STAR Algorithm and Data Products (ADP) Review

Suomi NPP Surface Type EDR Products

C. Huang, K. Song, M. Friedl, D. Sulla-Menashe
Y. Tang, M. Tsidulko, R. Mahoney, A. Sei

Presented by X. Zhan

1/7/2014

Presentation Outline

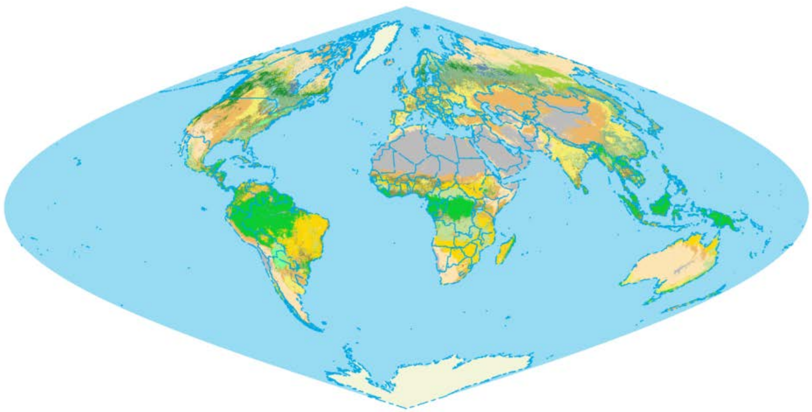
- Brief overview of Surface Type EDR algorithm and products
- Summary of beta maturity review results (2/20/2013)
- Status of VIIRS Quarterly Surface Type Intermediate Product (QST IP) development
- Summary and future plan

VIIRS Surface Type EDR Team

- Xiwu Zhan (NOAA/NESDIS/STAR)
  - Surface Type EDR team lead, User outreach
- Chengquan Huang (UMD/Geography)
  - Algorithm development lead
- Kuan Song (UMD/Geography)
  - Algorithm development
  - QST Product generation
  - User readiness
- Mark Friedl (Boston University)
  - Cal/Val lead
- Damien Sulla-Menashe (Boston University)
  - Ground truth data development
  - Product validation
  - MODIS land cover products as QST IP seed

User Needs for Surface Type Products

- Essential Climate Variable
  - Modeling studies (*Primary user: NCEP land team led by M. Ek*)
    - Land surface parameterization for GCMs
    - Biogeochemical cycles
    - Hydrological processes
  - Carbon and ecosystem studies
    - Carbon stock, fluxes
    - Biodiversity
- Feed to other VIIRS products
  - BRDF/Albedo
  - Land surface temperature (LST)
- Surface type change monitoring
  - Snow cover
  - Burned area (TBD)
  - Flooding/deforestation/etc (TBD)



L1RD Requirements

Table 4.5.4.2 - Surface Type (VIIRS)		
EDR Attribute	**Threshold**	**Objective**
SURF Applicable Conditions: 1. Both clear and partly cloudy sky conditions		
a. Horizontal Cell Size	1 km at Nadir	1 km at Edge of Scan
b. Mapping Uncertainty, 3 Sigma	5 km	1 km
c. Measurement Range	17 IGBP classes specified in Table 4.5.4.1	17 IGBP classes
d. Measurement Precision *	10%	0.1%
e. Measurement Accuracy *	70% correct for 17 types	2%
f. Refresh	At least 90% coverage of the globe every 24 hours (monthly average)	3 hrs.
		v2.0, 9/23/12

* *Current IDP product was designed to meet heritage NPOESS requirements. Beta evaluation is done against those heritage requirements. Precision and accuracy numbers are to be corrected in the JPSS L1RD.*

Existing Global Surface Type Products

Product	Sensor	Resolution	Data range	Classification method	Accuracy	Source
GlobCover	MERIS	300 m	2005	Classification and labeling	73%	Defourny et al., 2007
MODIS LC-Collection 5	MODIS	500 m	2005	Decision tree	74.8%	Friedl et al., 2010
MODIS LC-Collection 4	MODIS	1 km	2001	Decision tree	~75%	Friedl et al, 2002
DISCover	AVHRR	1 km	1992.4-1993.3	Clustering and labeling	66.9%	Scepan, 1999
UMD-1km	AVHRR	1 km	1992.4-1993.3	Decision tree	69% (training)	Hansen et al., 2000
GLC2000	SPOT-VEGETATION	1 km	1999.11-2000.12	Regional experts	68.6%	Mayaux et al., 2006

Overview of VIIRS Surface Type EDR

- Describes surface condition at time of each VIIRS overpass
- Produced for every VIIRS swath/granule
  - Same geometry as any VIIRS 750m granule
- Two major components
  - Gridded Quarterly Surface Type (QST) IP
    - Remapped to the swath/granule space for each VIIRS acquisition
    - Requires one full year of VIIRS data
    - Just developed, subject of this review
    - Currently use MODIS land cover (LC) as QST IP seed in the production system
  - Includes flags to indicate snow and fire based on
    - Active fire Application Related Product (ARP)
    - Snow EDR

History of ST EDR Related DRs

Date	DR #	Reason	Status
11/18/2011	4452	ST rename vegetation fraction	NDE provides separate GVF product in future
11/18/2011	4458	C5 Decision tree replacement	UMD to upgrade QST IP algorithm with SVM in future
11/18/2011	4459	QST IP goes to annual	To be approved in future
05/22/2012	4707	Update QST IP seed data	Completed with MODIS C5 LC
09/12/2012	4900	Remove new QST IP Seed fill values	Completed with update MODIS C5 LC
01/14/2013	7002	ST EDR Beta review	Completed, with a focus on ST EDR
03/2013	7090	ST EDR update for use of snow rolling tile	Approved, waiting for final DPE test
09/2013	7329	ST EDR controller cores without optional input VCM	Changed VCM to required input

Current Status of Surface Type EDR

- Beta review completed:
  - QST IP (gridded) copied and mapped to the EDR (swath) properly
  - Fire flag set properly according to the Fire ARP product
  - Snow flag set properly according to the Snow EDR
- Improvements since beta review:
  - Change snow fraction from 0.75 to 0.5 in aggregating Snow EDR from 375m to 750m before being copied to Surface Type EDR
  - Use rolling tile snow products when VIIRS Snow EDR not available
- QST IP is still based on MODIS Collection5 LC.

What to look for in this review

- Development of VIIRS QST IP
  - VIIRS data processing
    - Gridding
    - Compositing
    - Annual metrics generation
  - Training data collection
  - Classification
- QST IP assessment
  - Comparison with existing products
- Next steps

VIIRS QST IP Overview

- Global surface type product
  - Gridded, 1km
  - 17 IGBP surface type classes
  - Required typing accuracy ~70%
  - Generated quarterly to reflect recent changes
  - Based on gridded surface reflectance products (IDPS version still being evaluated)
  - Use decision tree (C5.0) classifier
  - Requires training data

Surface Types in QST IP

Classes separated by a threshold value more likely to be confused:

