MODIS Global Land Cover: Algorithm Refinements and Characterization of New Datasets

Mark Friedl[1], Damien Sulla-Menashe[1], Bin Tan[2], Annemarie Schneider[3], Navin Ramankutty[4], Adam Sibley[1] and Xiaoman Huang[1]

[1] Department of Geography and Environment (http://www.bu.edu/geography), Boston University, 675 Commonwealth Avenue, Boston, MA 02215

[2]Earth Resources Technology, Inc., 10810 Guilford Road, Suite 105, Annapolis Junction, MD 20701

[3]Center for Sustainability and the Global Environment, University of Wisconsin-Madison, 1710 University Avenue, Room 264, Madison, Wisconsin 53726

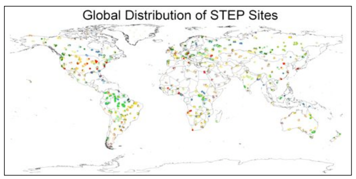
[4]Department of Geography and Program in Earth System Science, 627 Burnside Hall, 805 Sherbrooke Street W., Montreal, QC, H3A 2K6, Canada

Abstract

Information related to land cover is immensely important to global change science. Here we describe the datasets and algorithms used to create the Collection 5 MODIS Global Land Cover Type product, which is substantially changed relative to Collection 4. In addition to using updated input data, the algorithm and ancillary datasets used to produce the product have been refined. Most importantly, the Collection 5 product is generated at 500-m spatial resolution, providing a four-fold increase in spatial resolution relative to the previous version. In addition, many components of the classification algorithm have been changed. The training site database has been revised, land surface temperature is now included as an input feature, and ancillary datasets used in post-processing of ensemble decision tree results have been updated. Further, methods used to correct classifier results for bias imposed by training data properties have been refined, techniques used to fuse ancillary data based on spatially varying prior probabilities have been revised, and a variety of methods have been developed to address limitations of the algorithm for the urban, wetland, and deciduous needleleaf classes. Finally, techniques used to stabilize classification results across years have been developed and implemented to reduce year-to-year variation in land cover labels not associated with land cover change. Results from a cross-validation analysis indicate that the overall accuracy of the product is about 75% correctly classified, but that the range in class-specific accuracies is large. Comparison of Collection 5 maps with Collection 4 results show substantial differences arising from increased spatial resolution and changes in the input data and classification algorithm.

Refinements to the STEP Database

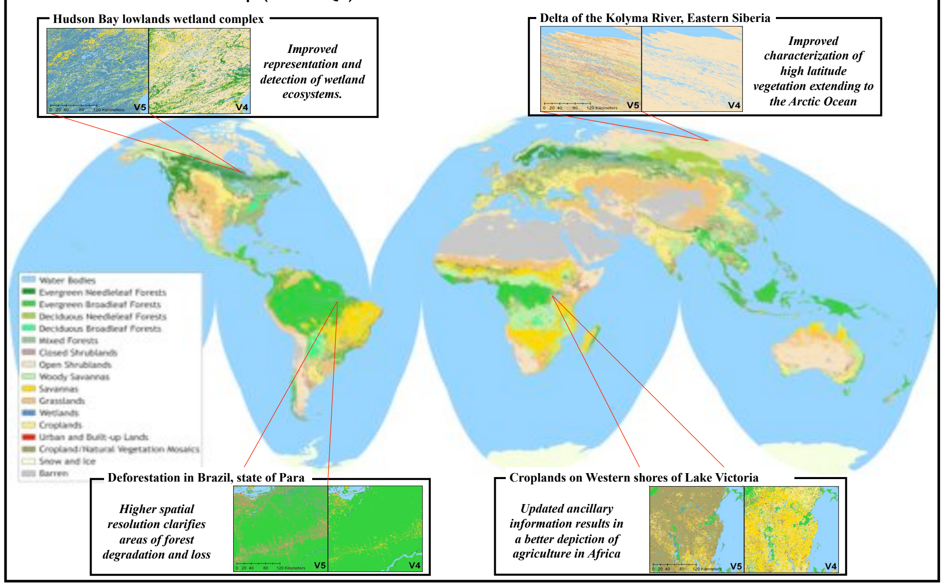
- The System for Terrestrial Ecosystem Parameterization (STEP) database provides training data for the MCD12Q1 product algorithm.
- In preparation for collection 5, the STEP database was analyzed and extensively revised to ensure representative sampling of land cover. Specifically, we examined:
  - Geographic and ecological sampling based on biogeographic realms, global biomes from Olson et al. (2001), and country level statistics.
  - Outlier and error analysis based on PCA of raw NBAR data, annual EVI profiles, and measures of internal heterogeneity.
- Results from these analyses lead to augmentation of the database for under-represented areas, reduction in size of large sites, and the removal of ambiguous or physically altered sites.



