

MOD44B: Vegetation Continuous Fields Collection 3, Version 3.0.0

USER GUIDE

Submitted as “Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields algorithm” to *Earth Interactions*, 2003.

M. C. Hansen¹, R. S. DeFries^{1,2}, J. R. G. Townshend^{1,3}, M. Carroll¹, C. Dimiceli¹, and R. A. Sohlberg¹

¹ Department of Geography, University of Maryland, College Park, MD 20742

² Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20742

³ Institute for Advanced Computer Studies, University of Maryland, College

Abstract

The first result of the MODIS (Moderate Resolution Imaging Spectroradiometer) vegetation continuous field algorithm, global percent tree cover, is presented. Percent tree cover per 500 meter MODIS pixel is estimated using a supervised regression tree algorithm. Data derived from the MODIS visible bands contribute most to discriminating tree cover. The results show that MODIS data yield greater spatial detail in the characterization of tree cover compared to past efforts using AVHRR data. This finer scale depiction should allow for using successive tree cover maps in change detection studies at the global scale. Initial validation efforts show a reasonable relationship between the MODIS estimated tree cover and tree cover from validation sites.

Introduction

Standardized maps of global forest cover serve many purposes, among them the ability to estimate parameters for use in biogeochemical modeling procedures (Bonan et al. 2002; DeFries et al. 2002), and their use in delineating remaining intact forest and woodland tracts for conservation and forestry concerns (Matthews 2001). Such maps also reveal land use intensification when compared to potential vegetation conditions, revealing the human impact on naturally forested ecosystems. Repeated efforts over time can document future alterations to forest ecosystems. The synoptic view of global satellite data sets affords the best possibility of creating such maps. Initial efforts (DeFries et al. 1999; DeFries et al. 2000; Zhu and Waller 2000; Hansen et al. 2002) have demonstrated this capability. Proportional per pixel tree cover estimates, or continuous fields of percent tree cover, are an improved thematic representation over discrete classifications (DeFries et al. 1995). Continuous field maps yield improved depictions of spatially complex landscapes and the ability to use successive depictions to measure change (DeFries et al. 2002; Hansen 2003).

This paper describes the creation of a new global percent tree cover map based on 500 meter MODIS data, and represents the finest scale global forest information to date. The MODIS sensor represents a significant gain in spatial detail due primarily to two facts. The first is the finer instantaneous field of view of MODIS (250 and 500 meters squared) as compared to heritage AVHRR instruments (1 kilometer squared). Secondly, due to the fact that MODIS was built with seven bands specifically designed for land cover monitoring, there is an improved spectral/spatial response compared to AVHRR. This

allows for greater accuracy in mapping due to more robust spectral signatures. It also aids in reducing background scattering from adjacent pixels as the MODIS land bands were designed to limit the impact of atmospheric scattering. The result is a data set which reveals far more spatial detail than previous efforts. Maps such as the MODIS global continuous field of percent tree cover map should be of use to more varied scientific applications than previous coarse-scale maps.

Data

This initial attempt using MODIS imagery employed approximately a year of data. The inputs consist of 8-day minimum blue reflectance composites which are made in order to reduce the presence of clouds in the data stream. However, this procedure can lead to the inclusion of pixels within areas of cloud shadow. To reduce the presence of cloud shadows, the data were converted to 40 day composites using a second darkest albedo (sum of blue, green and red bands) algorithm. The inputs date from October 31, 2000 to December 9, 2001. An extra forty day composite period was added to attempt to compensate for data gaps resulting from temporary sensor outages.

The seven MODIS land bands were used as inputs: (blue (459-479 nm), green (545-565 nm), red (620-670 nm) and near infrared (841-876 nm) and middle infrared (1230-1250 nm, 1628-1652 nm, 2105-2155 nm)). The MODIS composited data were transformed into annual metrics which capture the salient points in the phenologic cycle. A total of 68 metrics were derived from the composited data for bands 1-7 and the NDVI (Normalized Difference Vegetation Index). These were used as the inputs for estimating

percent tree cover. Metrics such as maximum annual NDVI or mean growing season red reflectance represent generic signatures which can be used to map global vegetation. The approach to deriving the metrics and training data is fully described in Hansen et al. (2002).