Closed vs open shrublands

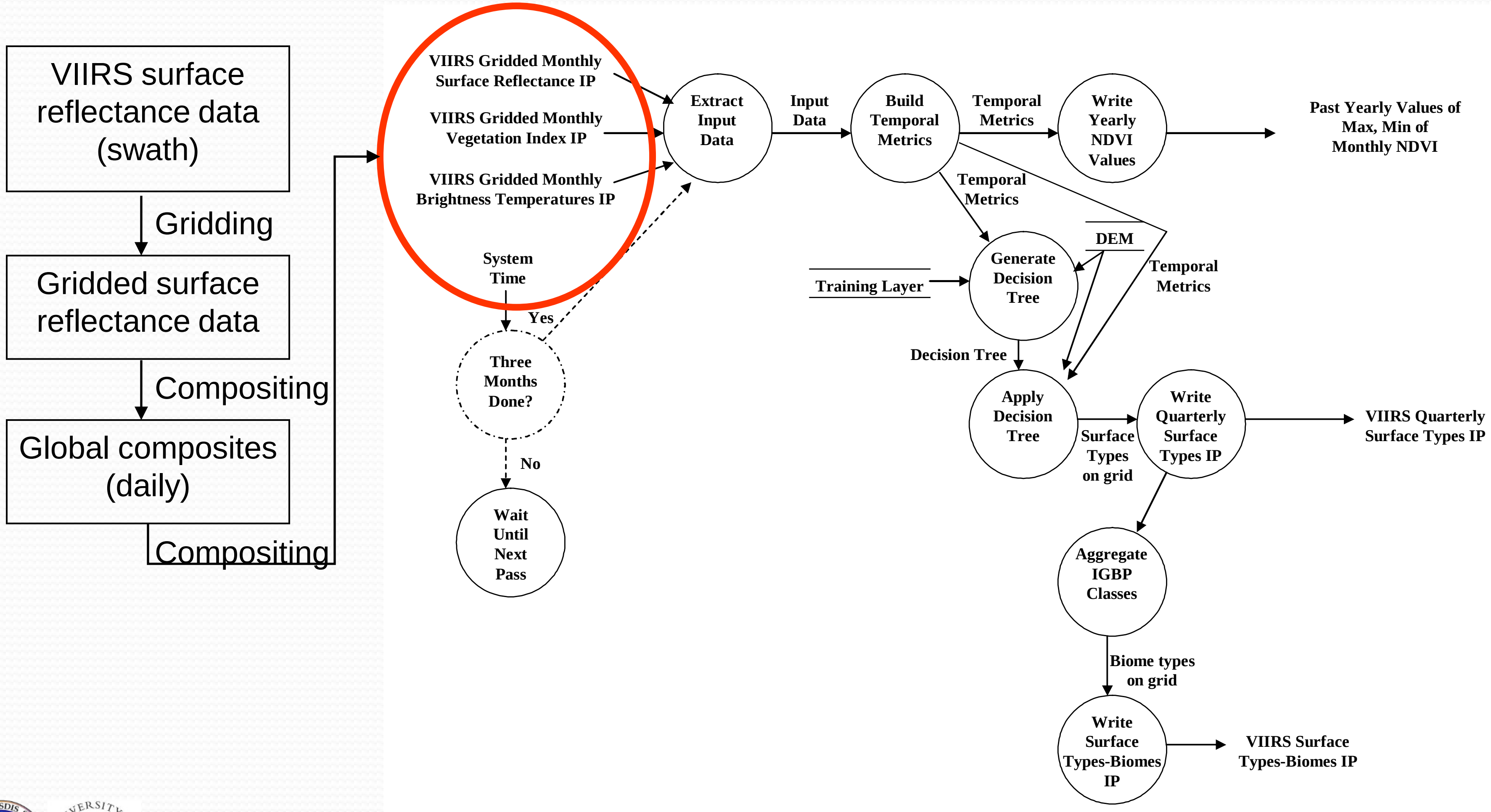
Forest vs woody savanna vs savanna

Cropland vs cropland/natural vegetation mosaics

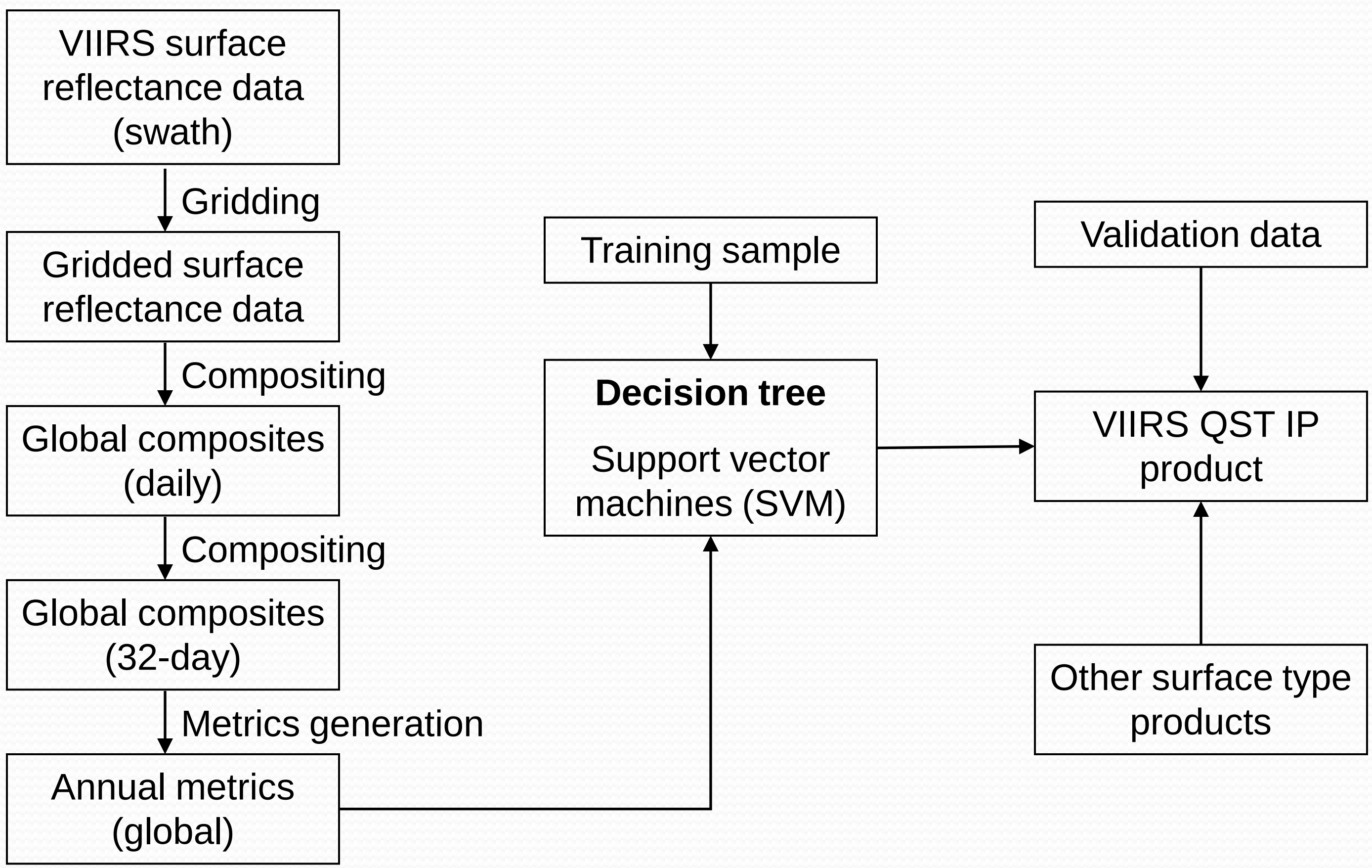
1	Evergreen Needleleaf Forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Almost all trees remain green all year. Canopy is never without green foliage.
2	Evergreen Broadleaf Forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Almost all trees remain green all year. Canopy is never without green foliage.
3	Deciduous Needleleaf Forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Consists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
4	Deciduous Broadleaf Forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Consists of seasonal broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
5	Mixed Forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Consists of tree communities with interspersed mixtures or mosaics of the other four forest cover types. None of the forest types exceeds 60% of landscape.
6	Closed Shrublands	Lands with woody vegetation less than 2 meters tall and with shrub canopy cover is >60%. The shrub foliage can be either evergreen or deciduous.
7	Open Shrublands	Lands with woody vegetation less than 2 meters tall and with shrub canopy cover is between 10-60%. The shrub foliage can be either evergreen or deciduous.
8	Woody Savannas	Lands with herbaceous and other understorey systems, and with forest canopy cover between 30-60%. The forest cover height exceeds 2 meters.
9	Savannas	Lands with herbaceous and other understorey systems, and with forest canopy cover between 10-30%. The forest cover height exceeds 2 meters.
10	Grasslands	Lands with herbaceous types of cover. Tree and shrub cover is less than 10%.
11	Permanent Wetlands	Lands with a permanent mixture of water and herbaceous or woody vegetation that cover extensive areas. The vegetation can be present in either salt, brackish, or fresh water.
12	Cropland	Lands covered with temporary crops followed by harvest and a bare soil period (e.g., single and multiple cropping systems. Note that perennial woody crops will be classified as the appropriate forest or shrub land cover type.
13	Urban and Built-up	Land covered by buildings and other man-made structures. Note that this class will not be mapped from the AVHRR imagery but will be developed from the populated places layer that is part of the Digital Chart of the World.
14	Cropland/Natural Vegetation Mosaics	Lands with a mosaic of croplands, forest, shrublands, and grasslands in which no one component comprises more than 60% of the landscape.
15	Snow and Ice	Lands under snow and/or ice cover throughout the year.
16	Barren	Lands exposed soil, sand, rocks, or snow and never has more than 10% vegetated cover during any time of the year.
17	Water Bodies	Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or salt water bodies

BOSTON UNIVERSITY

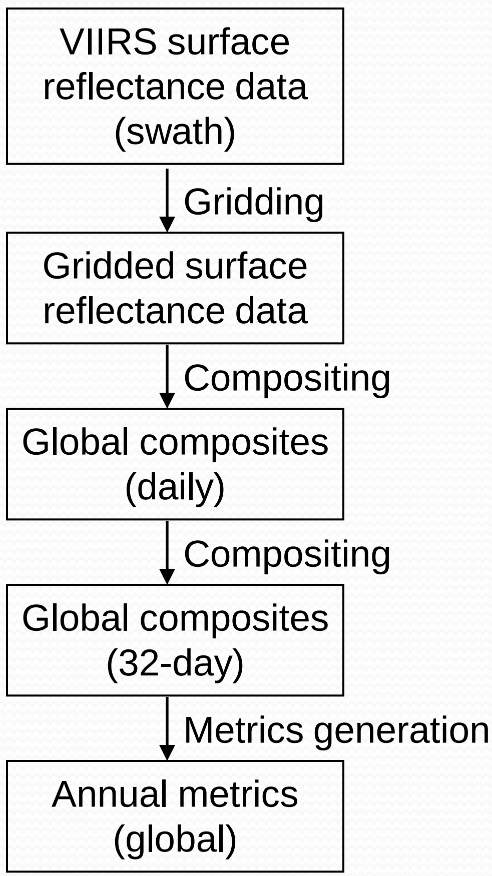
Data Flow of QST IP Algorithm



Overview of QST IP Algorithm and Product Assessment Approaches



Gridding, Compositing, Annual Metrics (GCAM) Overview

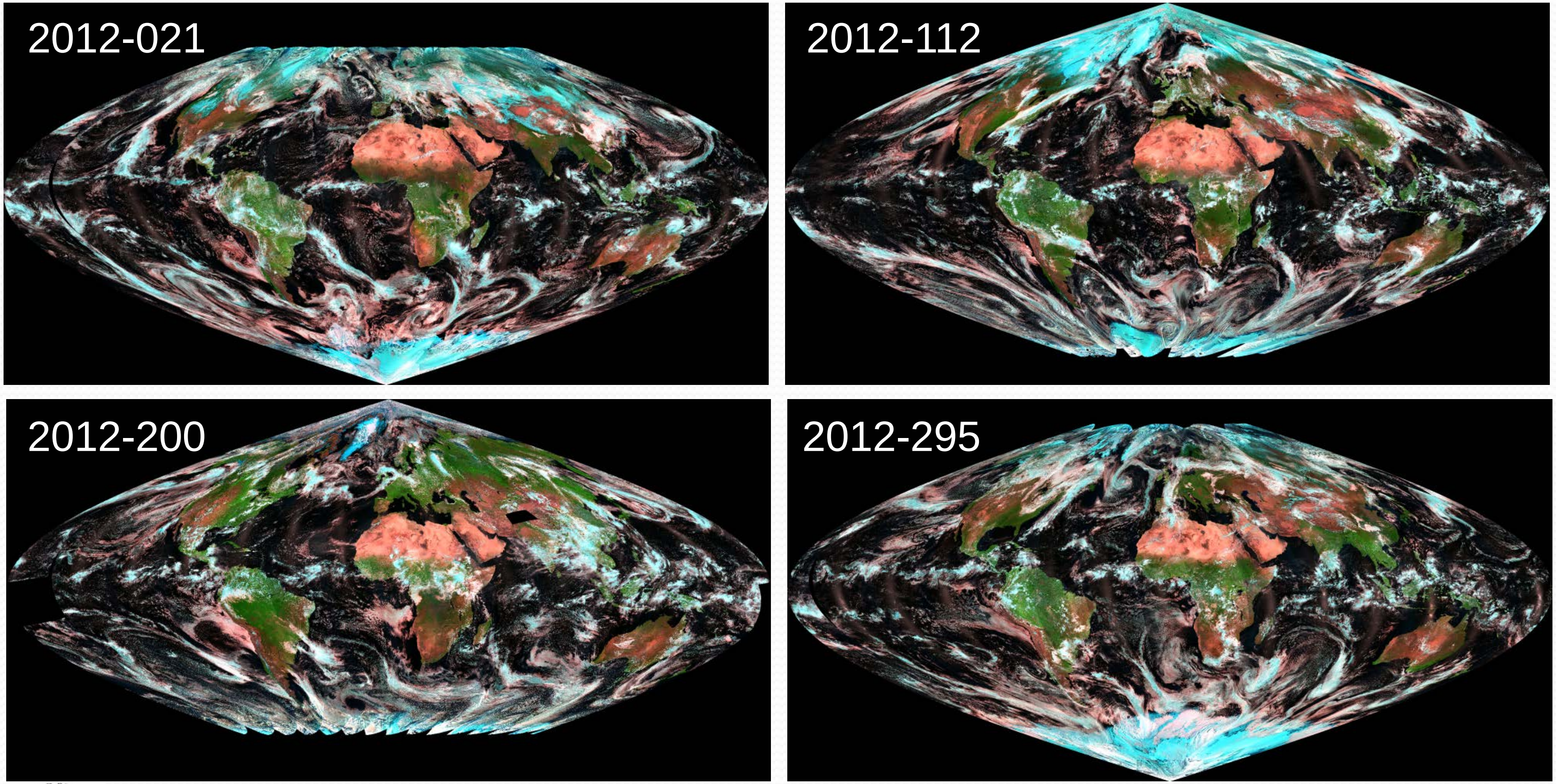


- Required before classification
  - Remove cloud/shadow
  - Annual metrics globally more consistent
- Should be done by IDPS, but gridded data not available until recently
  - processed in-house at UMD
  - Improvements possible
    - Composites comparable to or better than MODIS composites

Gridding

- Done in house at UMD
  - No gridded data from IDPS until late 2012
  - Solution provided by Land PEATE
    - MODIS like VIIRS data (5-minutes aggregation)
    - Gridding tool
  - Fully automated downloading and gridding scripts
  - All 2012 VIIRS data required by QST IP processed
    - ~880,000 files (80,000 granules x 11 bands), totaling ~150 TB
    - > 30,000 CPU hours

