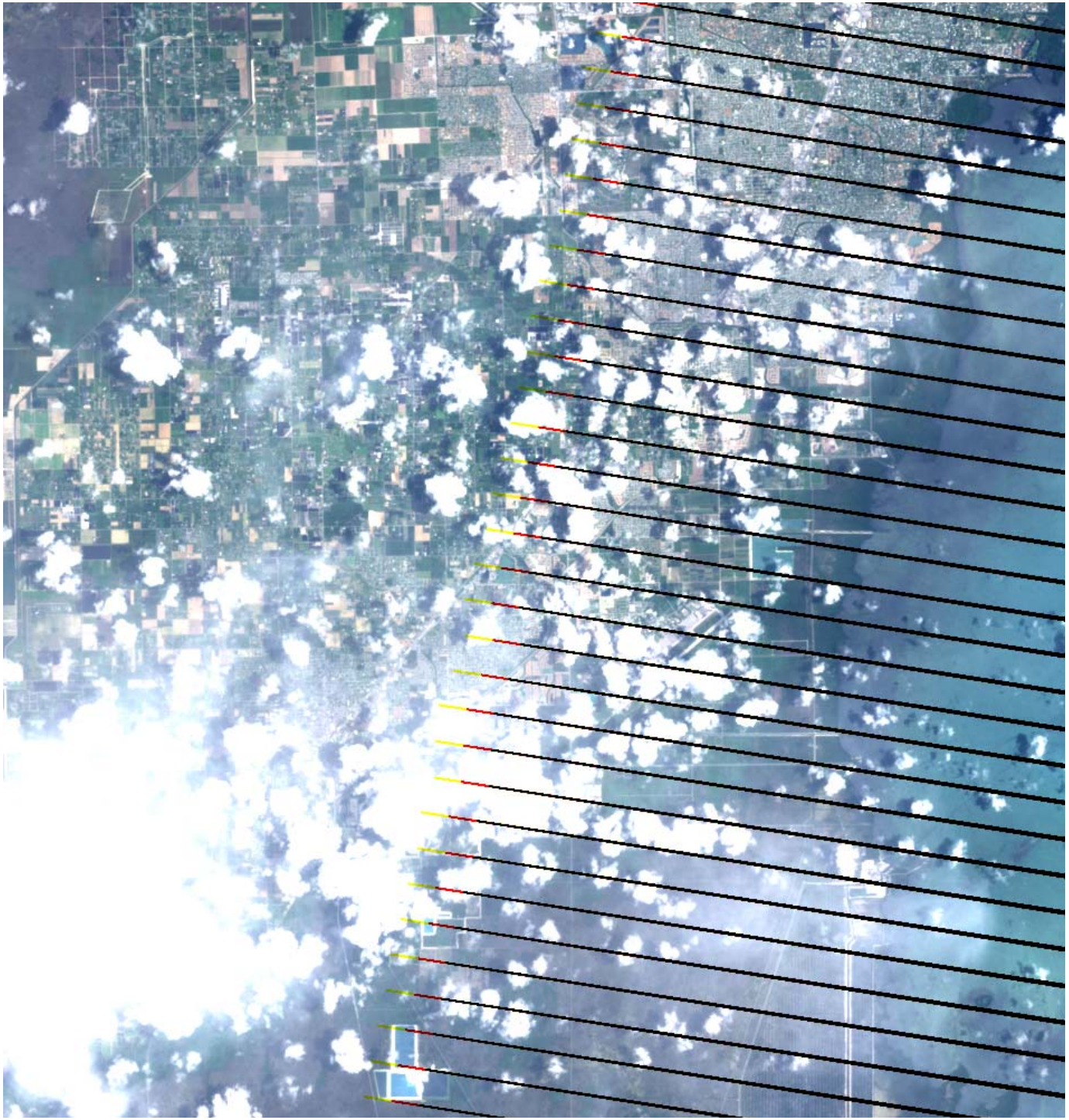
950 x 1000 30m pixels



ACCA



Classification Tree



True Color
TOA
Reflectance

Florida
July 20 '08

Diffuse
cloud

950 x 1000 30m pixels

Reprojection

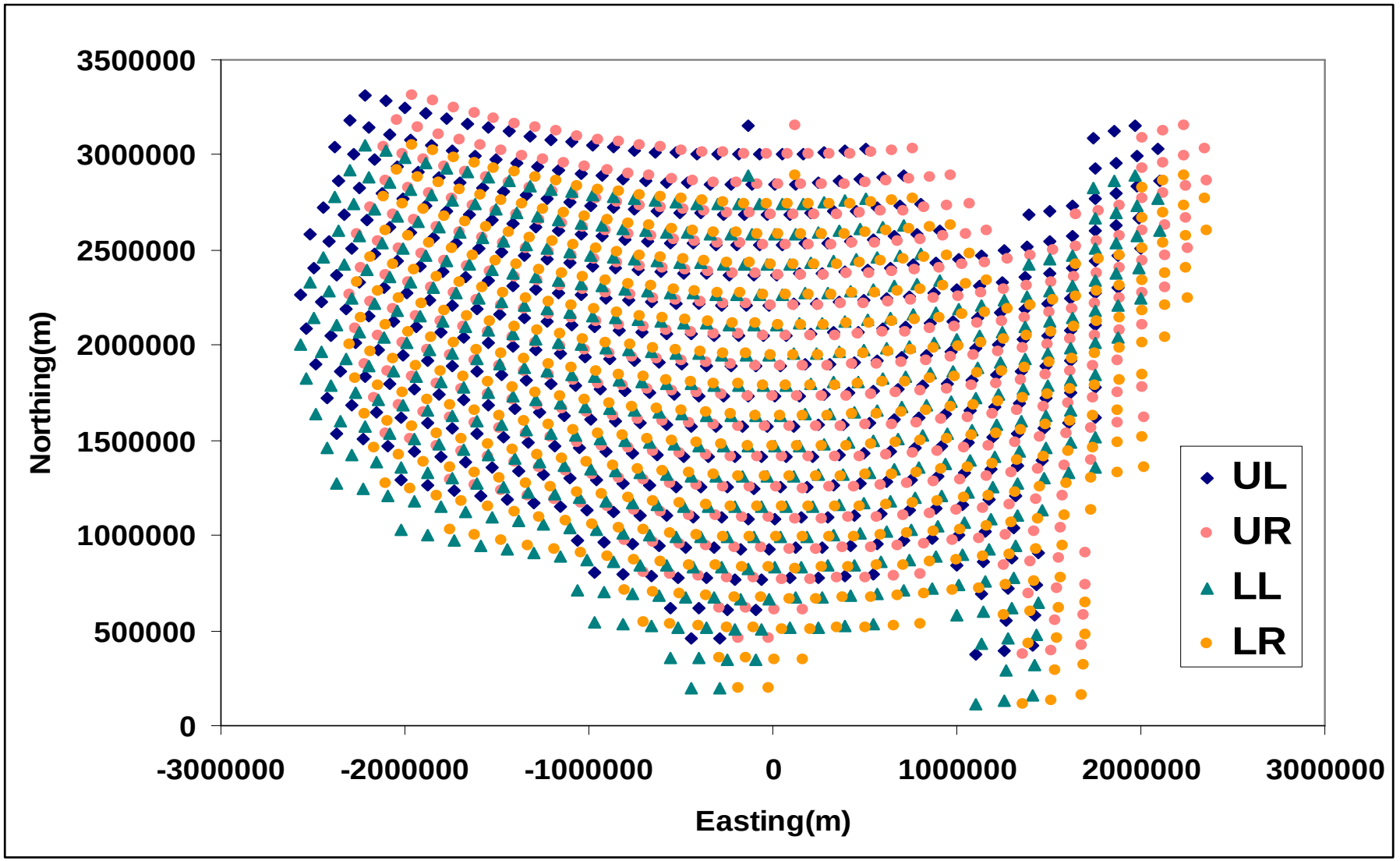
- Project Landsat L1T UTM data to a continental CONUS and Alaskan coordinate system

- Use the Albers projection
  - Suitable for continental projections
  - Equal area
  - USGS EROS NLCD heritage

CONUS	**Alaska**
Map Projection Name: **Albers Conical Equal Area** *First Standard Parallel:* **29.5** *Second Standard Parallel:* **45.5** *Longitude_of_Central_Meridian*: **-96.0** *Latitude of Projection Origin:* **23.0** *False Easting:* **0.0** *False Northing:* **0.0**	*Map Projection Name:* **Albers Conical Equal Area** *First Standard Parallel:* **55.0** *Second Standard Parallel:* **65.0** *Longitude_of_Central_Meridian:* **-154.0** *Latitude of Projection Origin:* **50.0** *False Easting:* **0.0** *False Northing:* **0.0**

- Projection code provided by General Cartographic Transformation Package (GCTP)

CONUS ETM+ scene corners in Albers

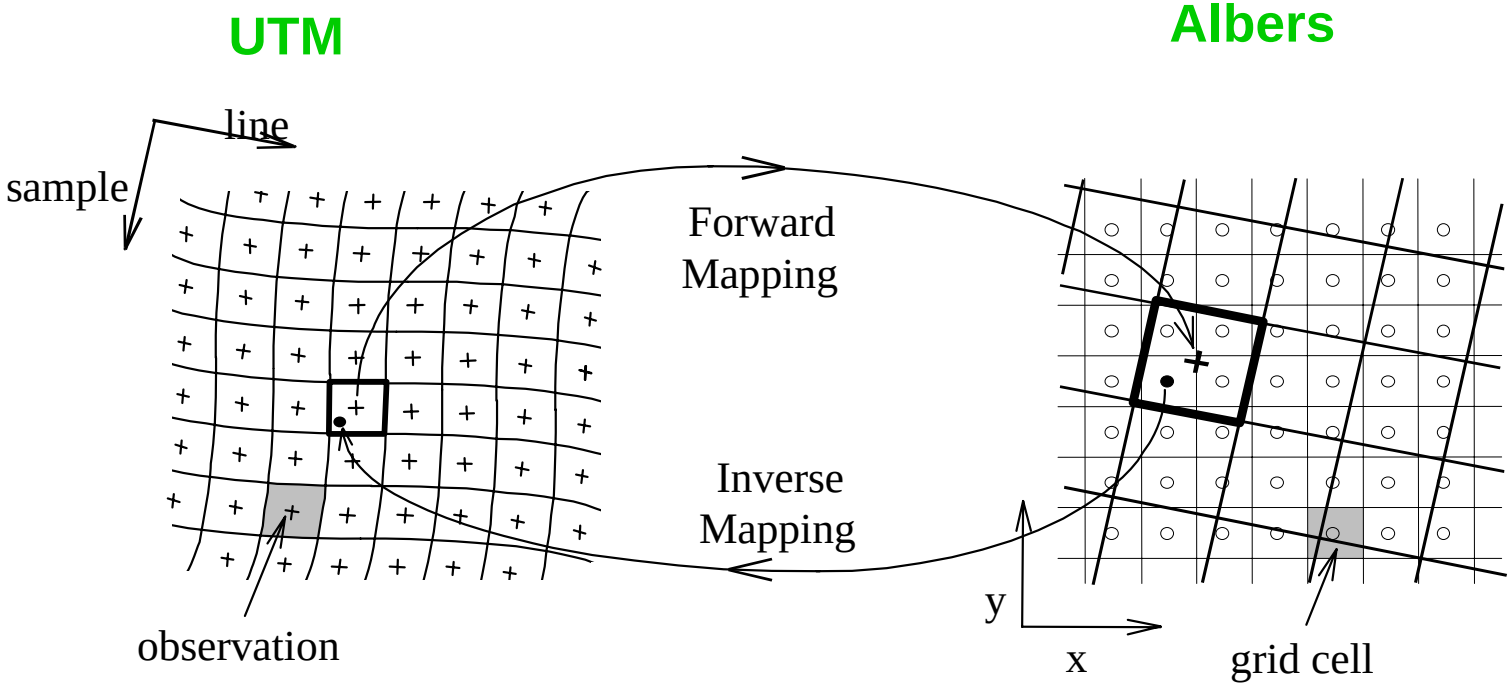


(Val Kovalassky)