Only the red and near infrared MODIS bands are close to the bandwidths in the long-term AVHRR sensor's record. In addition to these two bands, the AVHRR has one middle infrared and two thermal bands which record brightness temperatures and proved invaluable to mapping global land cover (Hansen et al. 2000). While the MODIS sensor has bands for measuring surface temperature (band 31 (10780-11280 nm) and band 32 (11770-12270 nm)), they are not currently processed for use in land cover mapping. There is a surface temperature product (Wan et al. 2002) which employs these bands, but its algorithms are land cover dependent, precluding its use in mapping surface cover. Bands 31 and 32 of the MODIS instrument are used to derive surface temperature. They mimic the AVHRR thermal bands and their inclusion in future reprocessing of the land products is recommended. If this is done, these bands will be used in future mapping efforts. In place of the missing MODIS thermal data, which act as a key regional stratification signal (Hansen et al. 2002), other features were included. First, a three region layer was included as a metric: extra tropical north (approximately 23 degrees north and above), tropical (approximately between plus and minus 23 degrees latitude), and extra tropical south (approximately 23 degrees south and below). Second, an archival 1km AVHRR channel 4 brightness temperature (10300 – 11300 nm) signal was used in metric form. These data are from 1995-96 and represent the most recent globally

processed thermal images at 1km for the AVHRR. The thermal information is used within the algorithm to regionally stratify the globe, as previously stated, and should not be affected by small-scale land cover change events since the time of the data's acquisition. The AVHRR metrics and regional layers are used alongside the MODIS data inputs.

The training data are derived by aggregating over 250 classified high-resolution Landsat images to the MODIS grid. The Landsat images were classified into four classes of tree cover, each class having a mean percent tree cover label. By averaging the Landsat tree cover strata to MODIS cells, a 500 meter continuous training data set was created. This training data set contains over a million pixels which were systematically sampled to create a final training data set of 271 149 pixels at the 500 meter MODIS resolution.

Methods

The MODIS continuous fields of vegetation cover algorithm is described in Hansen et al. (2002). It is an automated procedure which employs a regression tree algorithm (Venables and Ripley 1994). The regression tree is a non-linear, distribution-free algorithm which is highly suited for handling the complexity of global spectral land cover signatures. The training data are used as the dependent variable, predicted by the independent variables in the form of the annual MODIS metrics. Outputs from the regression tree are further modified by a stepwise regression and bias adjustment per Hansen et al. (2002). The derivation of tree cover in this way creates the possibility of

using subsequent depictions to measure change. Hansen et al. (2003) used such an approach in detecting change based on the long-term AVHRR 8km Pathfinder data set.

The output of the algorithm is percent canopy cover per 500 meter MODIS pixel. Here percent canopy refers to the amount of skylight obstructed by tree canopies equal to or greater than 5 meters in height and is different than percent crown cover (crown cover = canopy cover + within crown skylight). The canopy cover definition is used in vegetation modeling exercises where light availability is an important parameter. Foresters, on the other hand, largely employ crown cover in measuring forest density. Crown cover is a better measure when performing areal inventories and is the variable used in many forest accounting procedures.

To better understand this relationship, ongoing field work is being performed where both crown and canopy cover values are measured. Initial work suggests that the mean forest label used in deriving canopy cover (80 percent) training data corresponds to a 100 percent forested area in terms of crown cover. Figure 1 shows field data gathered from 4 different sites to test this assumption and shows a reasonable relationship. Of course, different tree types have different relationships between canopy and crown cover. Fir trees, for example, generally have little light availability within the canopy. Four subalpine fir sites from Colorado reveal a 0.9 ratio of canopy to crown cover. Broadleaf Kalahari woodland trees in Western Zambia, on the other hand, have a greater presence of within crown gaps and a 0.76 ratio. Information on stand species is not available at the global scale, however the 0.8 slope in Figure 1 suggests a reasonable estimate for

converting between canopy and crown cover. It is suggested that users interested in deriving the crown cover variable should divide the canopy cover layer by 0.8.