Global Mosaics of Gridded Daily Data



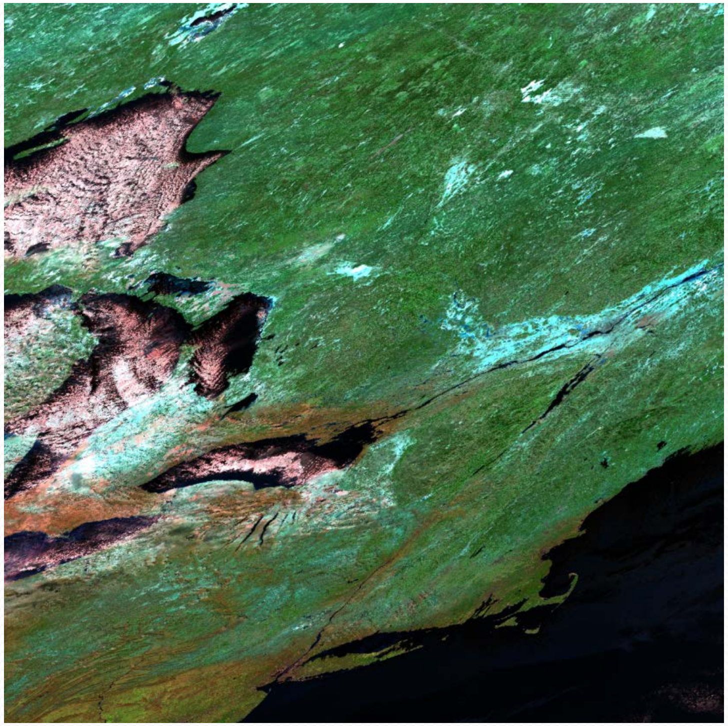
Compositing

- Required for reducing cloud/shadow contamination
- MODIS algorithm: “best” quality
  - Requires good quality flags and cloud mask
- Surface Type ATBD algorithm: Controlled View Angle Max NDVI
  - Max NDVI has many known issues
    - Tends to choose off-nadir view observations
    - Cloud selected over water
    - Problematic over non-vegetated areas
  - Improved Max NDVI
    - Compositing priority: vegetated > soil > water > snow > cloud

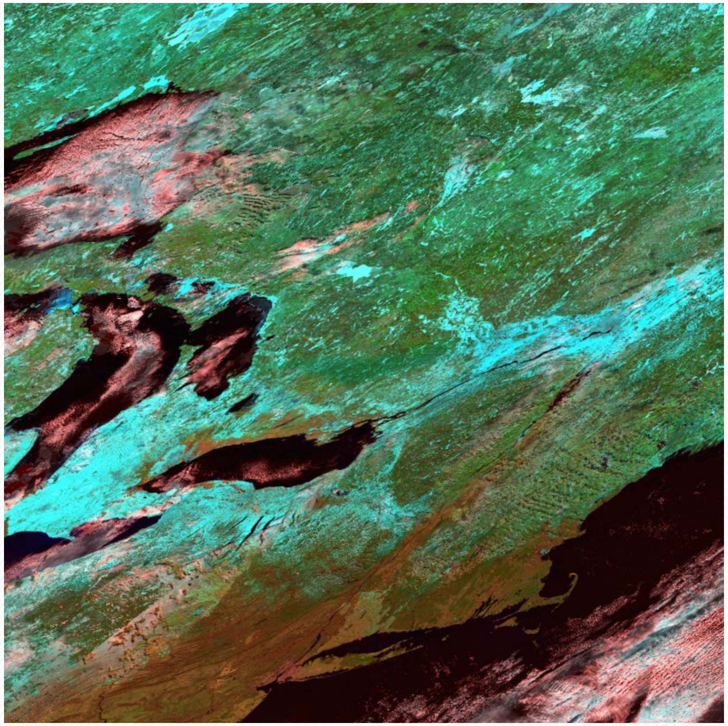
VIIRS Composites Comparable with Standard MODIS Composites (MOD09A1)

MODIS tile H12v04 (New York to Great lakes), 8-day (2012/193-200)