Summary of Major Changes to the MCD12 Land Cover Type Product (MCD12Q1)

	V4	V5
Inputs	• 32-day NBAR data (7 land bands) at 1-km resolution derived from two 16-day values • 32-Day EVI data at 1-km • Annual metrics (min, max, mean) for each of the bands	• 32-day NBAR (7 land bands), EVI and LST data at 500-m resolution derived from four 8-day values • Annual metrics (min, max, mean) for each of the bands
Ancillary Data	• MODIS V4 land-water mask • V4 urban mask • Prior probability layers based on Collection 3 data and ancillary data related global croplands intensity	• MODIS V5 land-water mask • Updated V5 urban layer • Prior probability layers based on Collection 4 data with inclusion of new agricultural intensity data
Algorithms	• Significant reliance on out-of-date prior probability layers • Ad hoc reduction of overestimated deciduous needleleaf forests and wetlands • Minimal stabilization of inter-annual variation	• Decreased reliance (weighting) on prior probability layers • Improved treatment/representation of deciduous needleleaf forests • Improved treatment/representation of wetlands • Improved techniques for stabilization of year-to-year variation in labels
STEP Database	• Largely based on older TM5 data • Large sites with significant internal heterogeneity. • Inadequate quality control • Relatively poor and misrepresentative sampling in key areas	• Updated to more recent (2000+) Landsat imagery and GoogleEarth© • Updated to conform to requirements for 500-m data • Extensive quality control including editing and removal of bad sites • Augmentation of sites in under-represented areas and classes

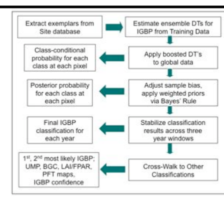
2005 IGBP Land Cover Map (MCD12Q1)



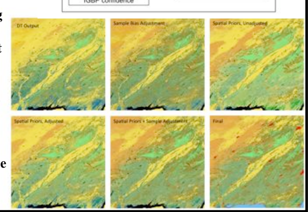
Classification Procedure

The flowchart to the right summarizes the land cover classification procedure:

1. Sites are extracted from the training database and used to train ensemble decision trees on the input features summarized in the adjacent table.
2. The decision trees perform the classification on the global data using a single year of inputs.
3. The results are adjusted to correct various biases originating from the training data and input features.
4. Finally the maps are cross-walked to other class schemes.

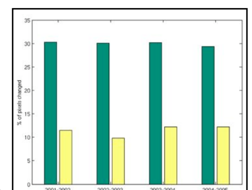


The figure to the right shows the intermediaries produced at each stage of processing for the tile h10v05 in the Southeastern United States.



Stabilization of Inter-annual Variability

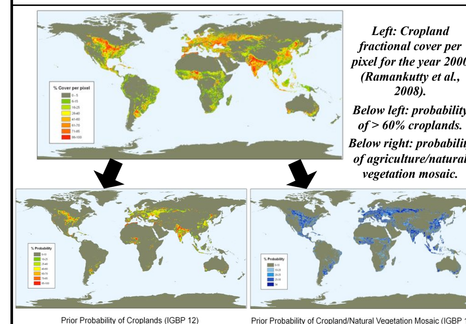
- There is substantial year-to-year variation in land cover labels resulting from mixtures of classes at 500-m spatial resolution and the difficulty in distinguishing spectrally similar classes.
- The MCD12Q1 algorithm imposes constraints on this variation. Over a three year window apparent changes in land cover classes are only retained if the confidence in the recent class is higher than in its alternatives thus perpetuating higher quality labels.