Reprojection - 2 mapping approaches

Forward: map ETM+ observations into Albers

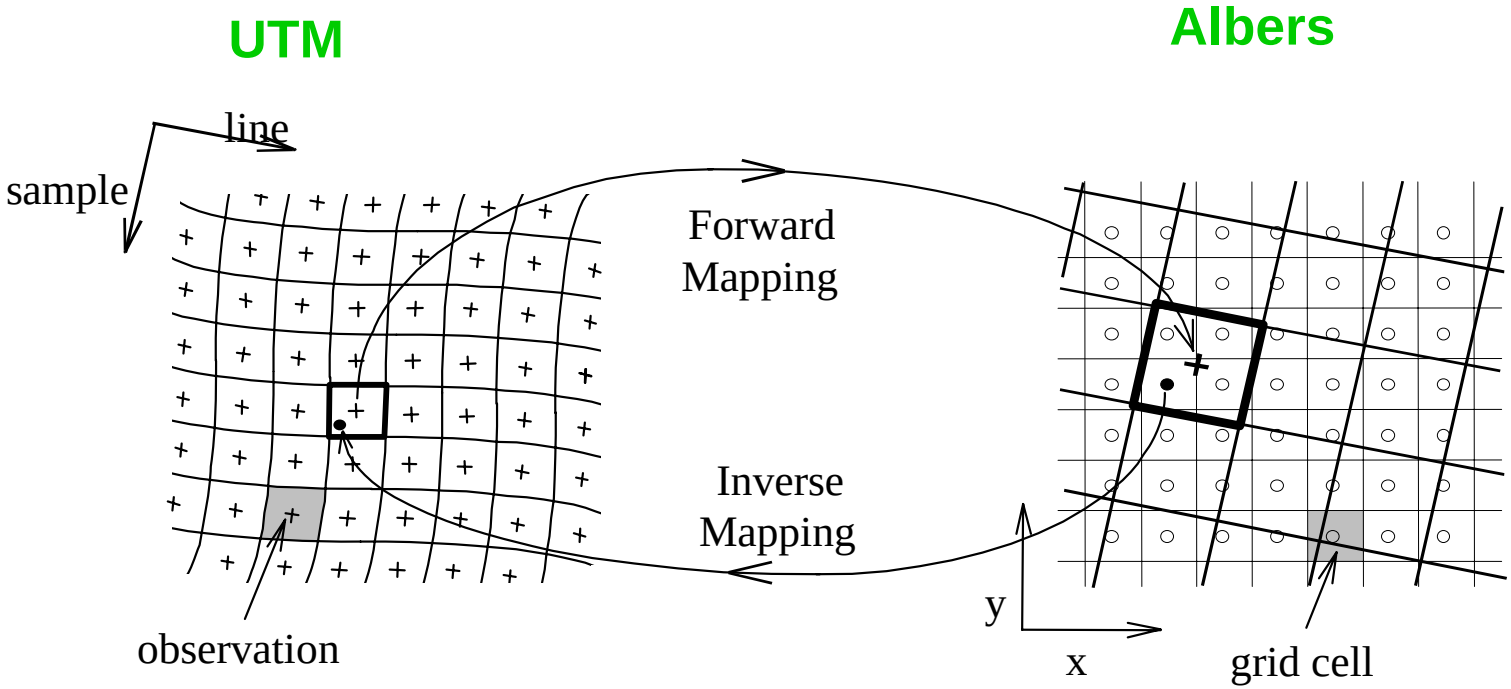
Inverse: map regular grid of Albers coordinates into the ETM+ image



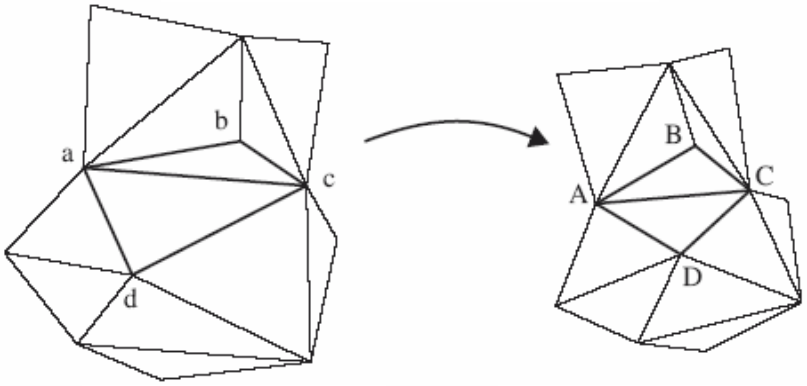
Reprojection: 2 mapping approaches

Forward: map ETM+ observations into Albers

Inverse: map regular grid of Albers coordinates into the ETM+ image *This mapping approach used as computationally less expensive, each Albers pixel only addressed once & no gaps in output.*



Piecewise linear triangle mapping to minimize General Cartographic Transformation Package (GCTP) transformation calls



Reference (Albers)

Image (UTM)

Select points sparsely across Albers

Use GCTP to compute corresponding UTM coordinates

Triangulate the Albers points

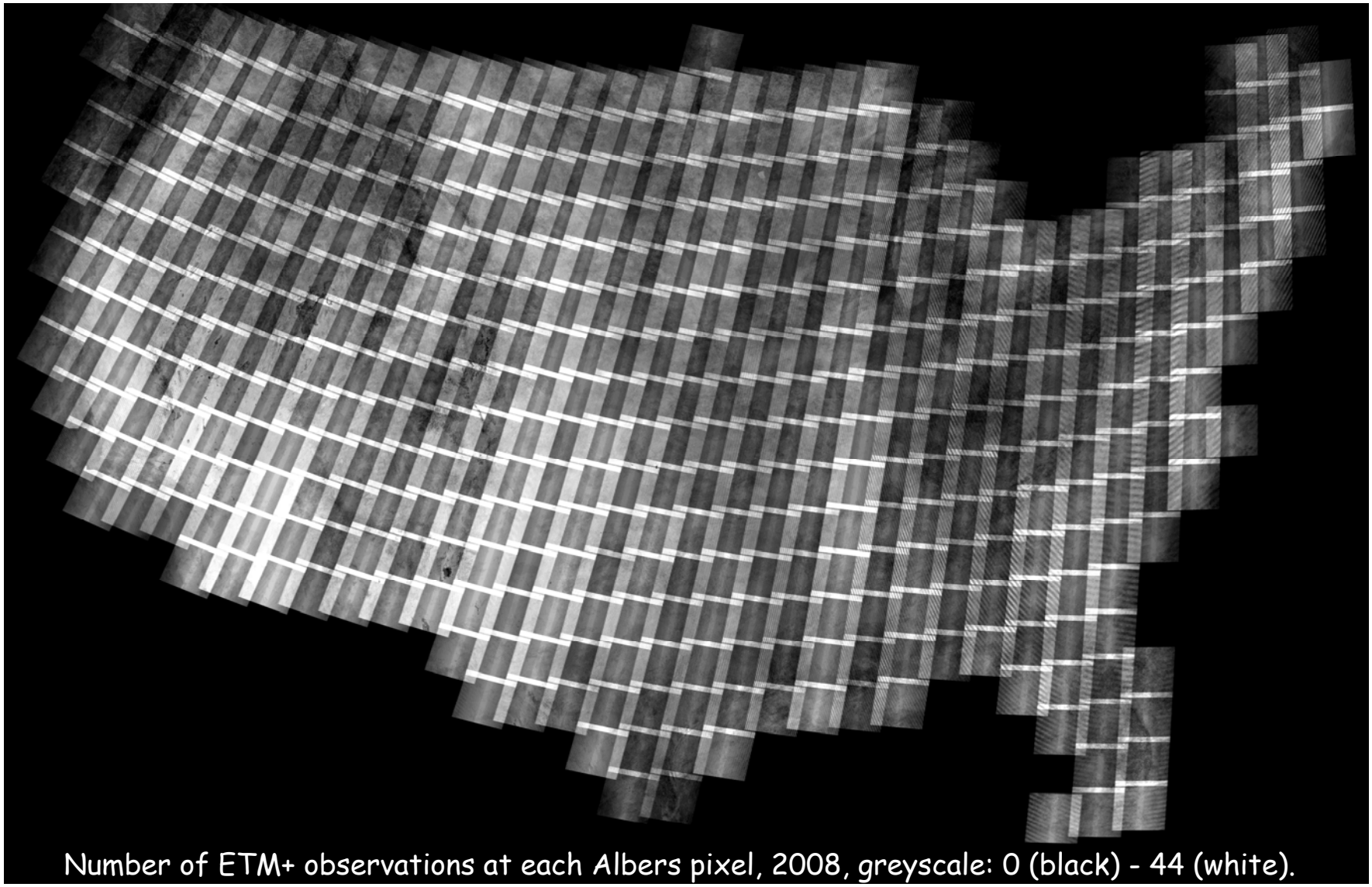
Apply the triangulation topology to the corresponding UTM points

Generate a set of corresponding reference and image triangles

Independently map regular grid of Albers pixel locations from corresponding reference to image triangles using triangle-to-triangle transformation

Nearest neighbor resample (select closest UTM ETM+ pixel)

Compositing: pixel level compositing - at each Albers pixel select the "best" observation in the compsiting period (month, season, year)



Number of ETM+ observations at each Albers pixel, 2008, greyscale: 0 (black) - 44 (white).

Landsat Compositing

- Select the 'best' representative day in the composite period (monthly, seasonal, annual)
  - minimal cloud and atmospheric contamination
  - preferably not shadow contaminated
  - select non-fill observations (until MODIS BRDF gap filling implemented)
  - store the associated processing path, cloud, QA, etc.