Better composites over water

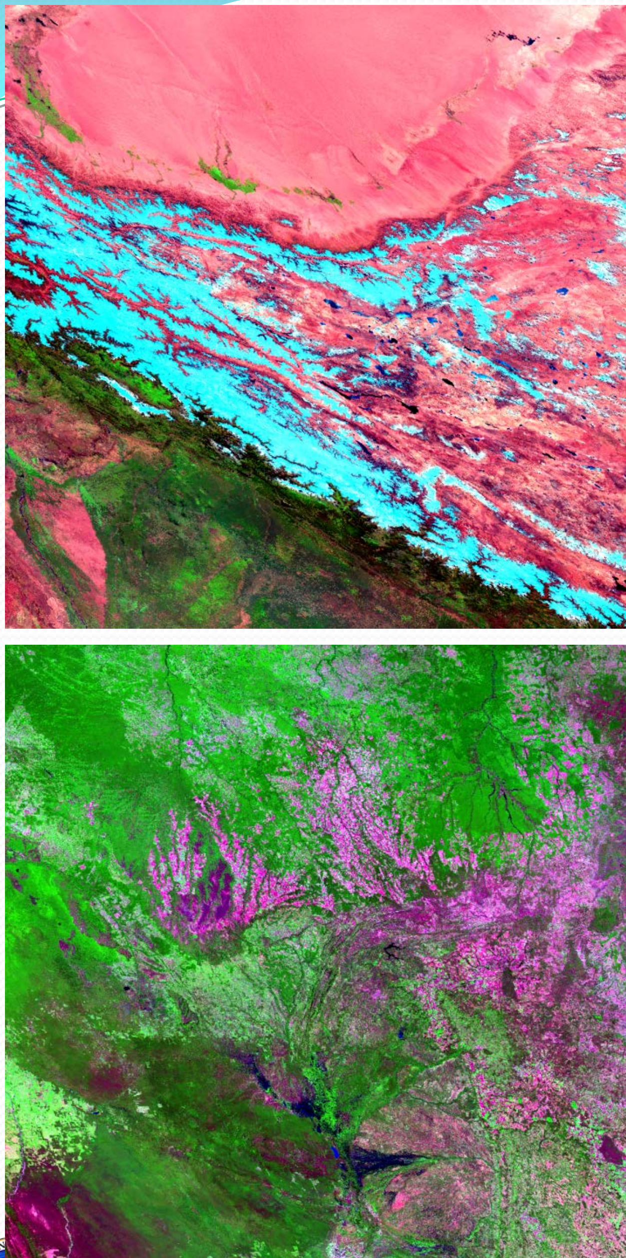


VIIRS, M11-7-5

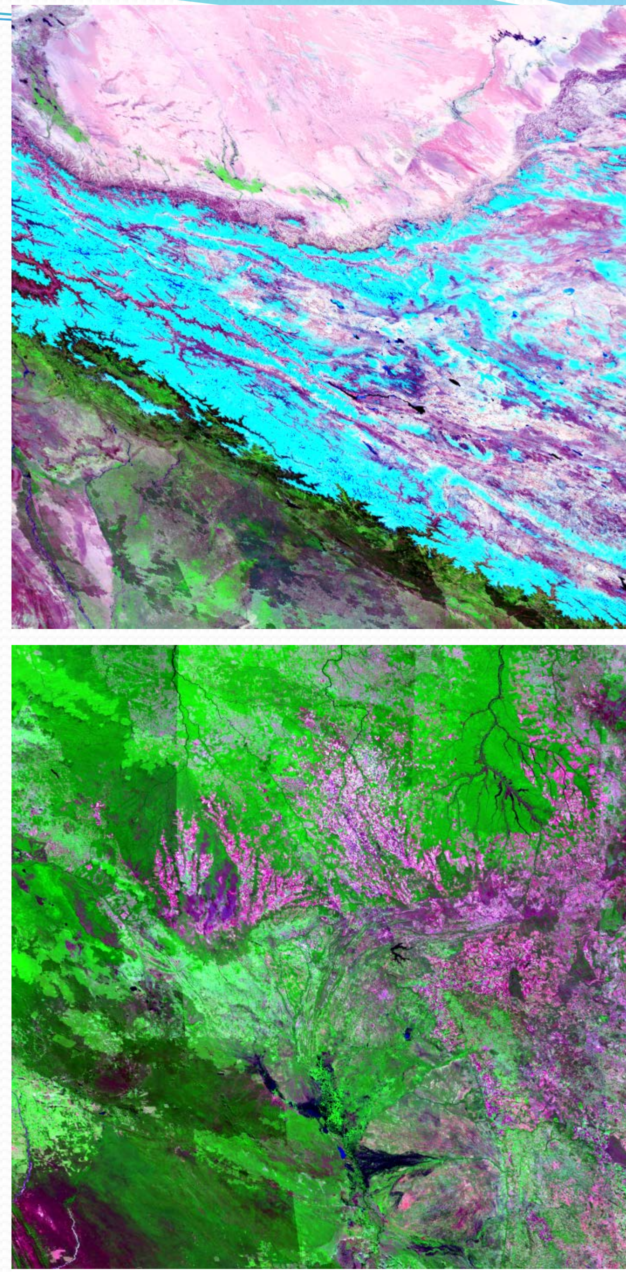


MODIS, B7-2-1

More Comparison with MODIS Composites



VIIRS, M11-7-5



MODIS, B7-2-1

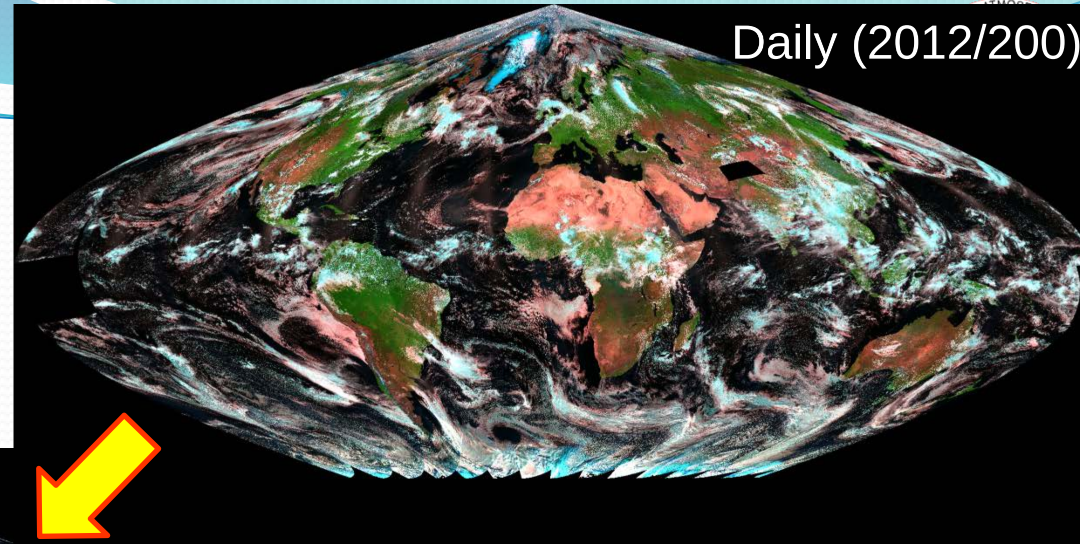
H12v04 (western China) 2012/193-200

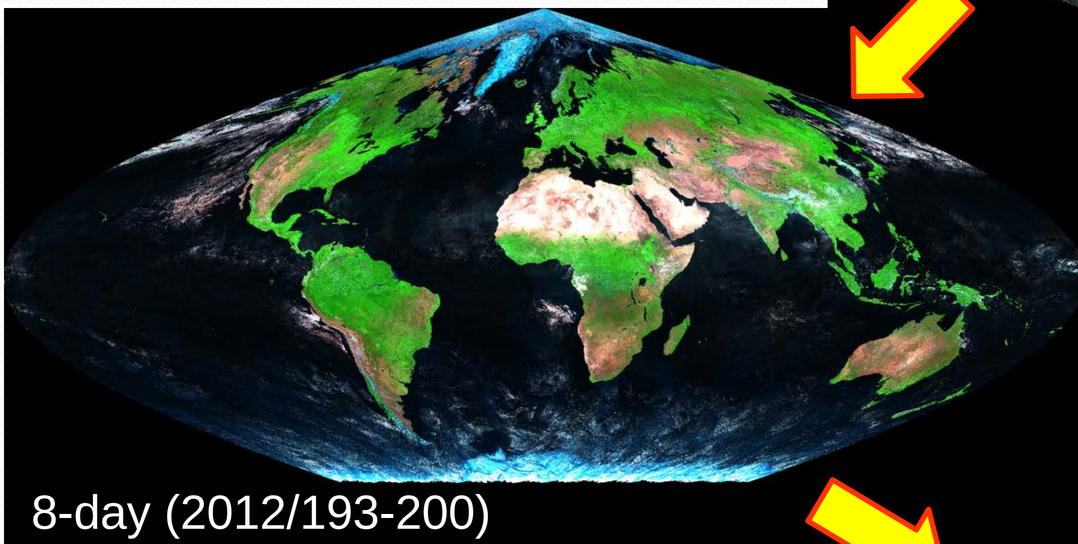
Less snow pixels selected

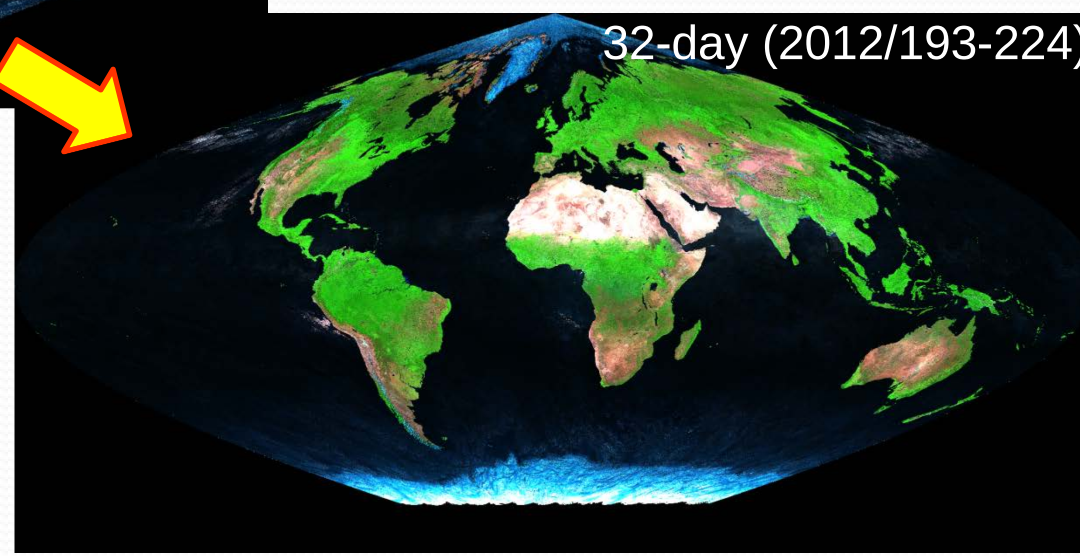
H12v10 (Brazil, Arc of Deforestation) 2012/193-200

Less artifacts (swath boundary)