Above: the proportion of global pixels changing from one year to the next with stabilization (yellow) and without (green).

Updated Ancillary Data Layers

- Ancillary data layers are used within the MCD12Q1 algorithm to aid classification results when spectral data from MODIS do not effectively separate classes. These data are derived from a variety of sources including the Collection 4 MOD12Q1 product, agricultural intensity data, and the MODIS land water mask.
- A 150 km x 150 km moving window was used to compute the approximate regional frequency of classes based on MODIS collection 4 data; this provides local likelihoods for each class at each pixel.
- To prescribe the likelihood of agriculture or agricultural mosaic, we used a new data set from Ramankutty et al. (2008; see below).

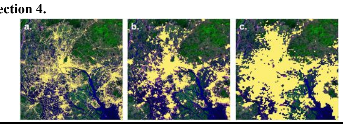


Left: Cropland fractional cover per pixel for the year 2000 (Ramankutty et al., 2008).

Below left: probability of > 60% croplands.

Below right: probability of agriculture/natural vegetation mosaic.

- Urban areas are classified separately using an ecoregion-based stratification (Schneider et al., 2009). Below is an example of the improvements in urban representation (yellow) for Guangzhou China: (a) Landsat-based classification, (b) MODIS Collection 5, (c) MODIS Collection 4.



Cross-Validation Accuracy Assessment

- A ten-fold cross-validation procedure was performed using the same classification methodology but keeping aside ten percent of the training data for testing.
- Low cross-validation accuracies (e.g., IGBP classes 5, 6, 8, 9, 14) are evidence of confusion among classes with similar ecological and spectral-temporal properties.

User and producer accuracies, standard errors, and 95% confidence intervals for Collection 5 IGBP classes based on cross-validation.

IGBP land cover class	Producer's accuracy (%)				User's accuracy (%)			
	PA	Std. err.	CI−	CI+	UA	Std. err.	CI−	CI+
1.	89.8	2.3	85.3	94.4	78.0	5.3	67.5	88.6
2.	92.6	2.4	88.0	97.2	83.1	3.2	76.8	89.5
3.	67.3	10.9	45.8	88.7	90.4	4.6	81.4	99.4
4.	68.9	6.2	56.7	81.0	75.9	5.3	65.6	86.3
5.	76.2	5.7	65.1	87.3	53.1	6.1	41.1	65.1
6.	63.4	5.9	51.9	74.9	47.0	5.5	36.1	57.8
7.	48.3	6.2	36.1	60.5	74.1	5.2	63.8	84.4
8.	45.2	4.1	37.2	53.3	34.3	4.5	25.4	43.2
9.	22.6	4.4	13.9	31.3	39.0	6.0	27.2	50.8
10.	73.6	4.1	65.7	81.6	55.9	4.2	47.6	64.2
11.	70.6	4.2	62.4	78.7	96.4	1.8	92.7	99.9
12.	83.3	2.0	79.4	87.1	92.8	1.5	89.8	95.8
14.	60.5	5.7	49.3	71.7	27.5	3.6	20.5	34.6
15.	75.6	10.9	54.4	96.9	96.8	2.3	92.2	100.0
16.	95.8	1.4	93.1	98.4	92.7	2.1	88.5	96.8
17.	96.6	1.9	92.8	100.0	99.3	0.4	98.6	100.0

References

Friedl, MA, D Sulla-Menashe, B Tan, A Schneider, N Ramankutty, A Sibley, and X Huang. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* 114: 168-182.

Olson, DM, E Dinerstein, ED Wikramanayake, ND Burgess, GVN Powell, EC Underwood, JA D'amico, I Itoua, HE Strand, JC Morrison, CJ Loucks, TF Allnutt, TH Ricketts, Y Kura, JF Lamoreux, WW Wettengel, P Hedao, and KR Kassem. 2001. Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience* 51(11): 933-938.

Ramankutty, N, AT Evan, C Monfreda, and JA Foley. 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, 22.

Schneider, A, MA Friedl, and D Potere. 2009. A new map of global urban extent from MODIS satellite data. *Environmental Research Letters* 4.