- Need to
  - handle band saturation
  - recognize errors of omission and commission in cloud mask
  - composite one image at a time to update composited mosaics in near-real time (i.e. same compositing result regardless of order of inputs)

- Less concerned with compositing over water at this stage

- BRDF and radiometric consistency, and gaps, will be corrected prior to compositing (not currently implemented in the processing chain)

- Reprojection & Compositing undertaken as one process

Compositing Approach
Complex but in essence:

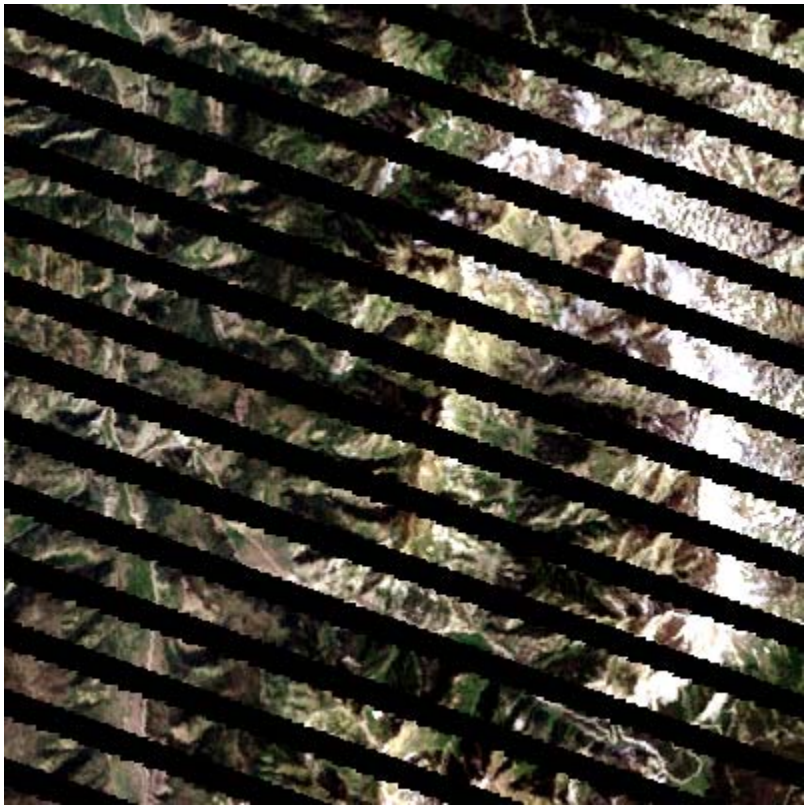
- Use Maximum NDVI criteria
- *But* if red or near-infrared reflectance saturated use Maximum $BT_{6\ low\ gain}$ criteria
- *But* if "dark" observation (NDVI & Band 7 thresholds) and one observation is cloudy use Maximum $BT_{6\ low\ gain}$ criteria

Monthly Compositing Example
Input 1: True Color TOA reflectance

July 5th 2008 Path 38, Row 30

400 x 400 30m pixels Albers

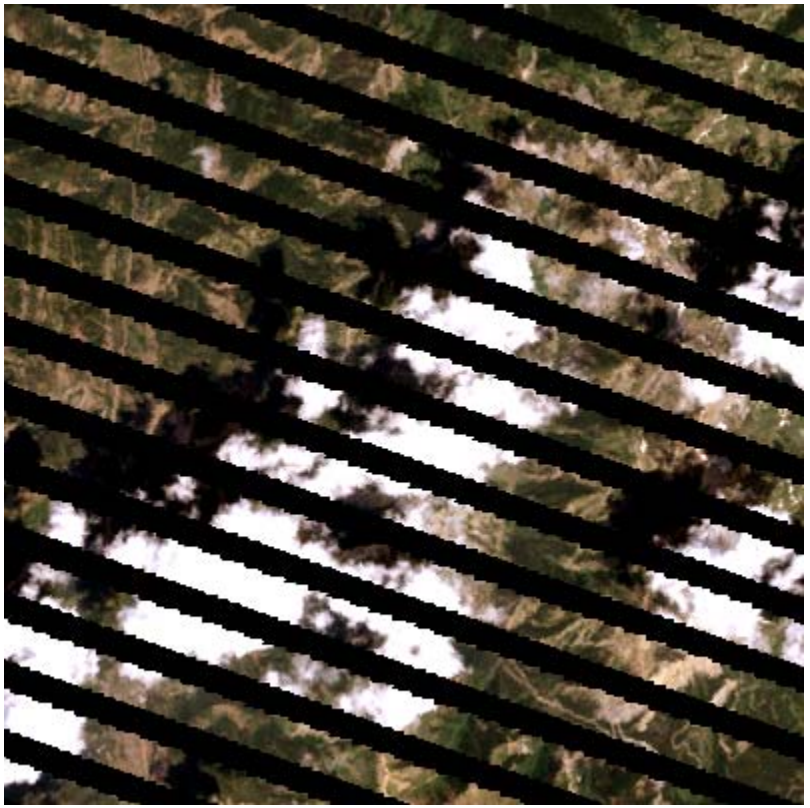
Monthly Compositing Example:
Input 2: True Color TOA reflectance



July 14th 2008 Path 37, Row 30

400 x 400 30m pixels Albers

Monthly Compositing Example:
Input 3: True Color TOA reflectance



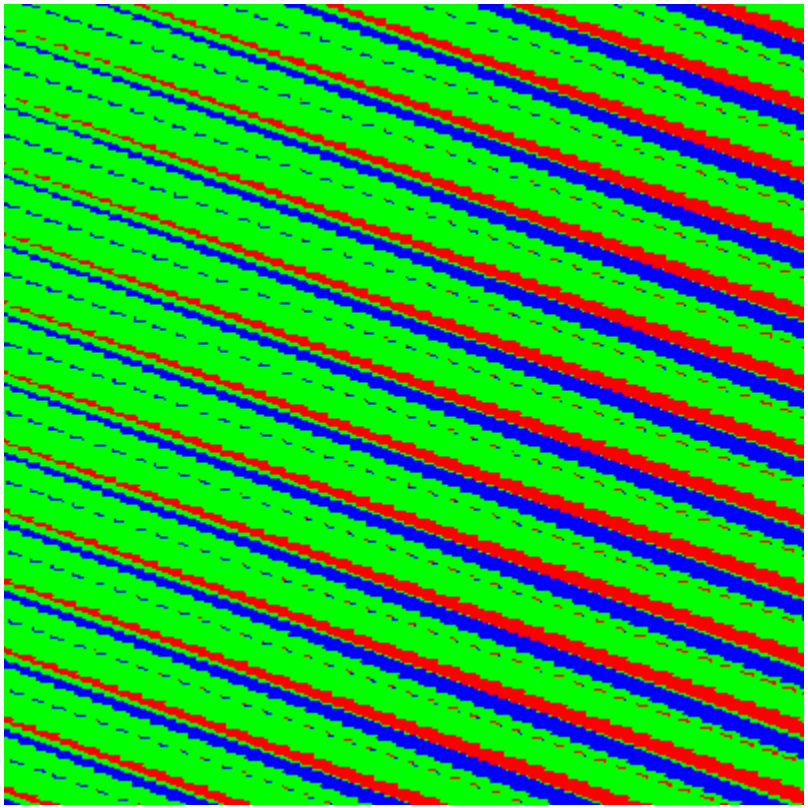
July 30th 2008 Path 37, Row 30

400 x 400 30m pixels Albers

Monthly Composite
Output: True Color TOA reflectance

400 x 400 30m pixels Albers

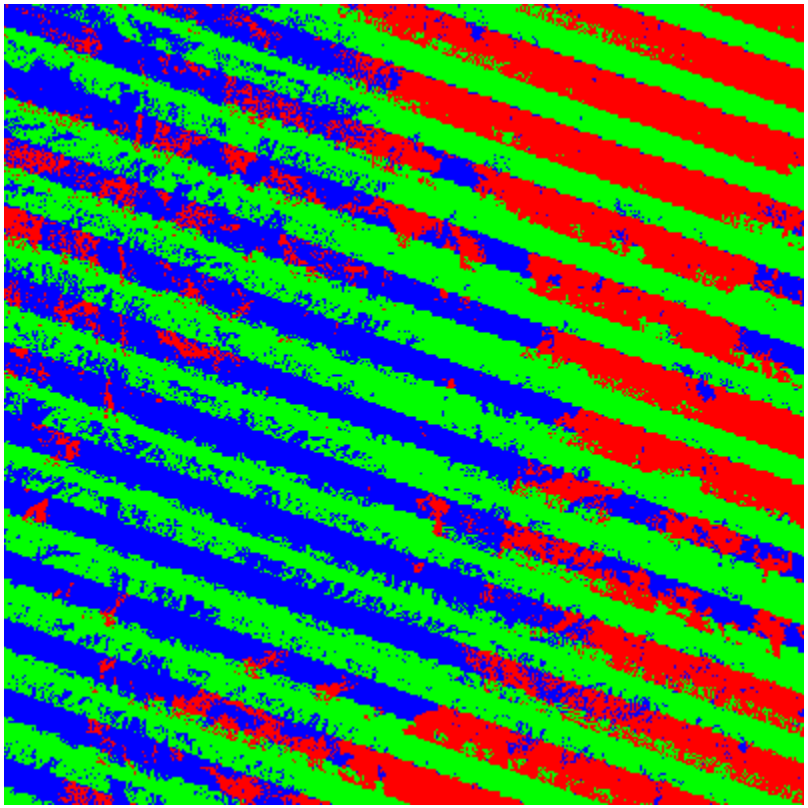
Monthly Composite
Output: Number of observations composited



B = 1, G = 2, R = 3

400 x 400 30m pixels Albers

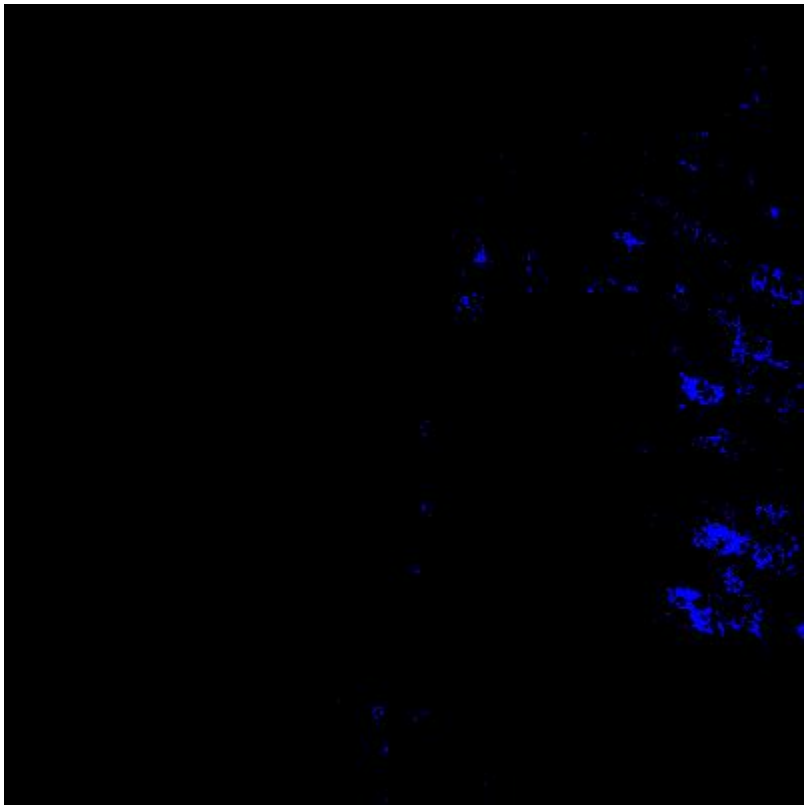
Monthly Composite
Output: Day of year composited