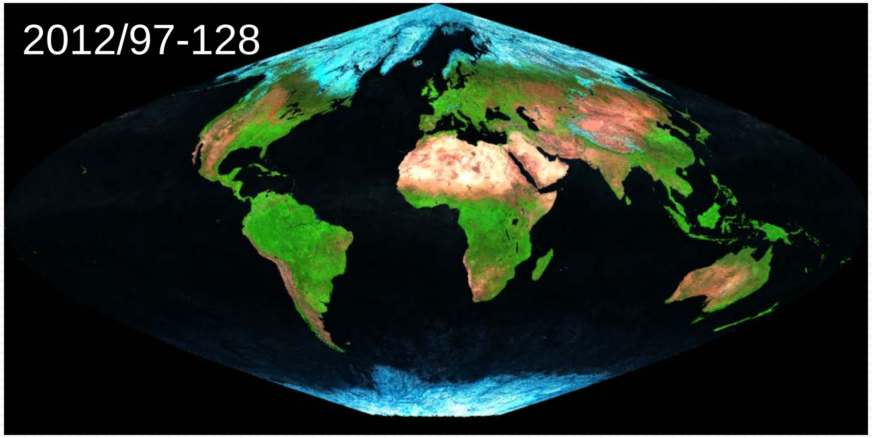
Cloud Reduction through Compositing

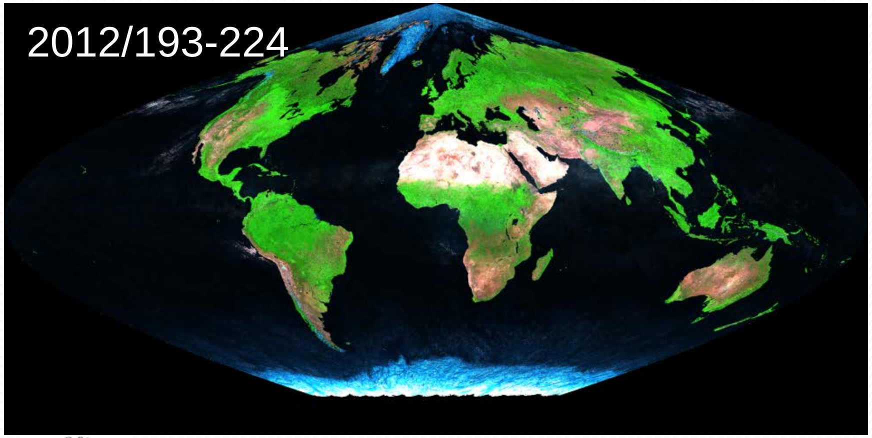


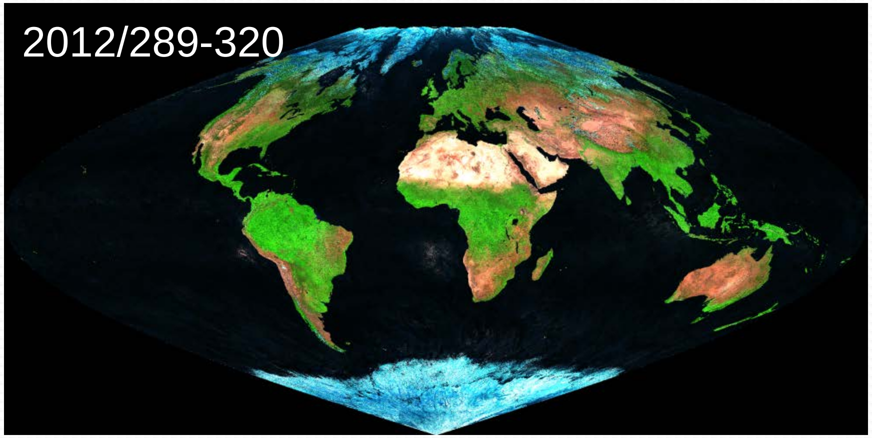




Global Mosaics of 32-day Composites





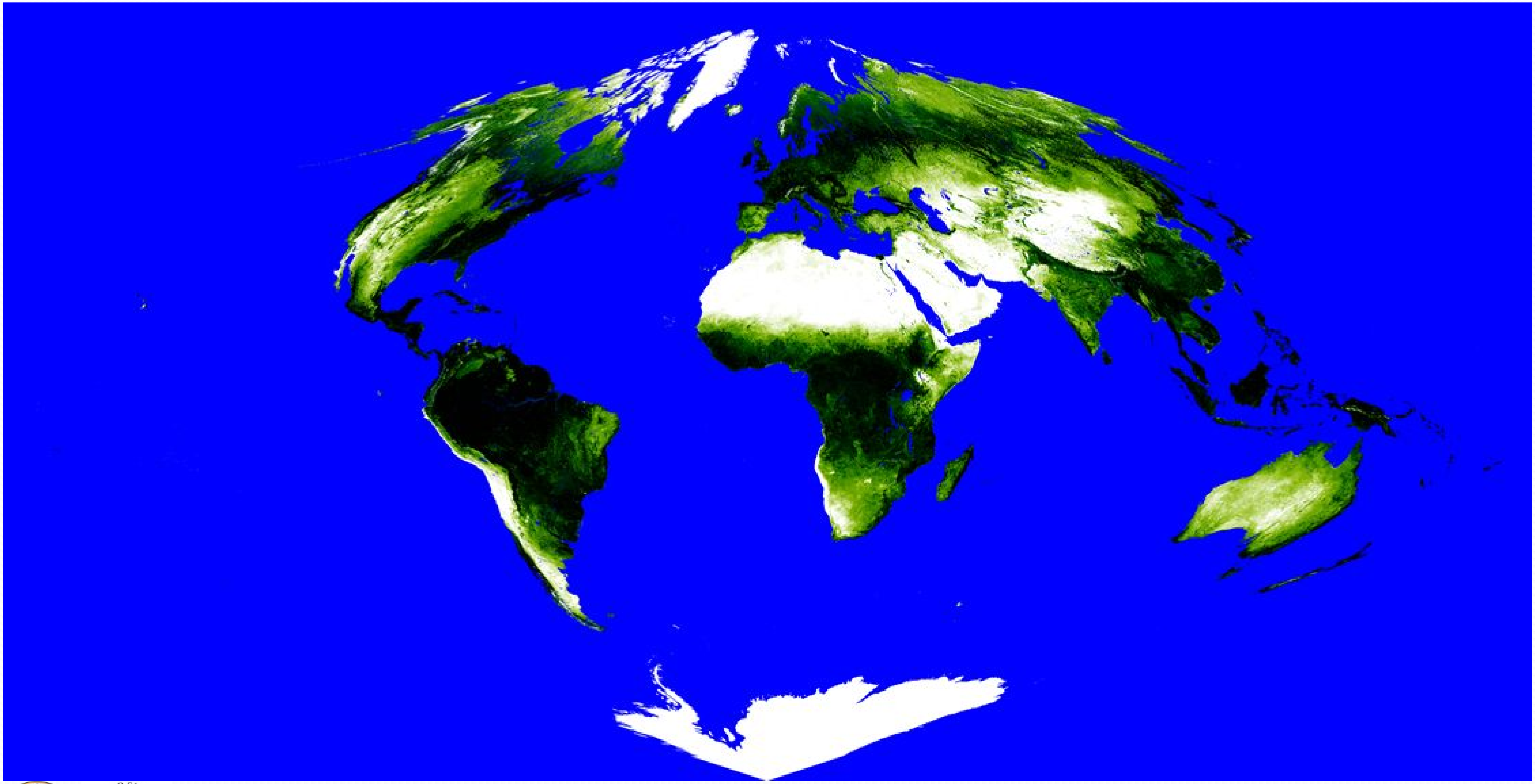


BOSTON UNIVERSITY

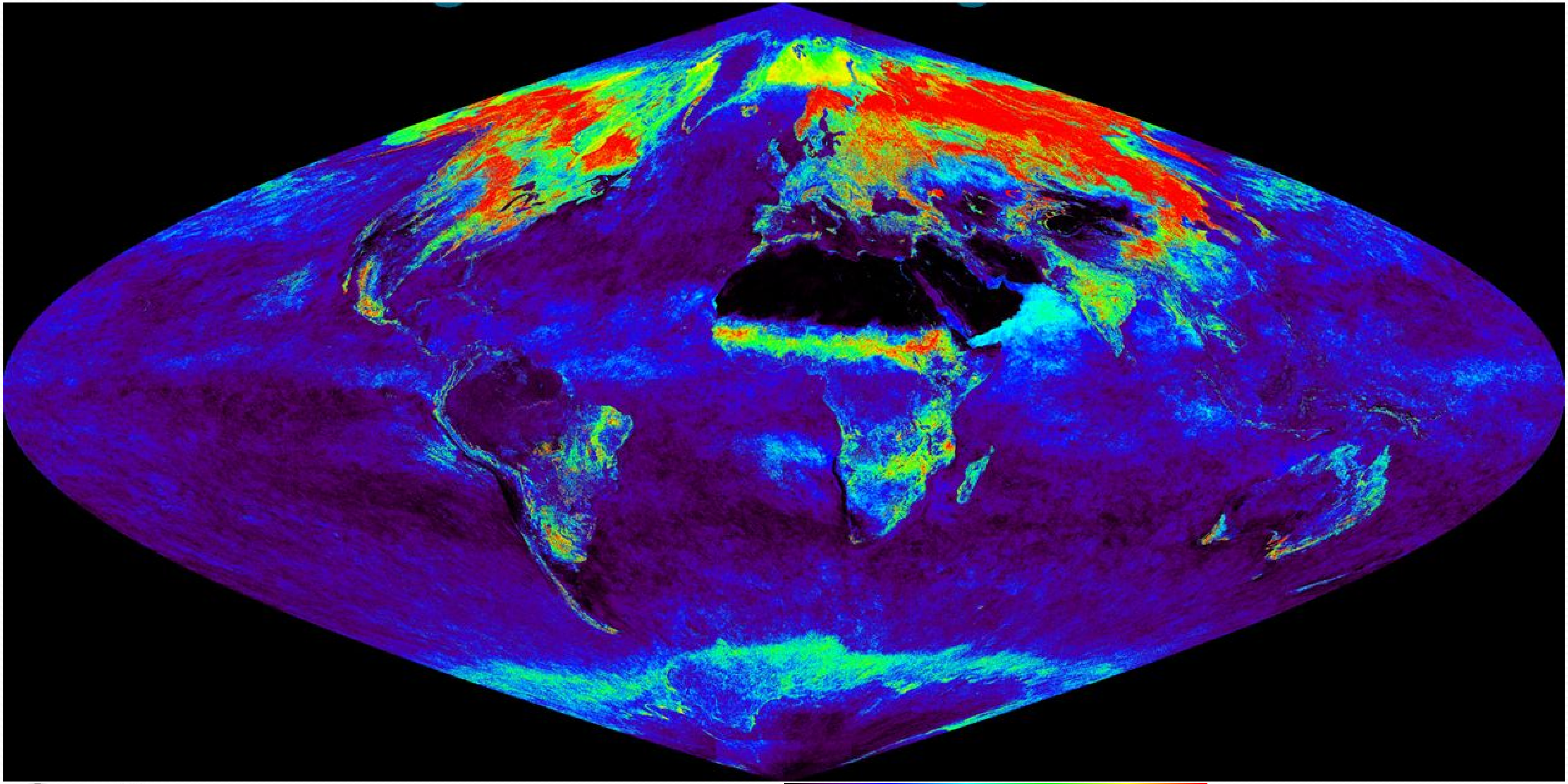
Use of Annual Metrics

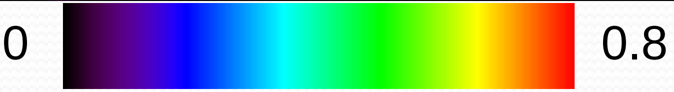
- Based on metrics developed by Hansen et al for AVHRR land cover and MODIS VCF
  - Further cloud removal in the 32-day composites
  - Normalize differences between the northern and southern hemisphere
  - Better reflect vegetation phenology and other surface characteristics
    - Some metrics could be sensitive to residual clouds in the 32-day composites

Annual Metrics from VIIRS – 2012 Median NDVI

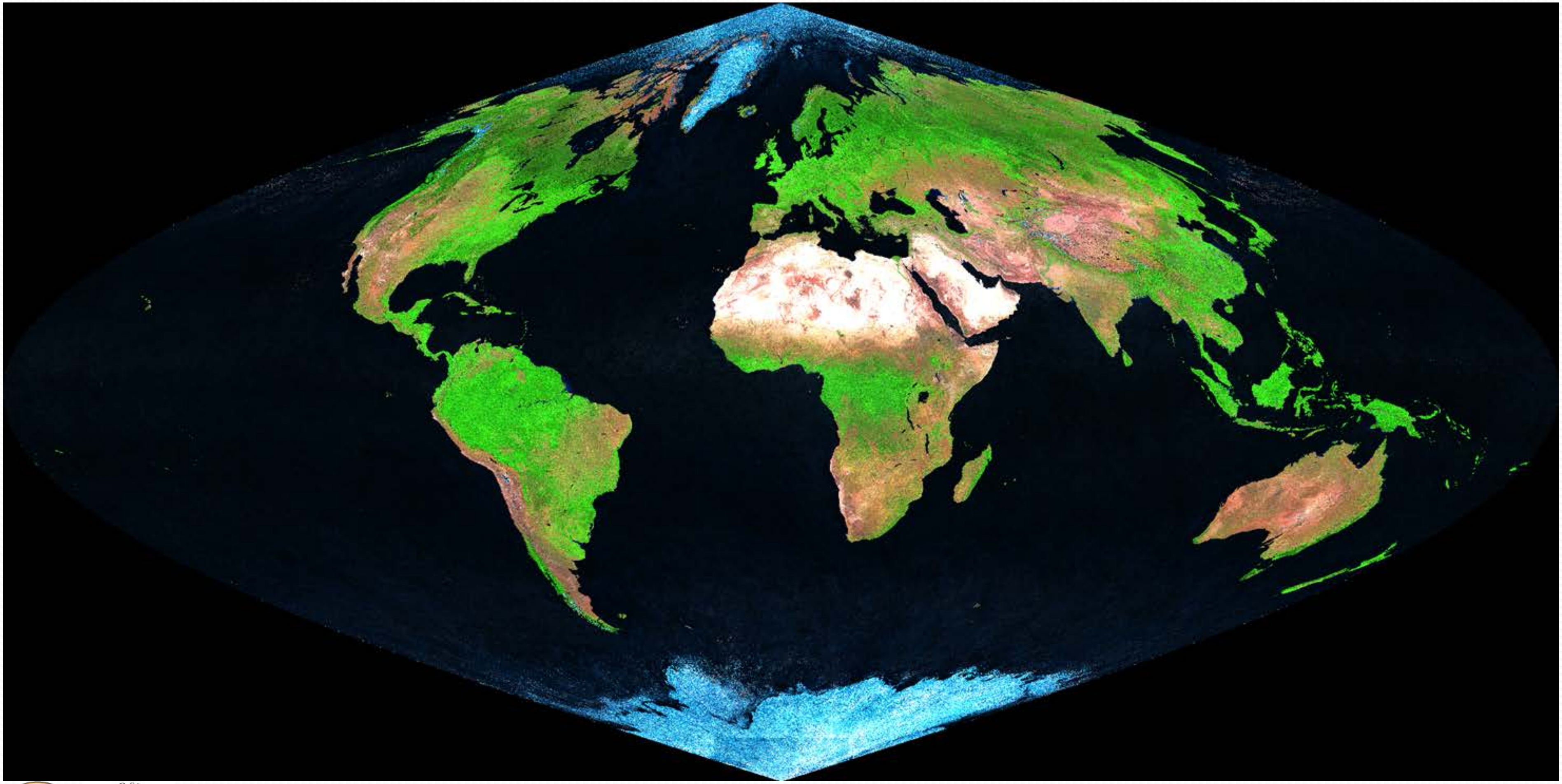


VIIRS 2012 Growth Season NDVI Variation

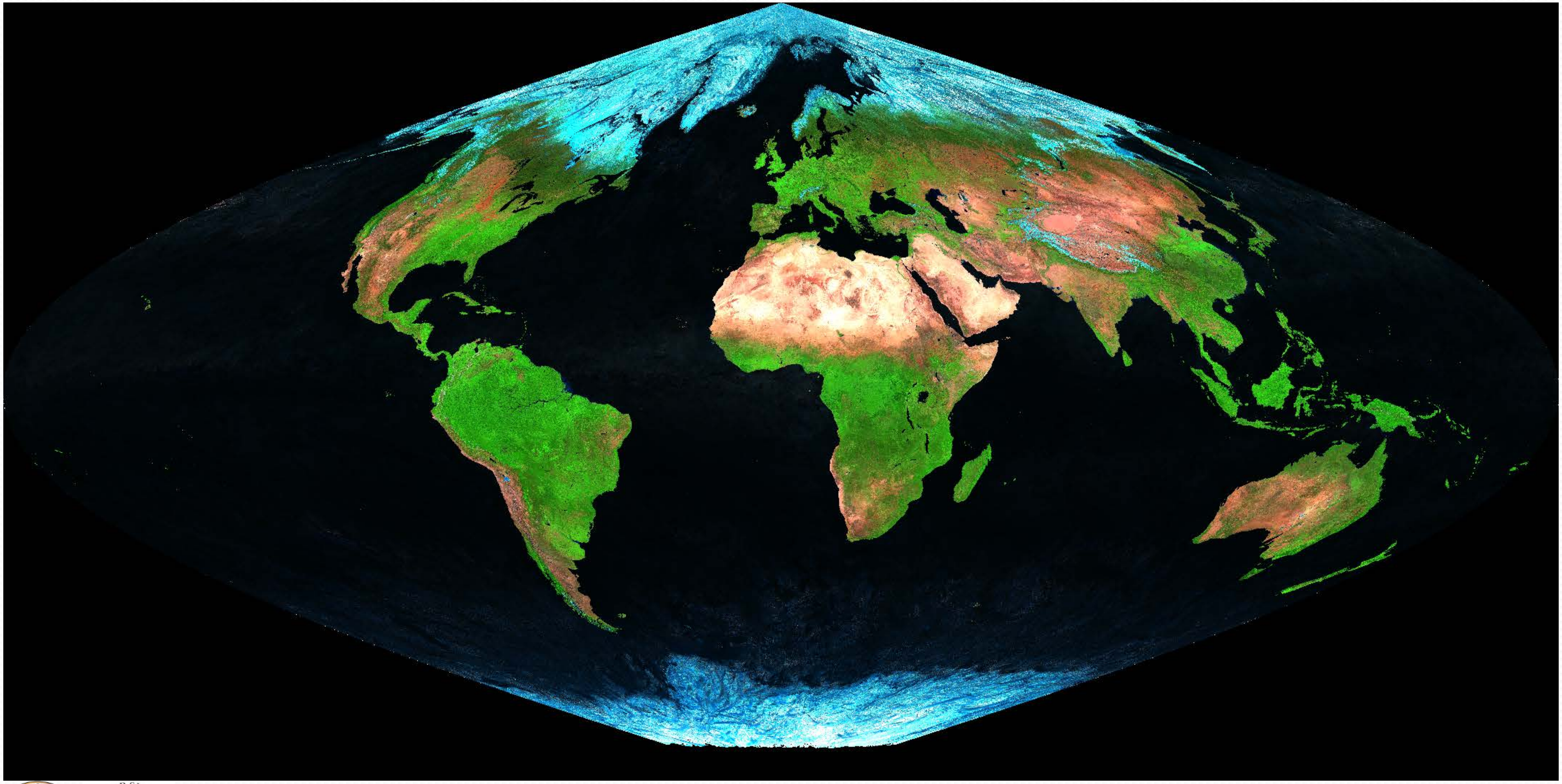




Median of the Three Warmest 32-day Composites (M11, M7, M5 in RGB)