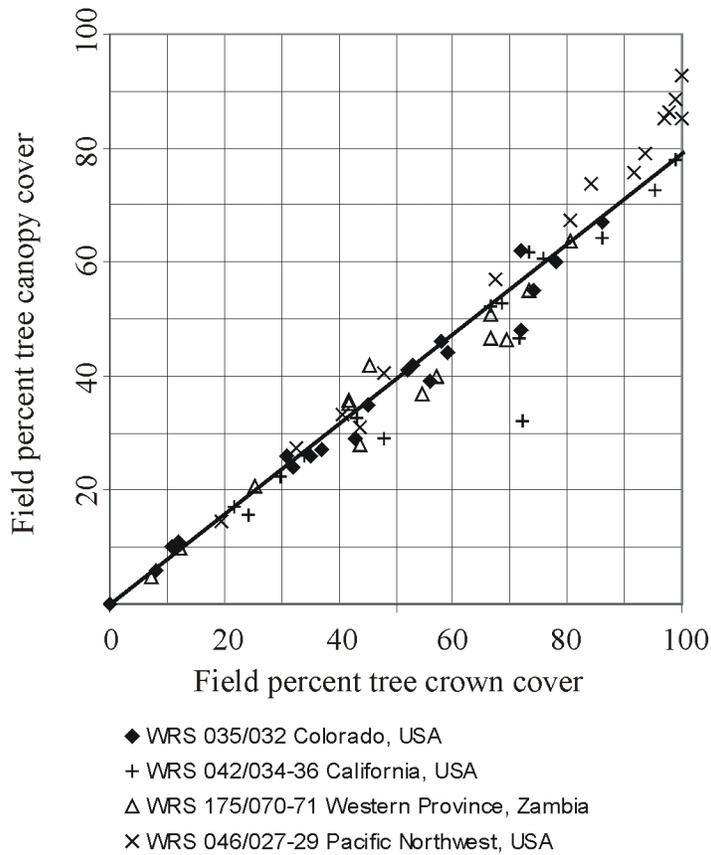


Figure 1. Plot of validation field data from four test areas, each covering a wide range of tree cover density, where $y=0.79x$ and $R^2 = 0.95$. See Hansen et al. (2002a) for example from Zambia test area.

Results

The resulting regression tree yielded 109 terminal nodes. The largest node in terms of surface area maps most of the tropical broadleaf evergreen forest and accounts for over 20 percent of dense (>40%) tree cover globally. This is a fairly homogeneous cover type with a characteristic signature. The regression tree delineated two subclasses of this

forest type which represent more confused spectral signatures: persistently cloudy areas, and areas of regrowth/disturbance. The next largest node in terms of tree cover maps dense needleleaf boreal forest. The final map is shown in Figure 2.

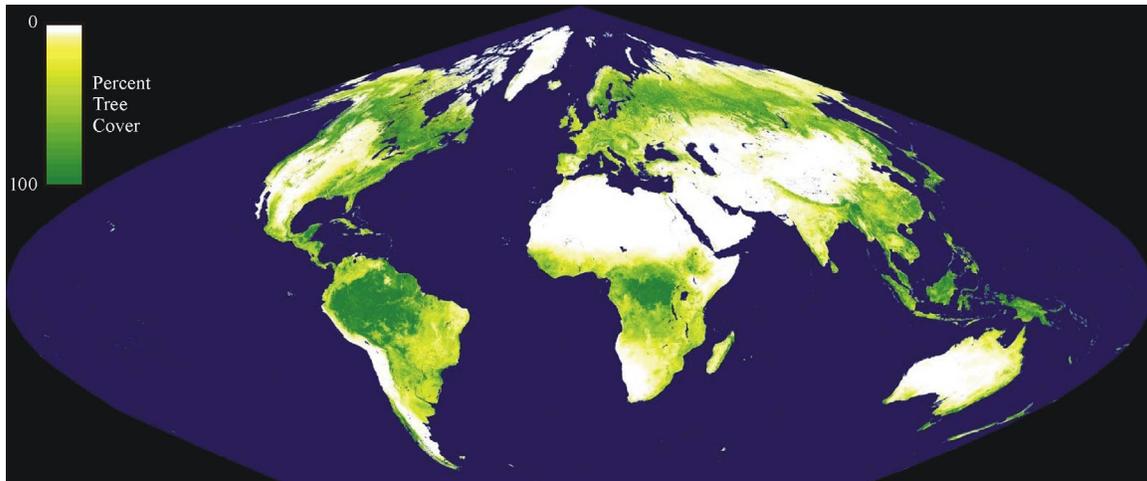


Figure 2. Final percent tree cover map in the MODIS Integerized Sinusoidal map projection.

The regression tree object can be studied to reveal which spectral information drives the tree cover characterization. Table 1 shows which metrics add most to reducing the overall sum of squares in delineating tree cover strata. From Table 1, it can be seen that the mean red reflectance corresponding to the three greenest composite periods metric contributes most to mapping tree cover. Of the overall reduction in the sum of squares, this metric alone contributes nearly 70 percent of the reduction. The first split in the regression tree uses this metric and this single split accounts for 60 percent of the reduction in the sum of squares. This metric is plotted against the resulting estimated tree cover for a 1000 pixel sample in Figure 3. While this metric alone cannot map global tree cover, it is clear that increasing canopy density is correlated with lower red

reflectance values due to the combined effects of canopy shadowing and chlorophyll absorption.