B = July 5th, G = July 14th, Red = July 30th

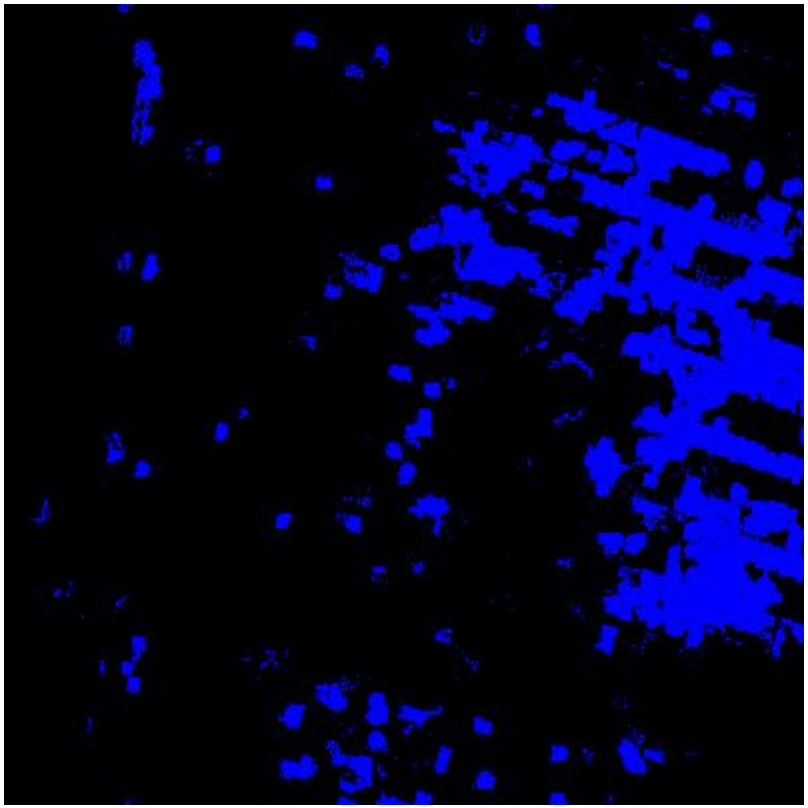
400 x 400 30m pixels Albers

Monthly Composite
Output: Composited pixels flagged as cloudy by classification tree



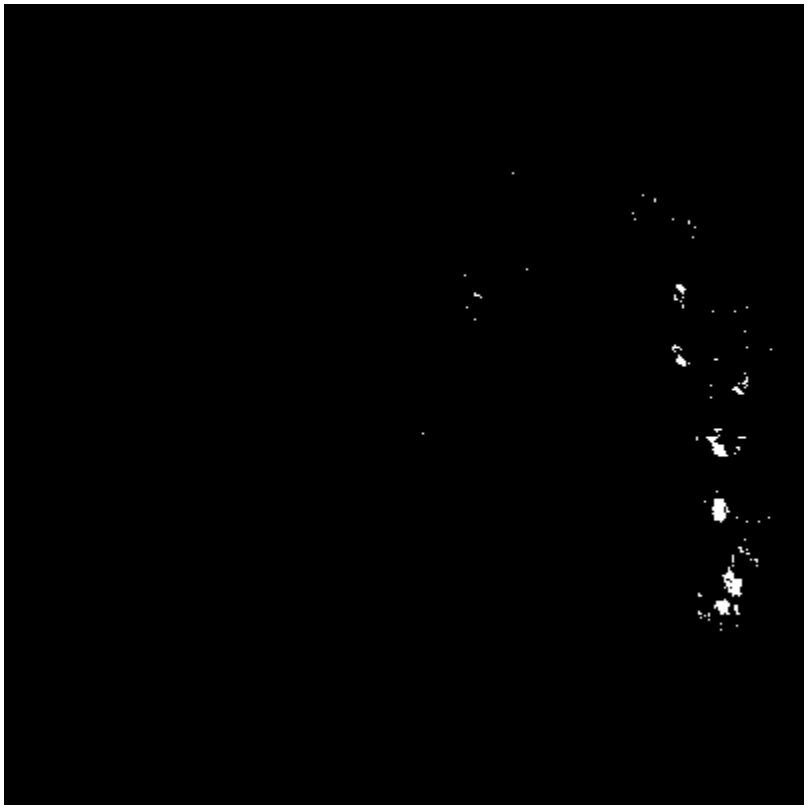
400 x 400 30m pixels Albers

Monthly Composite
Output: Composited pixels flagged as cloudy by ACCA



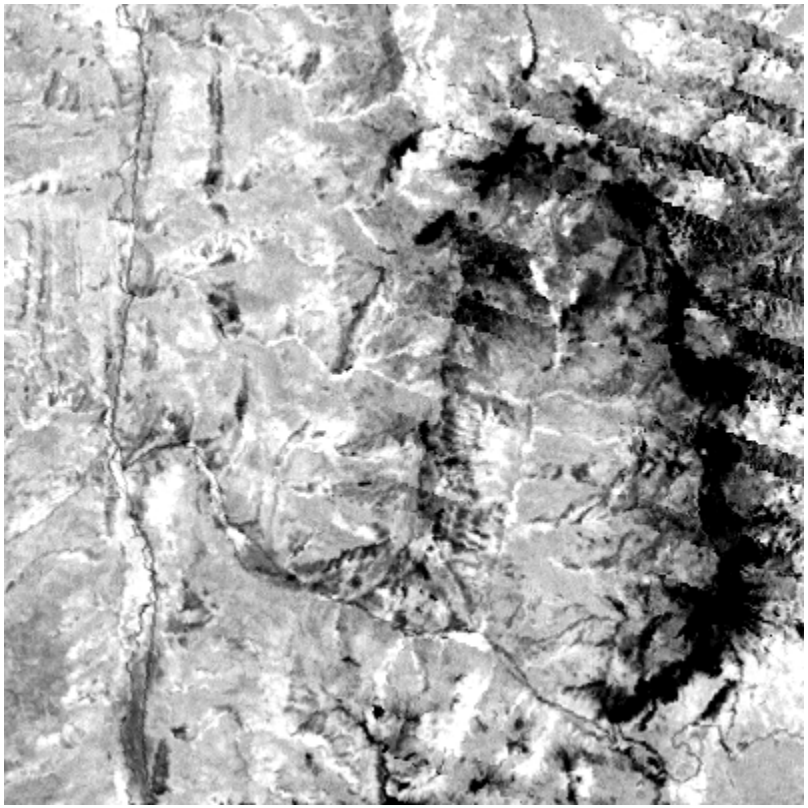
400 x 400 30m pixels Albers

Monthly Composite
Output: Saturation (any band)



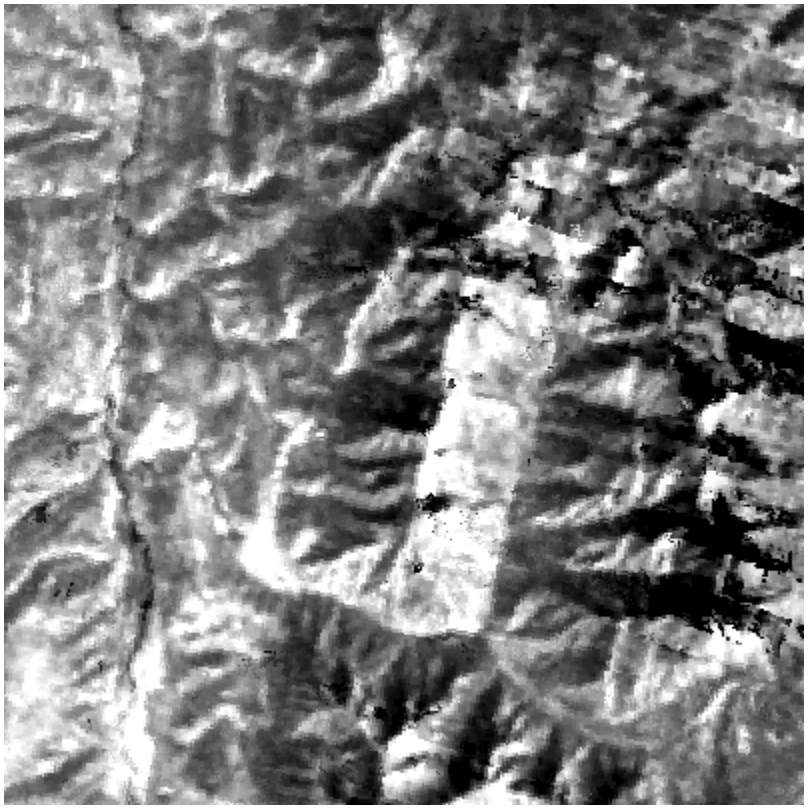
400 x 400 30m pixels Albers

Monthly Composite
Output: NDVI



400 x 400 30m pixels Albers

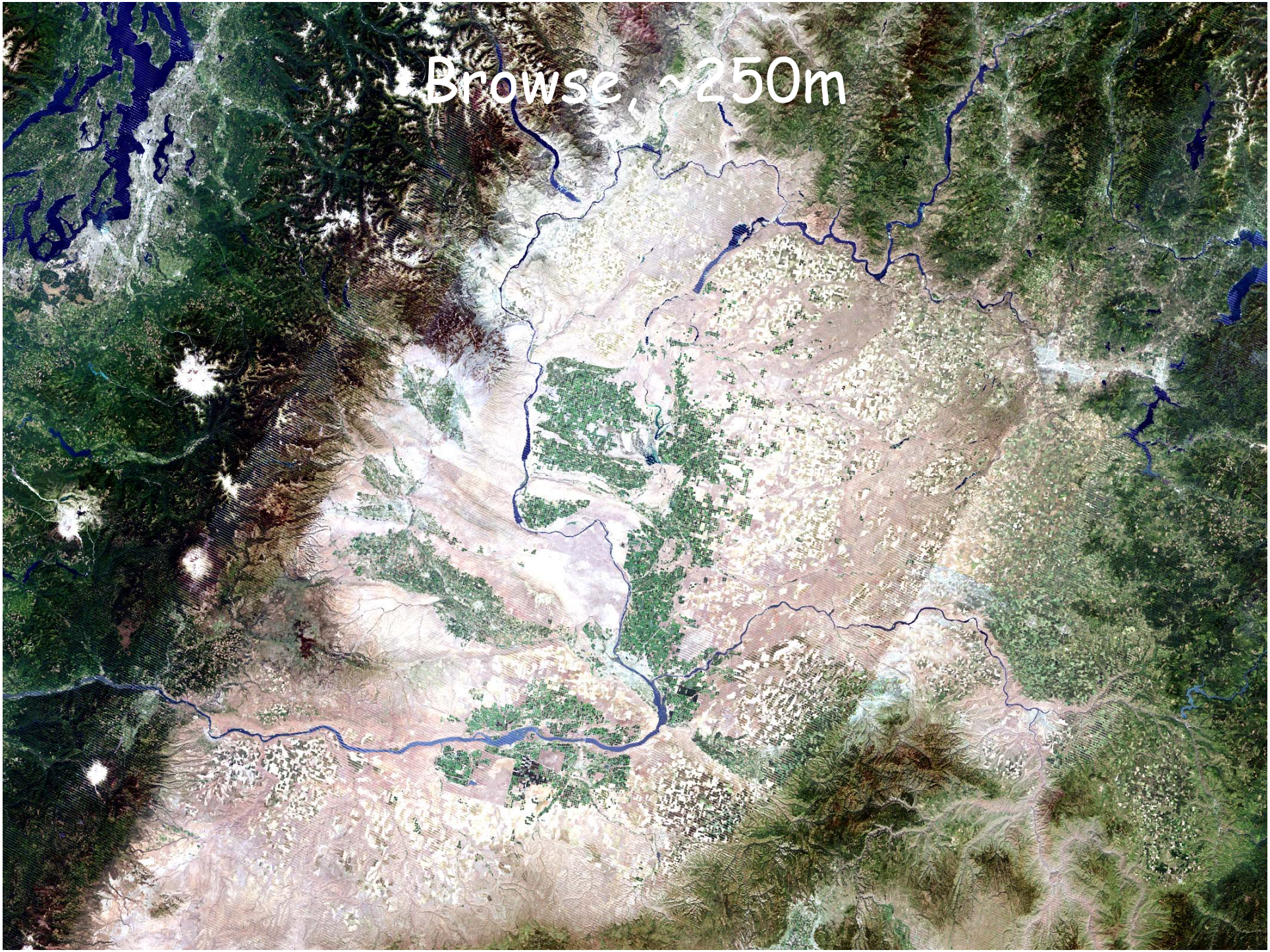
Monthly Composite
Output: Brightness Temperature