Late Spring (97-128) 32-day Composite



Annual Metrics (76 Total) Used

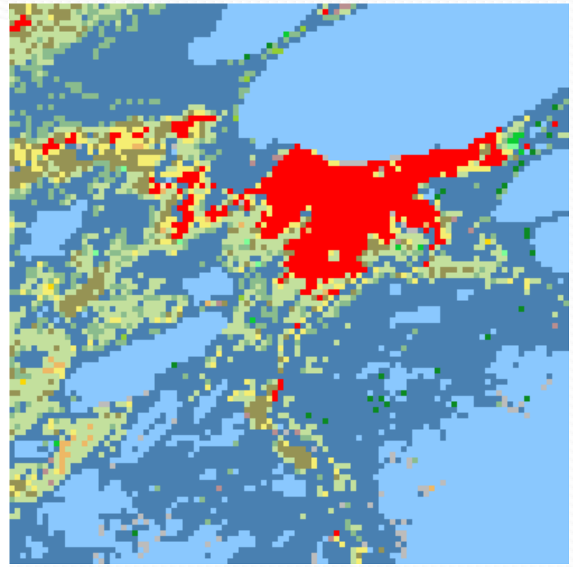
- NDVI Metrics:
  - Second maximum monthly NDVI in a year
  - Minimum NDVI value in the highest 7 monthly NDVIs
  - Median NDVI in the highest 7 monthly NDVIs
  - Range of NDVI in the highest 7 monthly NDVIs
  - Median NDVI in the 5 warmest months
  - NDVI of the warmest month
- Reflectance Metrics (VIIRS M bands 3-8, 10, 11, 15):
  - Second maximum reflectance, during the 7 greenest in the year
  - Minimum reflectance, during the 7 greenest months
  - Median reflectance of the 7 greenest months
  - Range of reflectance, among the 7 greenest months
  - Median reflectance of the 5 warmest months
  - Reflectance of the month with the second highest NDVI in the year
  - Reflectance of the warmest month

Training Data Distribution

Class	Training Points
1	518
2	937
3	190
4	738
5	673
6	204
7	1114
8	854
9	697
10	1398
11	1179
12	3739
13	
14	579
15	314
16	674

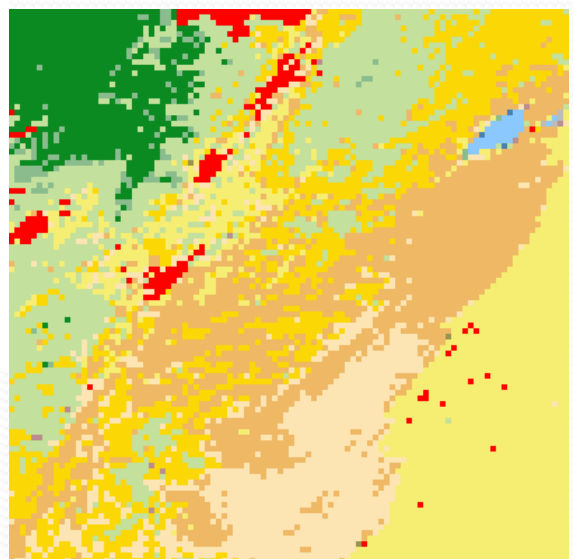
Legend

- Evergreen Needleleaf Forest
- Evergreen Broadleaf Forest
- Deciduous Needleleaf Forest
- Deciduous Broadleaf Forest
- Mixed Forest
- Closed Shrublands
- Open Shrublands
- Woody Savannas
- Savannas
- Grasslands
- Permanent Wetlands
- Croplands
- Urban and Built-Up
- Cropland/Natural Vegetation Mosaic
- Snow and Ice
- Barren or Sparsely Vegetated
- Water Bodies

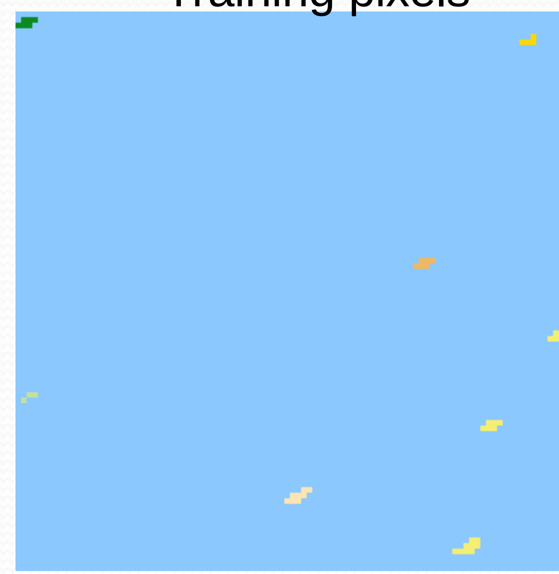


Land Cover in Louisiana

Training pixels



Land Cover in CA



Training pixels

These training points were provided by Boston University, in 500m resolution. They were majority-sampled into 1000m resolution for VIIRS classification

Classification Algorithm – Decision Tree

- Easy-to-understand decision rules
- Improved accuracy through ensemble (boosting)
- Robust for global applications
  - Base algorithm for several existing global surface type products (Hansen et al., 2000; Friedl et al., 2002; 2010)
  - Specified in ATBD

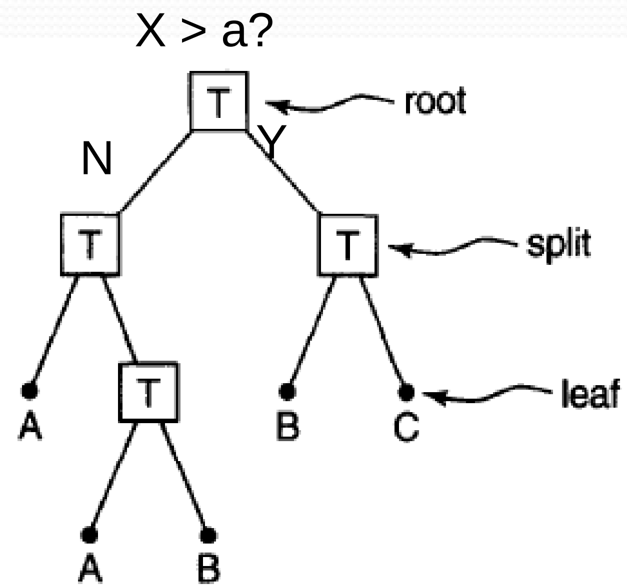


Figure 1. A decision tree classifier. Each box is a node at which tests (T) are applied to recursively split the data into successively smaller groups. The labels (A, B, C) at each leaf node refer to the class label assigned to each observation.

Decisions Made Based on Information Theory

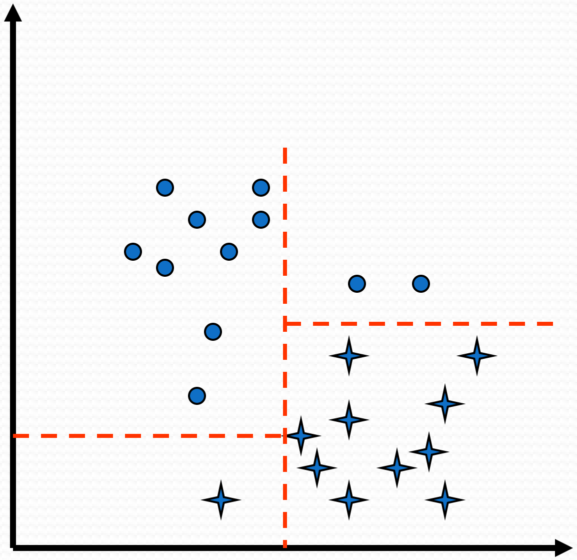
- For each variable, try all possible ways to split the training samples
- For each split, calculate the following:

$$Info(T) = -\sum_{j=1}^{k} \frac{C_j}{|T|} \times \log_2\left(\frac{C_j}{|T|}\right) \qquad Info_X(T) = -\sum_{i=1}^{n} \frac{|T_i|}{|T|} \times Info(T_i)$$