Table 1. Percent contribution to the overall reduction of sum of squares in the regression tree structure aggregated in the following ways: a) regression tree splits aggregated by the band used in the metrics with total number of splits in parentheses, b) splits aggregated by individual metric with total number of splits in parentheses and only the best ten metrics shown, c) metrics used in the 10 best individual splits in reducing overall sum of squares. These splits are highlighted in the tree structure of Figure 4. For mean metrics the span of time is listed by number of 40-day composites. For MODIS bands 1-7, metrics are derived by looking at dark albedo values. For example, *mean 1-3 band 6* represents the mean of the 3 darkest band 6 composite values. NDVI and temperature means are based on finding the maximum ranked composites. For example, *mean 1-5 NDVI* is the 5 highest NDVI composites averaged. The only exceptions to this are the metrics binned using NDVI as a reference. For example, in *mean 1-3g band 1*, the *g* indicates that the band 1 values are found which correspond to the 3 greenest composites based on ranked NDVI values. Thus, *mean 1-3g band 1* is a mean red reflectance value which corresponds to the 3 greenest composites.

Percent reduction in overall sum of squares per metrics aggregated by band		Percent reduction in overall sum of squares per individual metric		Percent reduction in overall sum of squares per regression tree split	
Band 1 metrics (9)	68.1	Mean1-3g band 1 (4)	67.7	Mean 1-3g band 1	60.1
Band 3 metrics (21)	9.6	Mean1-3g band 3 (2)	6.7	Mean 1-3g band 1	7.3
Band 4 metrics (20)	5.4	Mean 1-3 temperature (6)	4.4	Mean 1-3g band 3	6.7
Temperature metrics (13)	5.1	Mean 1-3 band 6 (3)	3.6	Mean 1-3 band 6	3.3
NDVI metrics (14)	4.5	Rank 3 band 3 (16)	2.7	Mean 1-3 temperature	2.4
Band 6 metrics (5)	3.8	Mean 1-3 band 4 (5)	2.7	Mean 1-3 temperature	1.8
Band 5 metrics (10)	1.3	Mean 1-8 NDVI (3)	2.2	Mean 1-3 band 4	1.4
Regions (7)	1.0	Mean 1-8 band 4 (5)	1.6	Mean 1-8 NDVI	1.2
Band 2 metrics (5)	1.0	Mean 1-5 NDVI (6)	1.3	Mean 1-5 NDVI	1.1
Band 7 metrics (4)	0.2	Regions (7)	1.0	Mean 1-8 NDVI	0.9

Of interest is the fact that the visible MODIS bands (red, green and blue) all contribute significantly while the near and middle infrared bands largely do not. Only band 6 performs comparably with the visible bands, and is critical to mapping inundated grasslands. Figure 4 shows the top levels of the regression tree and how band 6 is used to map these grasslands. While the infrared bands do not feature prominently, NDVI, derived using the near-infrared, is useful as seen in Table 1. It should also be noted that

many of the most used metrics of the visible bands are binned using NDVI to identify greenest times of the year.