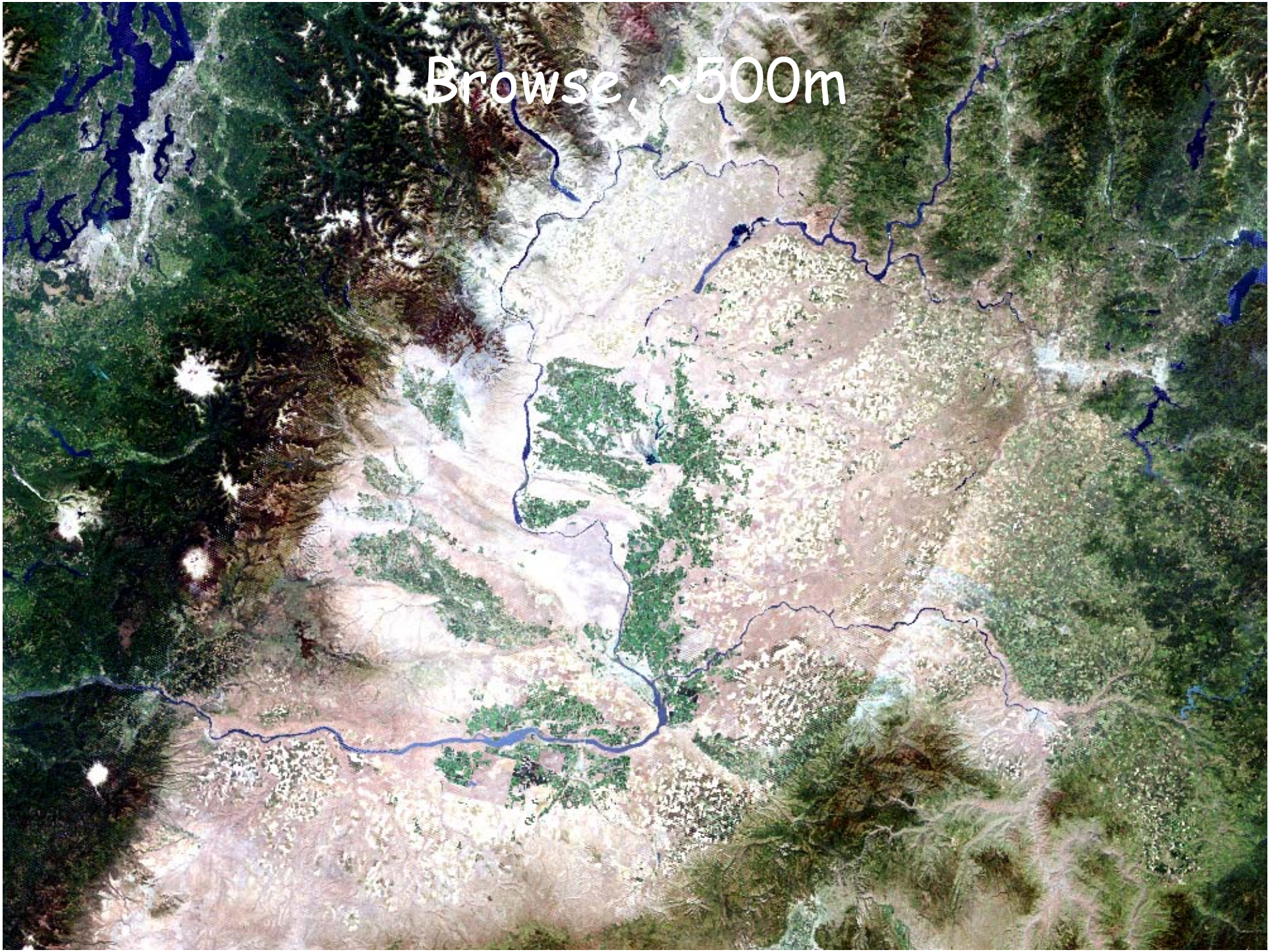


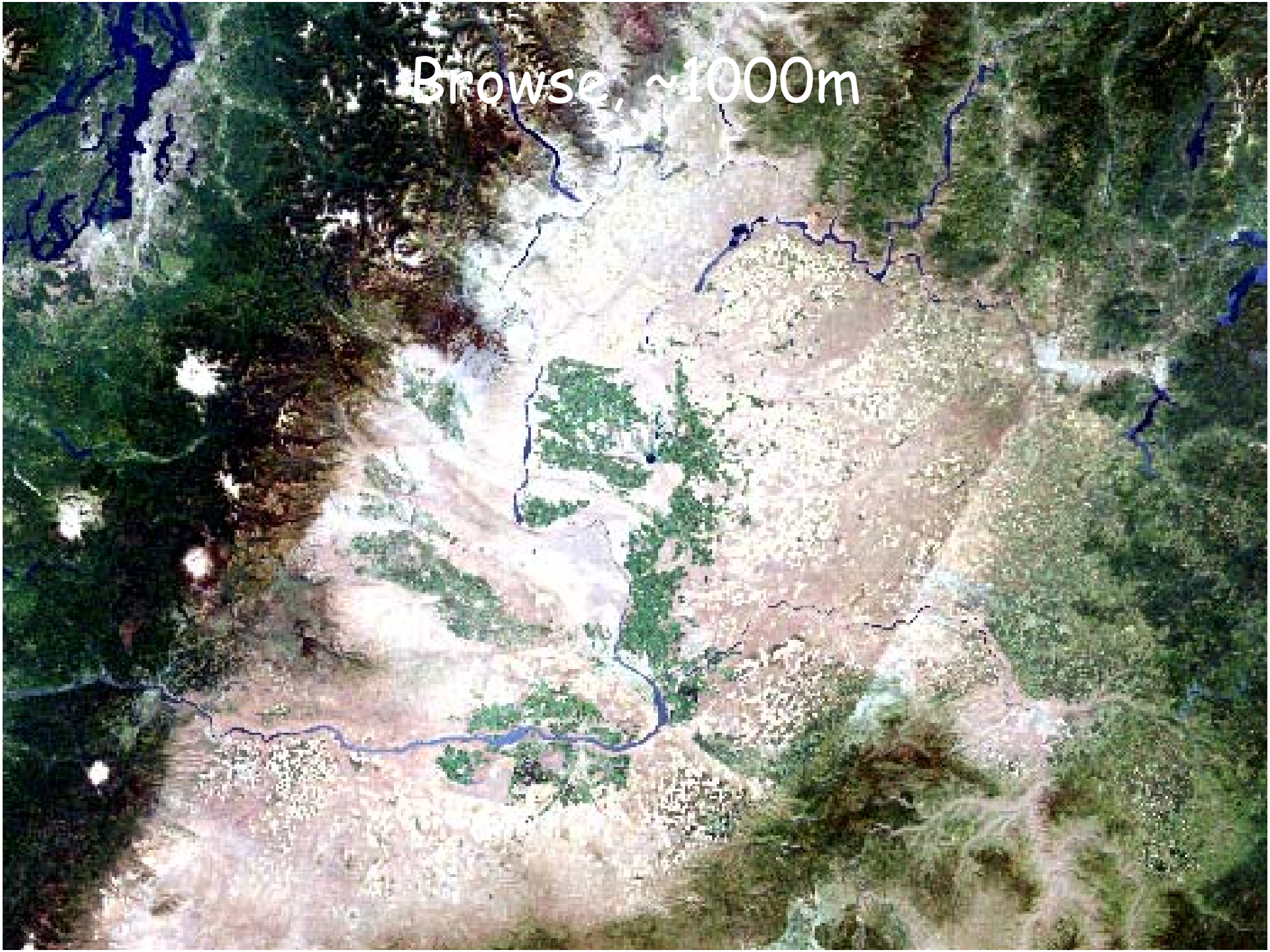
400 x 400 30m pixels Albers

Browse Generation

- To enable synoptic
  - quality assessment
  - browse JPG imagery for internet ordering
- Monthly, Seasonal and Annual
  - CONUS
  - Alaska
- Generated at variety of spatial resolutions
  - median based non-fill information preserving generalization technique
  - all CONUS browse in one file - 2GB file size limitation
- Multi-resolution browse JPEGs with fixed contrast stretching to enable consistent temporal comparison