$$Gain(X) = Info(T) - Info_X(T)$$

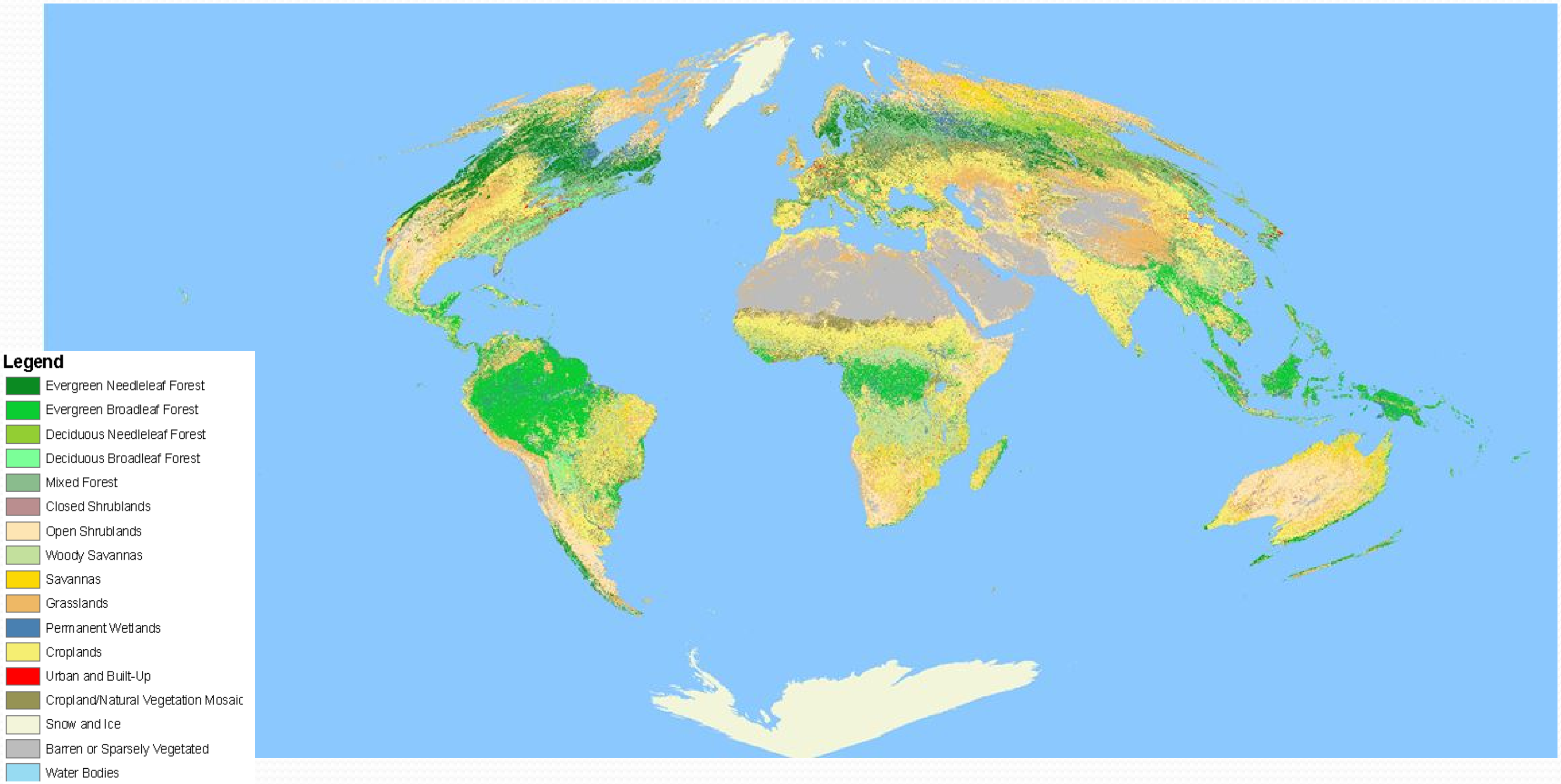
$$Split\_Info(X) = -\sum_{i=1}^{n} \frac{|T_i|}{|T|} \times \log_2\left(\frac{|T_i|}{|T|}\right)$$

$$Gain\_Ratio(X) = Gain(X) / Split\_Info(X)$$



- The selected split maximizes the Gain_Ratio
  - Different DT algorithms may use different criteria
- Applied recursively

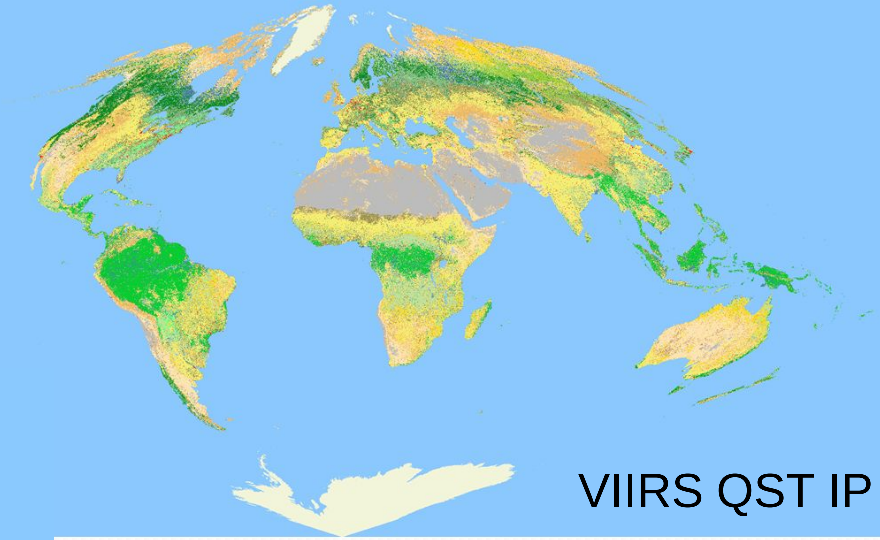
First VIIRS QST IP from 2012 VIIRS Data



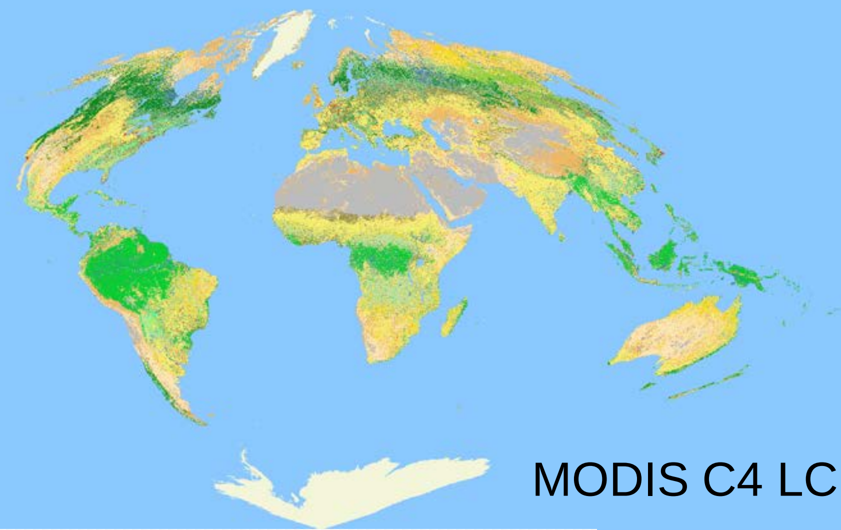
Comparison with MODIS Land Cover Products

- MODIS Collection 4 (C4) land cover product
  - Derived using annual MODIS data
  - Decision tree
  - Reported accuracy: ~75%
  - Initial seed in IDPS
- MODIS Collection 5 (C5) product
  - Derived using multi-year MODIS data
  - Decision tree
  - Post-classification process
  - Use class prior probability
  - Reported accuracy: 74.8%
  - Currently used as seed in IDPS

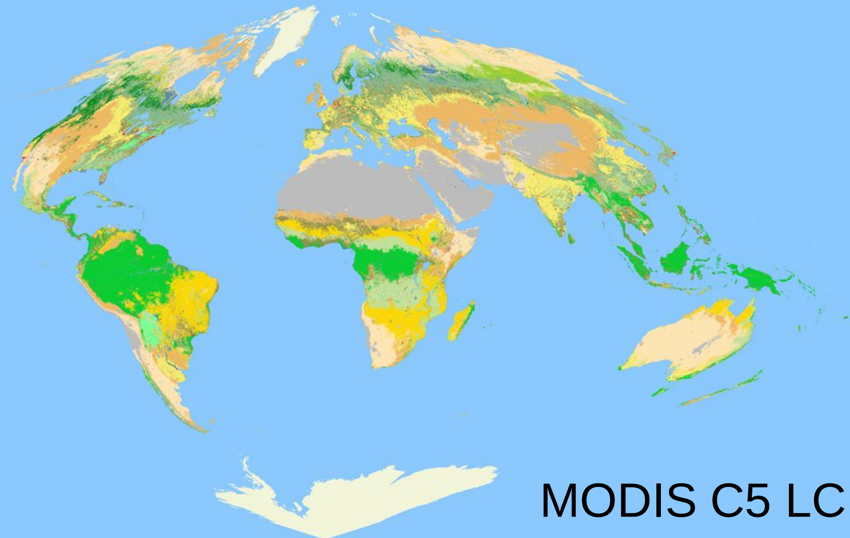
Comparison with MODIS C4/C5 LC



VIIRS QST IP



MODIS C4 LC



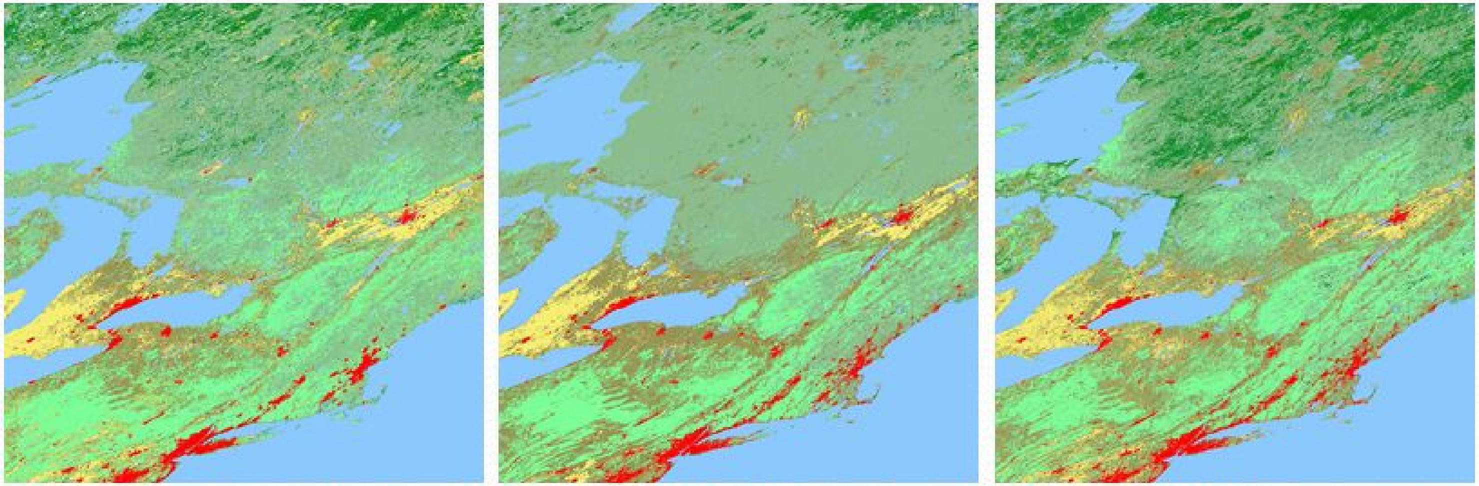
MODIS C5 LC

Legend

- Evergreen Needleleaf Forest
- Evergreen Broadleaf Forest
- Deciduous Needleleaf Forest
- Deciduous Broadleaf Forest
- Mixed Forest
- Closed Shrublands
- Open Shrublands
- Woody Savannas
- Savannas
- Grasslands
- Permanent Wetlands
- Croplands
- Urban and Built-Up
- Cropland/Natural Vegetation Mosaic
- Snow and Ice
- Barren or Sparsely Vegetated
- Water Bodies