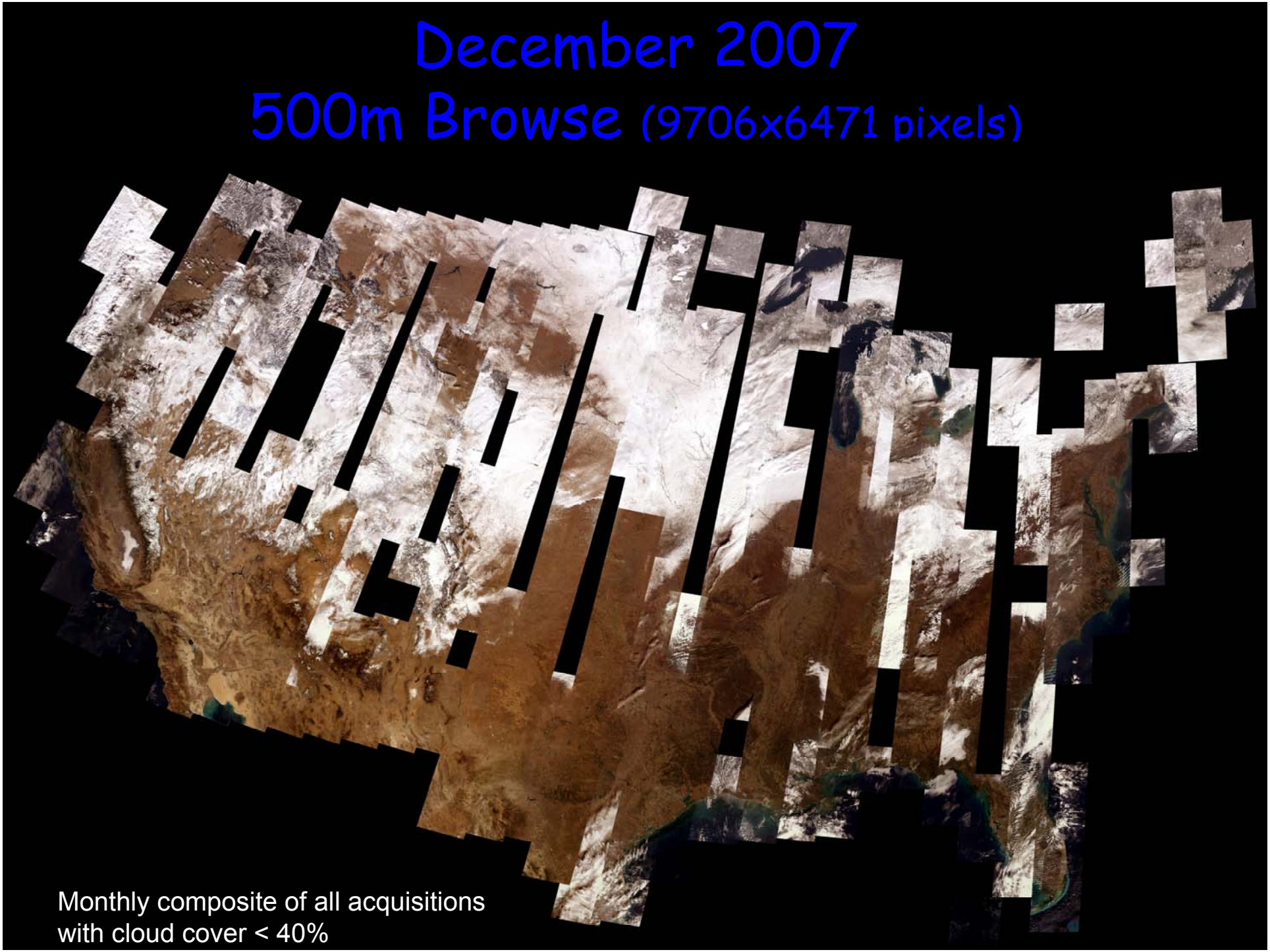

Browse, ~250m


Browse, ~500m


Browse, ~1000m

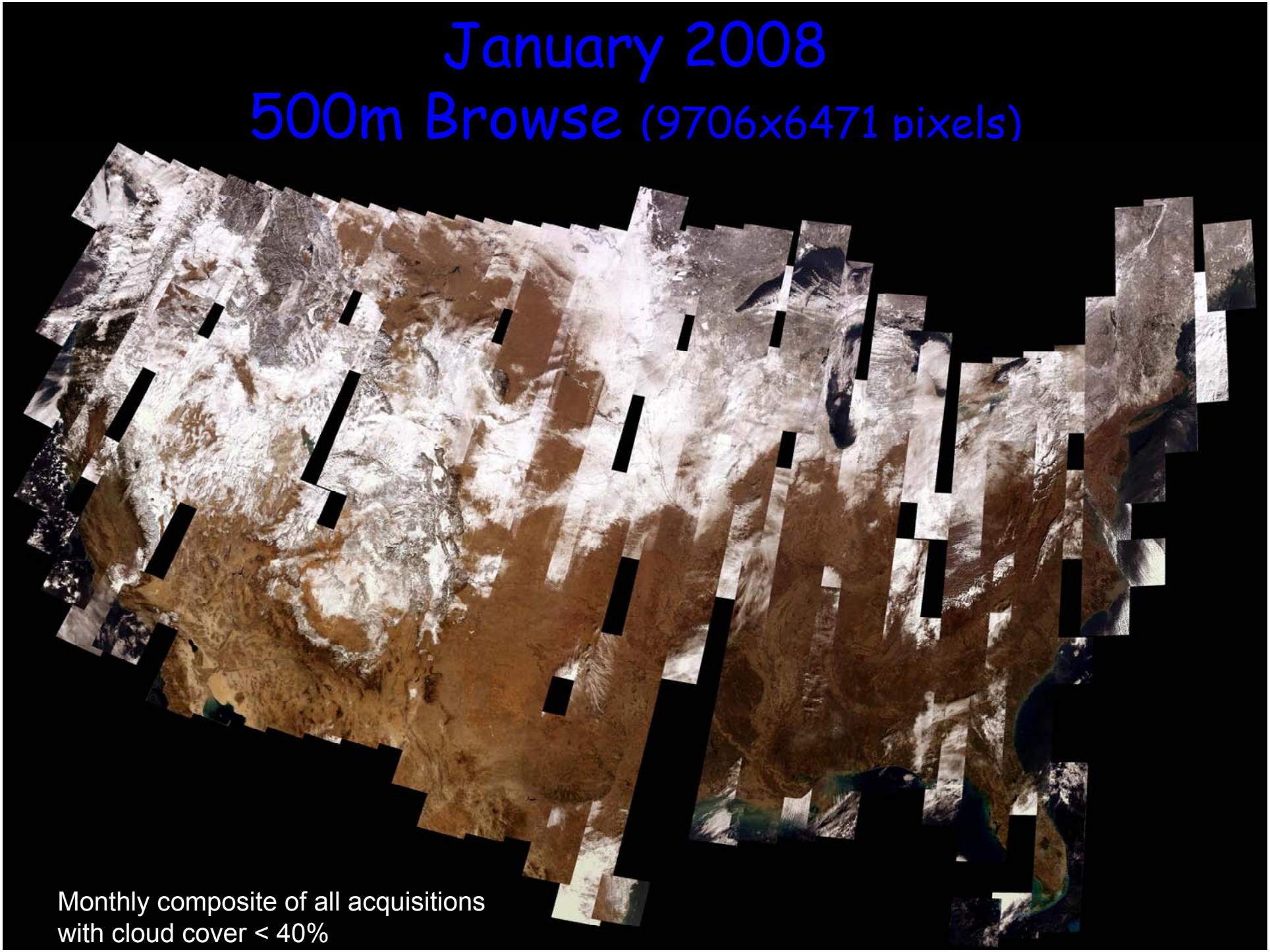
2008 Composited Mosaic Product Results

December 2007
500m Browse (9706x6471 pixels)



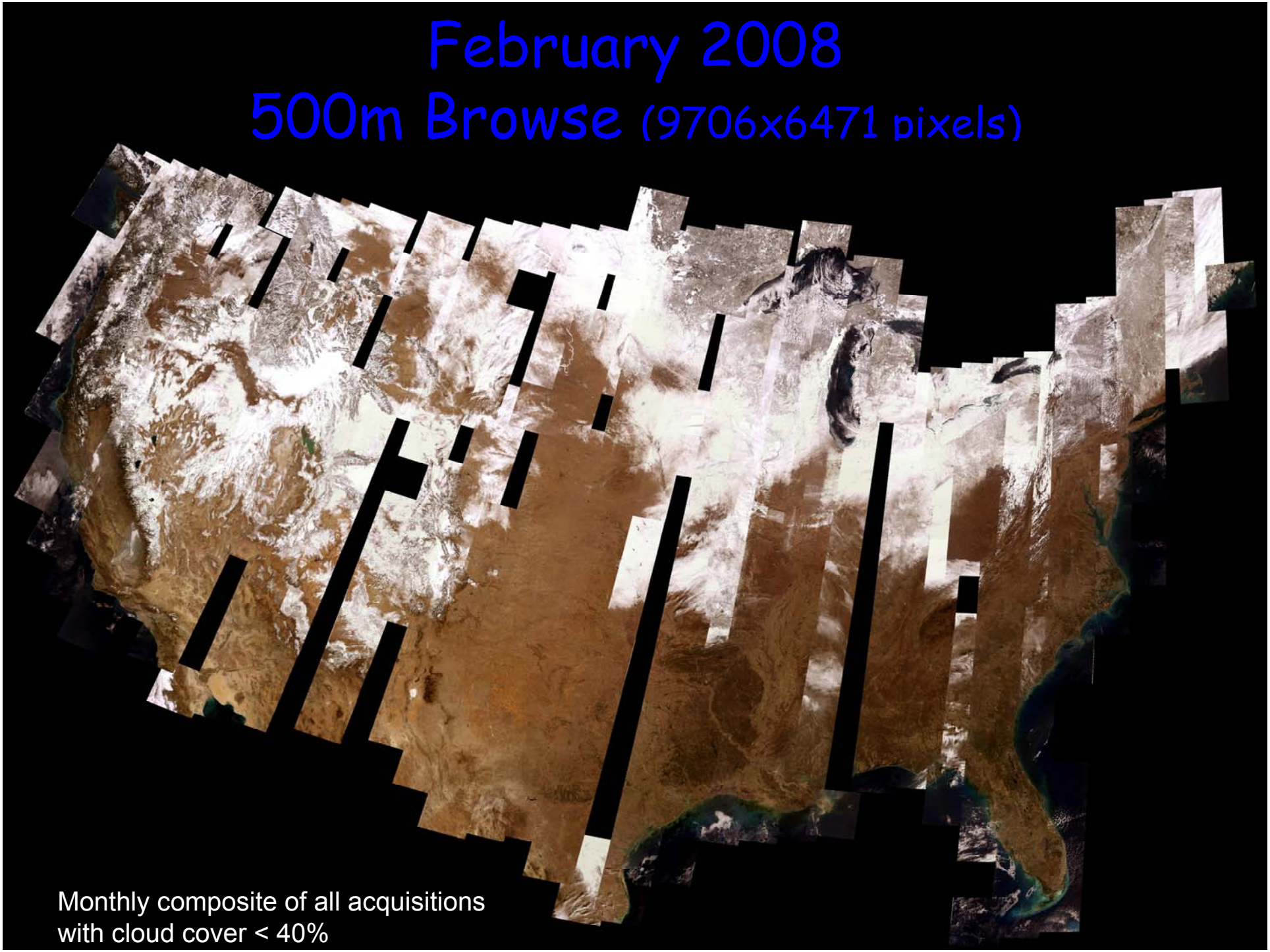
Monthly composite of all acquisitions
with cloud cover < 40%

January 2008
500m Browse (9706x6471 pixels)



Monthly composite of all acquisitions
with cloud cover < 40%

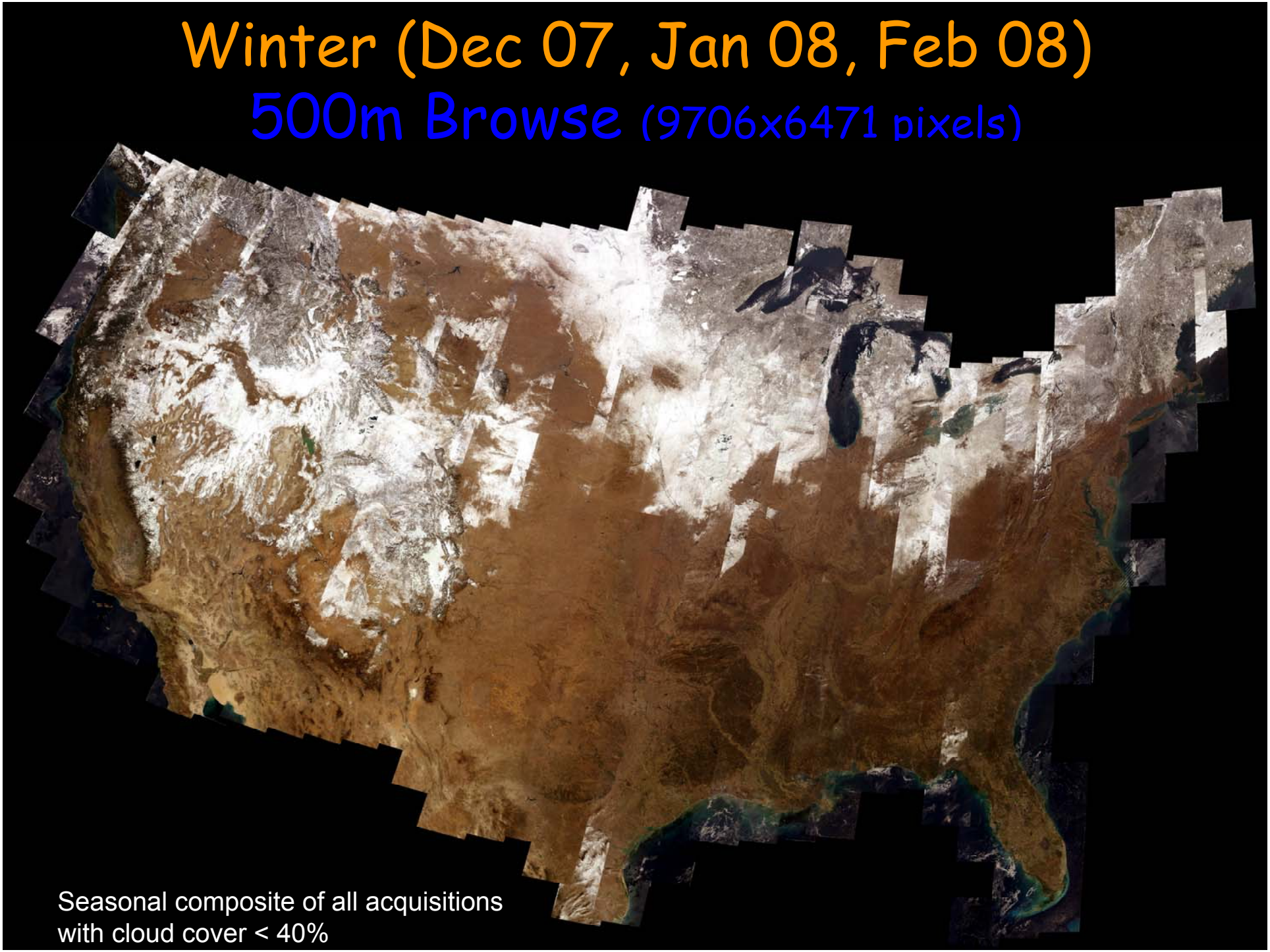
February 2008
500m Browse (9706x6471 pixels)



Monthly composite of all acquisitions
with cloud cover < 40%

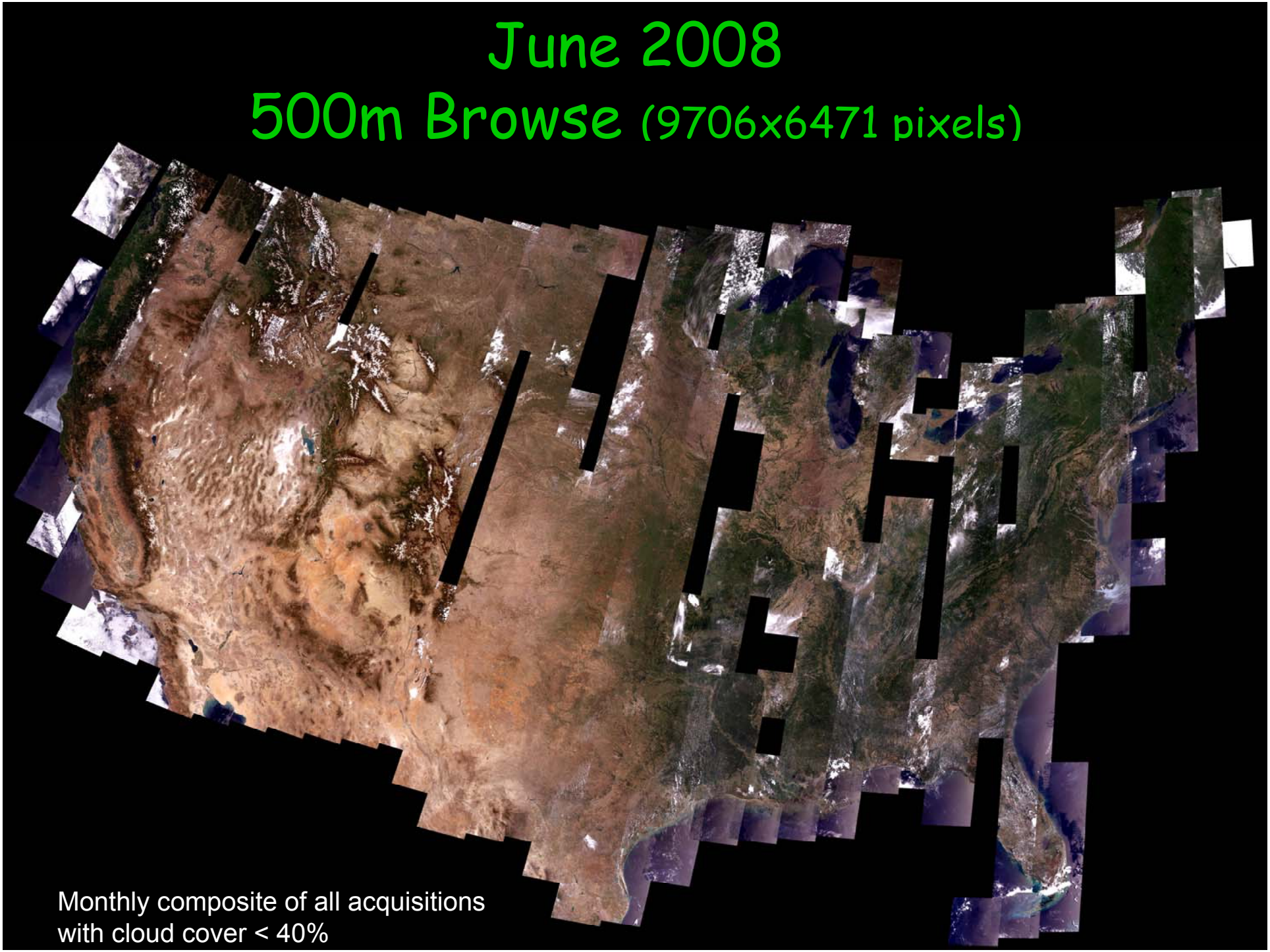
Winter (Dec 07, Jan 08, Feb 08)

500m Browse (9706x6471 pixels)



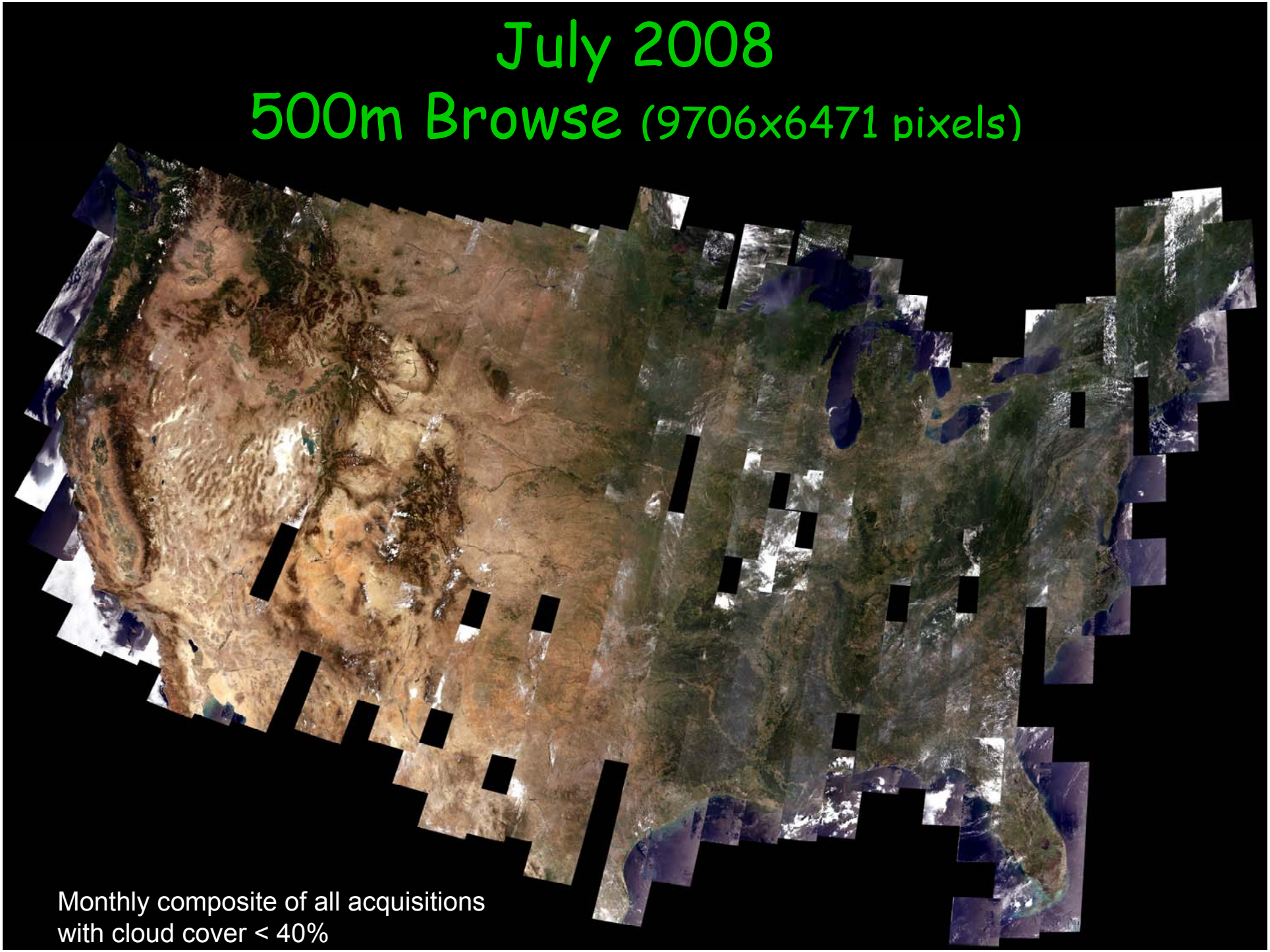
Seasonal composite of all acquisitions
with cloud cover < 40%

June 2008
500m Browse (9706x6471 pixels)