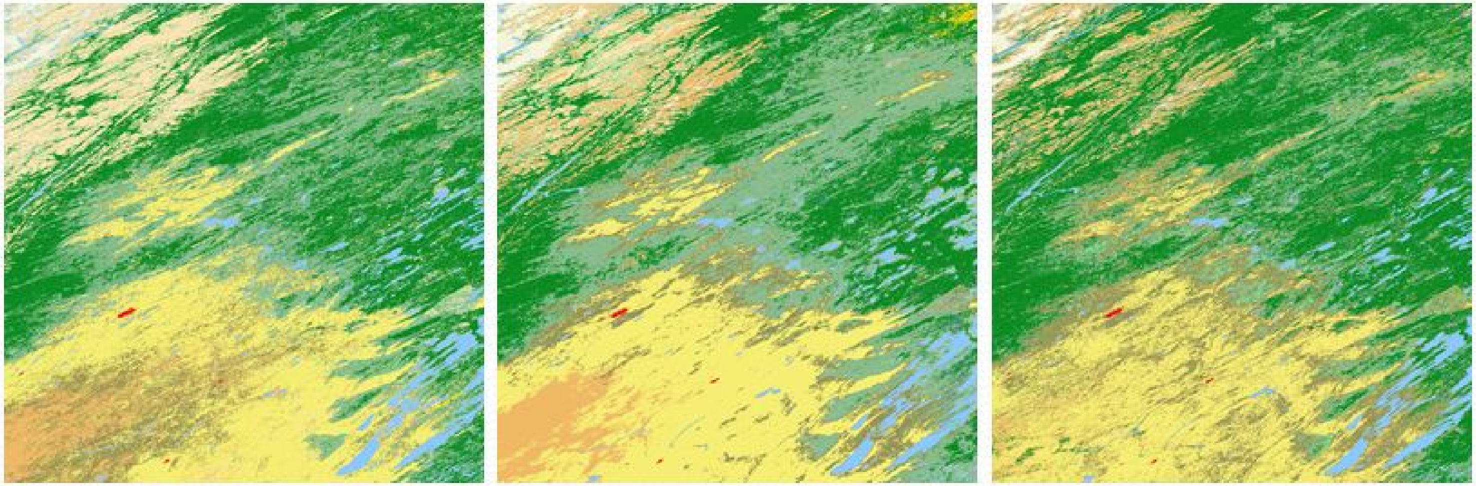
Detailed Comparisons

MODIS C4 MODIS C5 VIIRS QST IP



Great Lakes Region, MODIS Tiling system H12V04

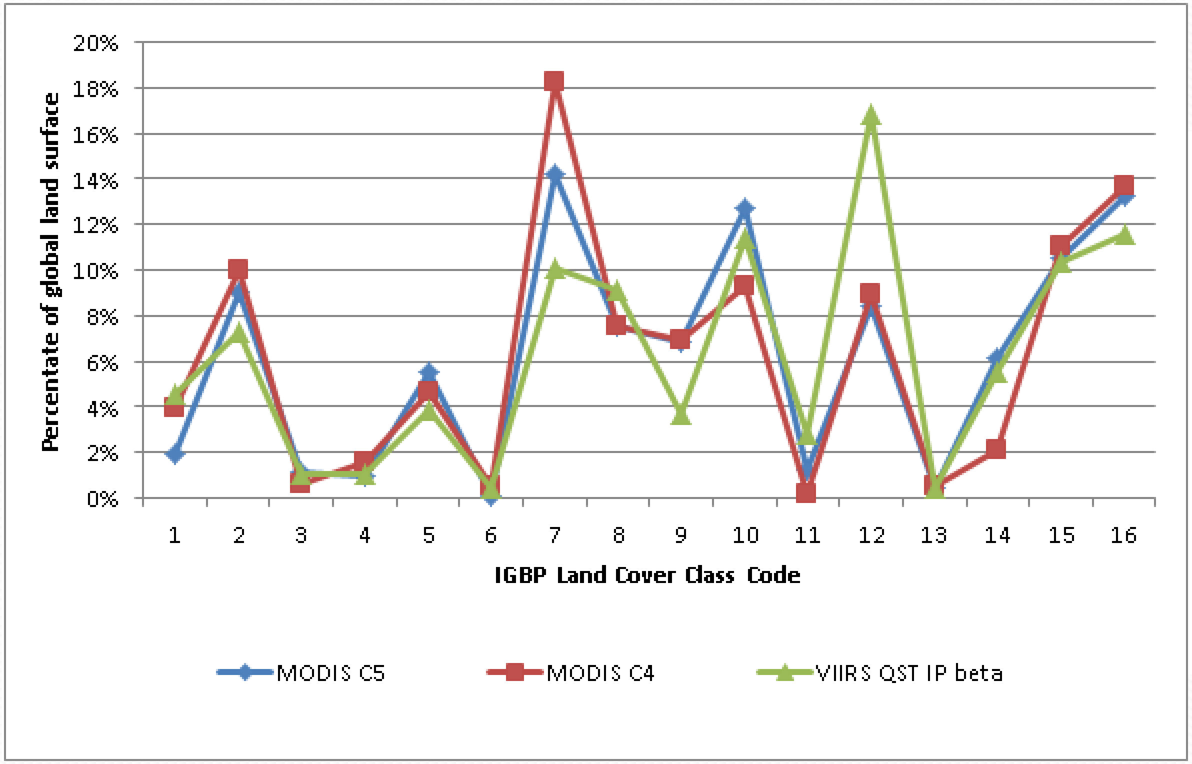
MODIS C4 MODIS C5 VIIRS QST IP



Western Canada Region, MODIS Tiling system H11V03

BOSTON UNIVERSITY

Comparison of ST proportions



Legend

1 Evergreen Needleleaf Forest
2 Evergreen Broadleaf Forest
3 Deciduous Needleleaf Forest
4 Deciduous Broadleaf Forest
5 Mixed Forest
6 Closed Shrublands
7 Open Shrublands
8 Woody Savannas
9 Savannas
10 Grasslands
11 Permanent Wetlands
12 Croplands
13 Urban and Built-Up
14 Cropland/Natural Vegetation Mosaic
15 Snow and Ice
16 Barren or Sparsely Vegetated
Water Bodies

Surface Type Provisional Status Summary

- ST EDR
  - Operational algorithm (IDPS) validated
  - Issues identified and addressed
  - Further improvements possible with
    - Improvements to Fire, Snow, and other VIIRS products
    - Updates to QST IP
    - Timely detection of major land cover change processes: flooding, burn scar, large scale deforestation

Surface Type Provisional Status Summary

- QST IP
  - In-house implementation of ATBD algorithms successful
    - Gridding/compositing/annual metrics (GCAM)
    - Classification
  - Newly collected training data
    - Original training data were of limited use (error, out of date, undesirable distribution)
    - New data from BU
    - Additional data collected by UMD
  - First QST IP produced using 2012 VIIRS data
    - Quality likely close to MODIS land cover products

Provisional Consideration and Future Plan

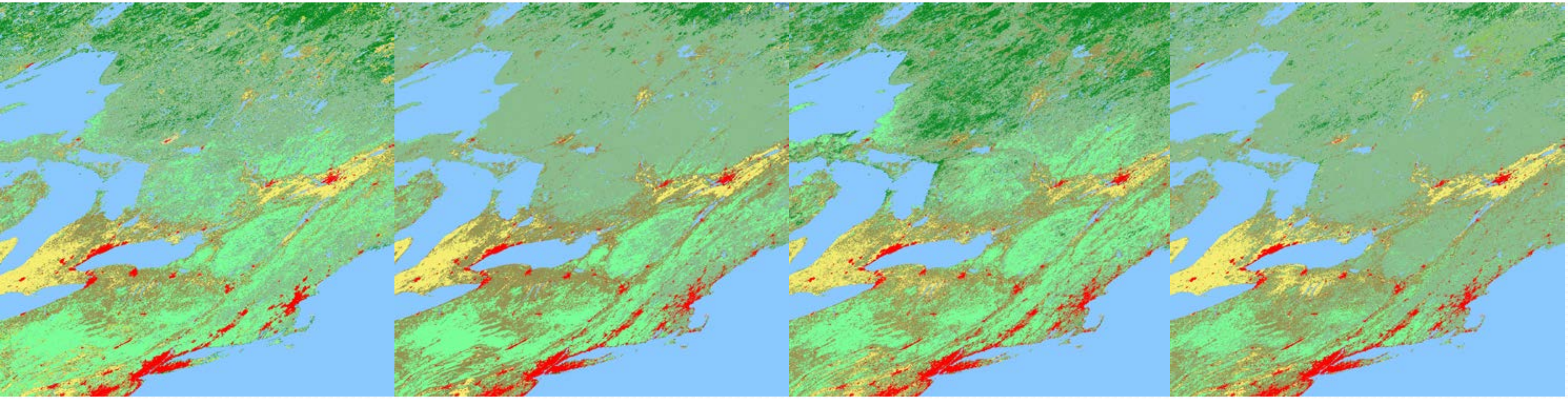
- Gridding/compositing/annual metrics (GCAM) huge job
  - Better done on systems with lots of CPU, bandwidth, and storage space
  - Compositing and annual metrics have room to improve
    - Especially over sparsely or non-vegetated surfaces
- Better cloud removal
  - Residual clouds apparent in some areas
  - Usefulness of some metrics questionable
  - Better compositing algorithm can help (e.g., use QA flags/cloud masks)
  - Use multi-year data
    - 2-years VIIRS data already exist
    - Can add MODIS data

Provisional Consideration and Future Plan

- Training data improvement
  - More representative
  - Timely update of changes
- Classification algorithm improvements
  - SVM generally more accurate than DT
  - Post-classification improvements (e.g., BU's approach of incorporating class probability in MODIS C5)

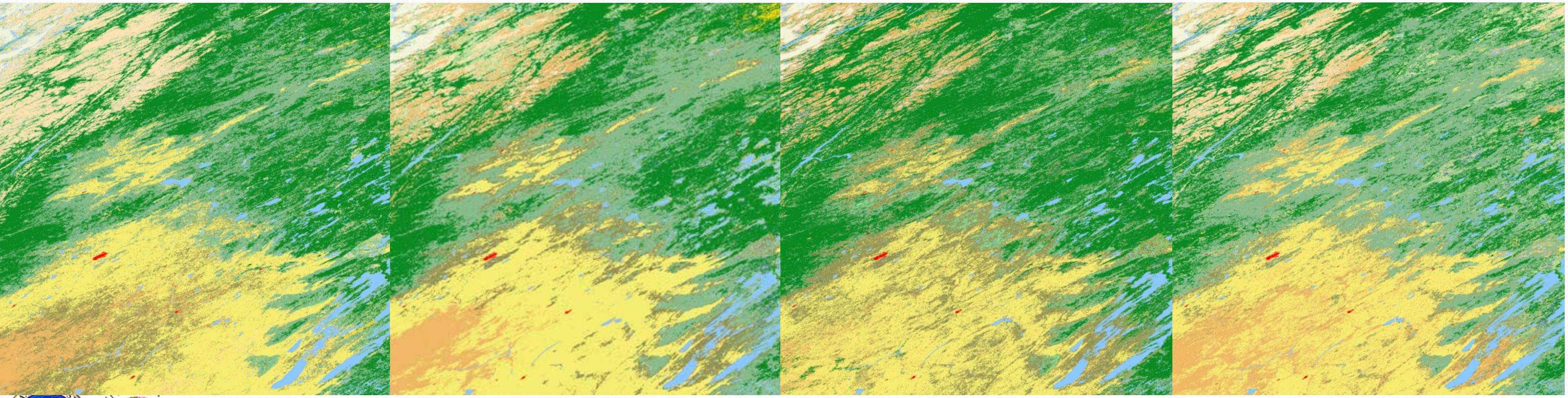
Initial Test Run Results Using SVM

MODIS C4 | MODIS C5 | VIIRS DT | VIIRS SVM



Great Lakes Region, MODIS Tiling system H12V04

MODIS C4 | MODIS C5 | VIIRS DT | VIIRS SVM



Western Canada Region, MODIS Tiling system H11V03

Provisional Consideration and Future Plan

- Take advantage of additional bands
  - 375m imagery bands allow better estimation of subpixel cover
  - DNB useful urban
- More comprehensive assessment
  - Better use of freely available high resolution data (e.g. Google Earth, Landsat)