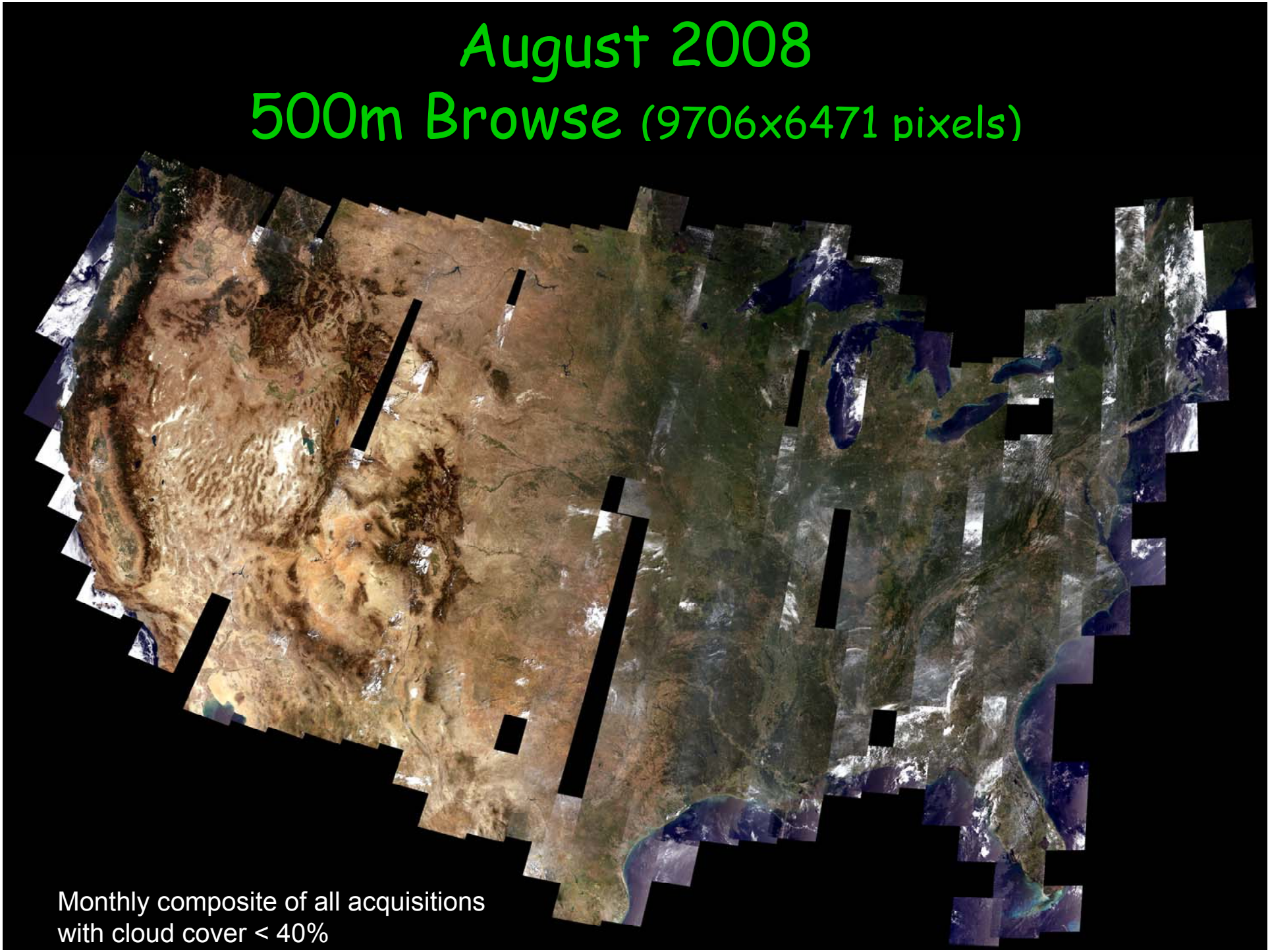
Monthly composite of all acquisitions
with cloud cover < 40%

July 2008
500m Browse (9706x6471 pixels)



Monthly composite of all acquisitions
with cloud cover < 40%

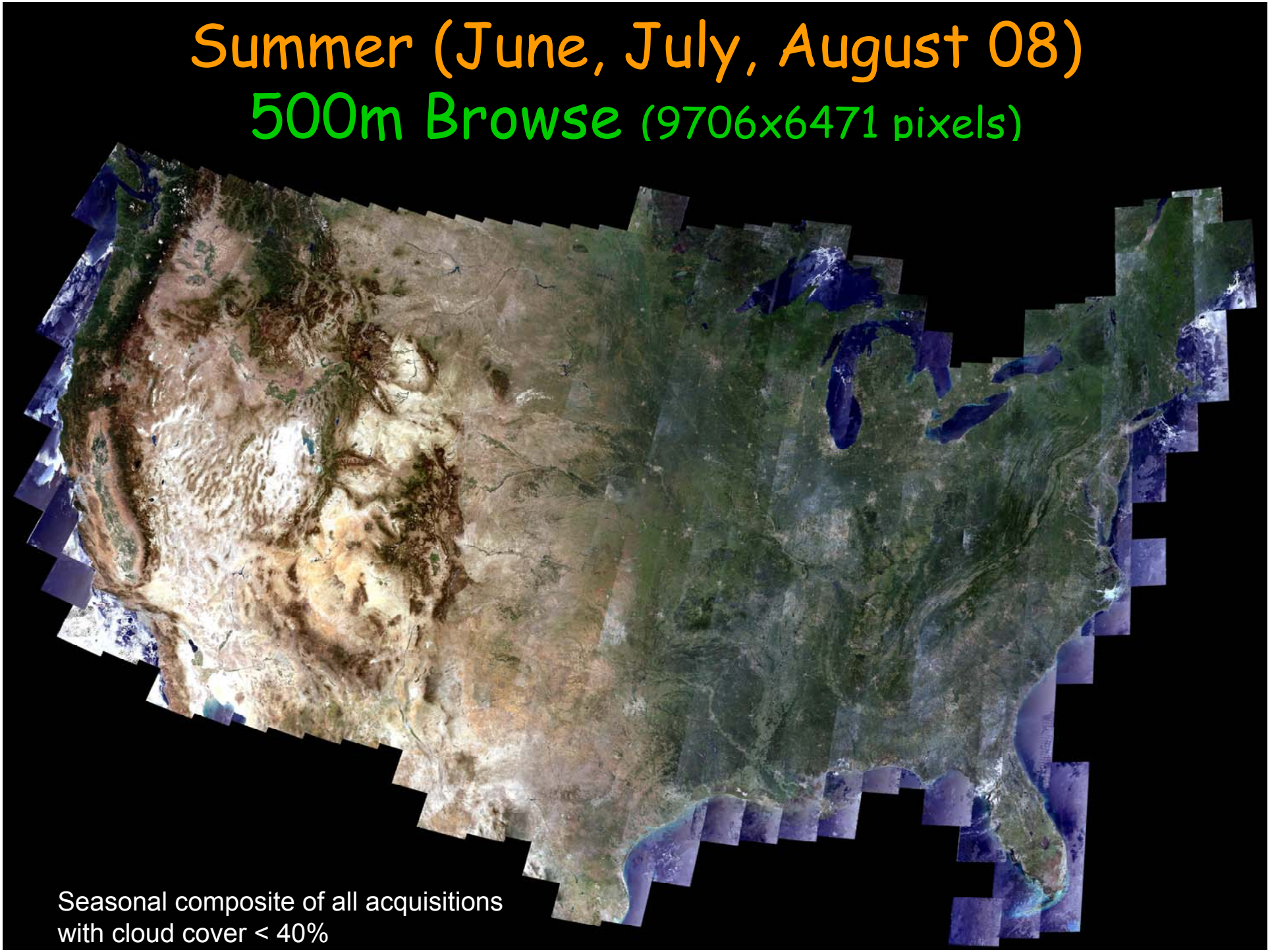
August 2008
500m Browse (9706x6471 pixels)



Monthly composite of all acquisitions
with cloud cover < 40%

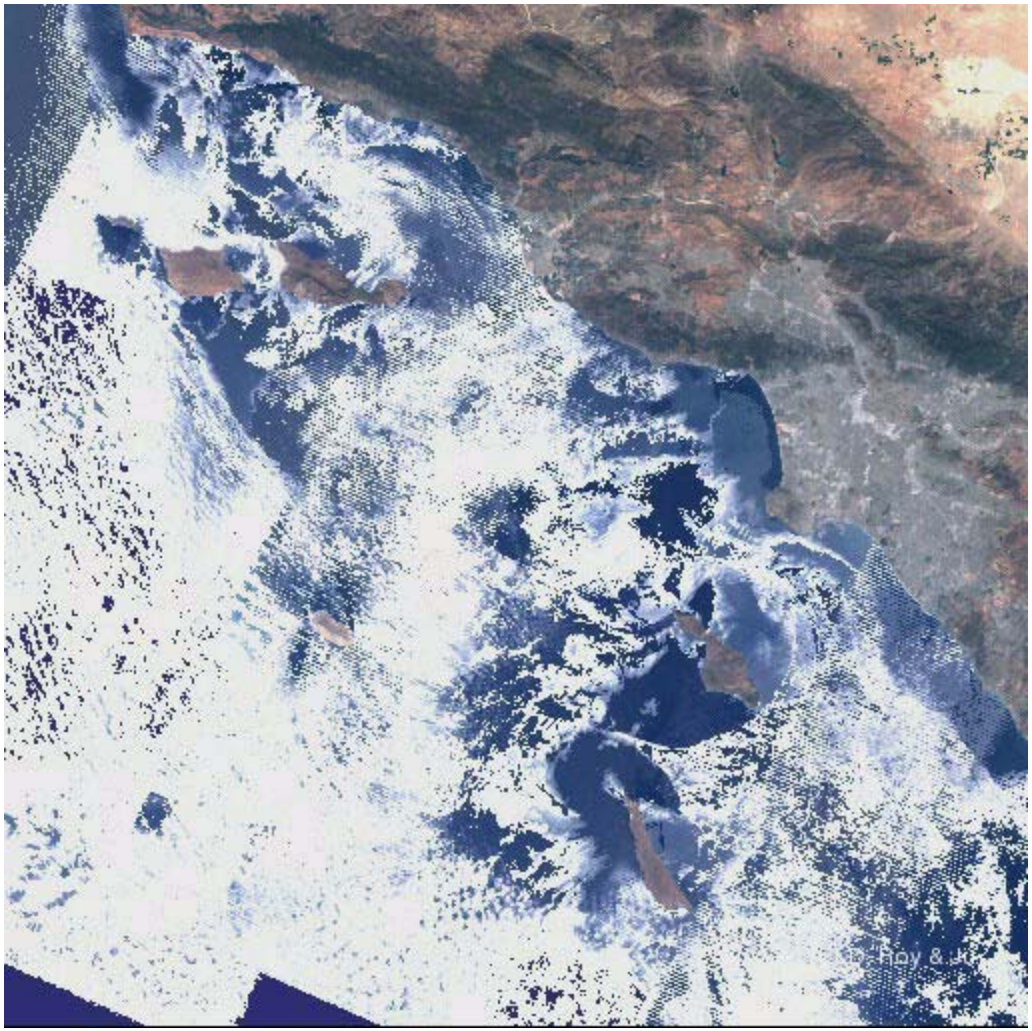
Summer (June, July, August 08)

500m Browse (9706x6471 pixels)



Seasonal composite of all acquisitions
with cloud cover < 40%

Movie: Fly from Los Angeles to New England, Summer 2008, 500m Landsat browse



Movie: Fly from Seattle to Houston, Summer 2008, 500m Landsat browse

Ongoing Work

- Gap Filling & Radiometric/BRDF normalization – using MODIS BRDF (we presented on this previously)
- Atmospheric Correction – using MODIS (Eric Vermote)
- Monthly and Annual Landcover characterization (% Tree, %Herbaceous, %Bare Ground, %Water, %Snow/Ice) – using MODIS (Matt Hansen)
- Advocacy to increase number of USGS web enabled acquisitions for CONUS and Alaska – *either* ~80% cloud cover threshold (instead or 40%) *or* use path/row/seasonal specific cloud cover thresholds (*have done the proof of concept research)*. We need temporally rich Landsat time series to have improved composited mosaic products (with an 80% cloudy scene, the non-cloudy 20% can still be composited)
- WELD project web Site
- Make “black step processed” monthly/seasonal/annual mosaic products available
- User outreach – community endorsement
- NASA reporting - ATBD