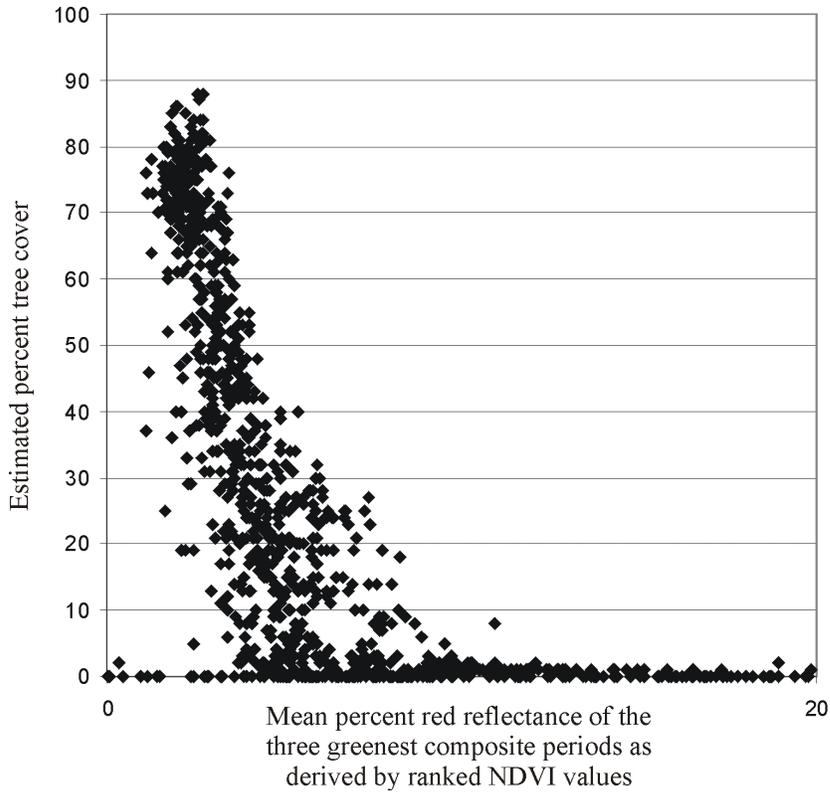


Figure 3. Plot of band 1 metric (mean red reflectance of 3 greenest compositing periods) versus estimate tree cover for a sample of 1000 pixels.

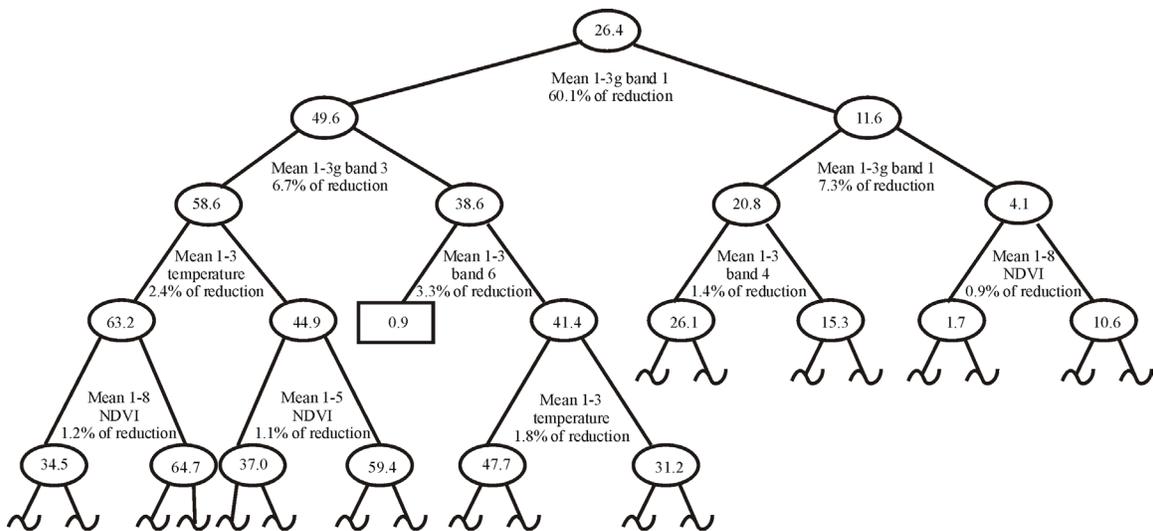


Figure 4. Regression tree object before stepwise regression and bias adjustment steps. Estimated percent tree cover at the intermediate nodes is shown within the ellipses. The metric used at each split is shown below each of these nodes along with corresponding percent of reduction in the overall sum of squares. Only one terminal node is shown as a rectangle, ~ indicates more splits in the lower portion of the regression tree.

The thermal signal of the AVHRR was used repeatedly, as seen in Table 1, and underscores the need to include the MODIS thermal signal in the gridded land products. The regional stratification was not as useful, only accounting for 1 percent of the overall reduction in the sum of squares. Two kinds of metrics were of little use: amplitude metrics for measuring the absolute spectral change of cover through the growing season and metrics associated with single peak greenness dates were largely unused by the regression tree.

The map has greatly increased spatial detail as compared to AVHRR-derived maps. Figure 5 shows two areas as examples. The human imprint on the landscape is more readily seen as compared to the AVHRR example. There is the reasonable expectation that consecutive comparisons of annual maps should reveal change. Discrete breaks in

tree cover due to administrative status, such as national park, government owned lands, and trans-national variations in land use intensity are clearly evident throughout the map. Figure 6 shows a region of southeastern Africa where differential land use intensification is visible across national boundaries. The rich detail present should be of use to land managers working at a regional scale and in need of an internally consistent map. Fire history is present as well, particularly in the boreal zone as shown in Figure 7 by the number of quasi-elliptical patterns which correspond to known fire scars. Further analysis of these data should reveal if this kind of map can be used to determine likely succession patterns, especially when other vegetation continuous field layers, such as leaf type are generated.

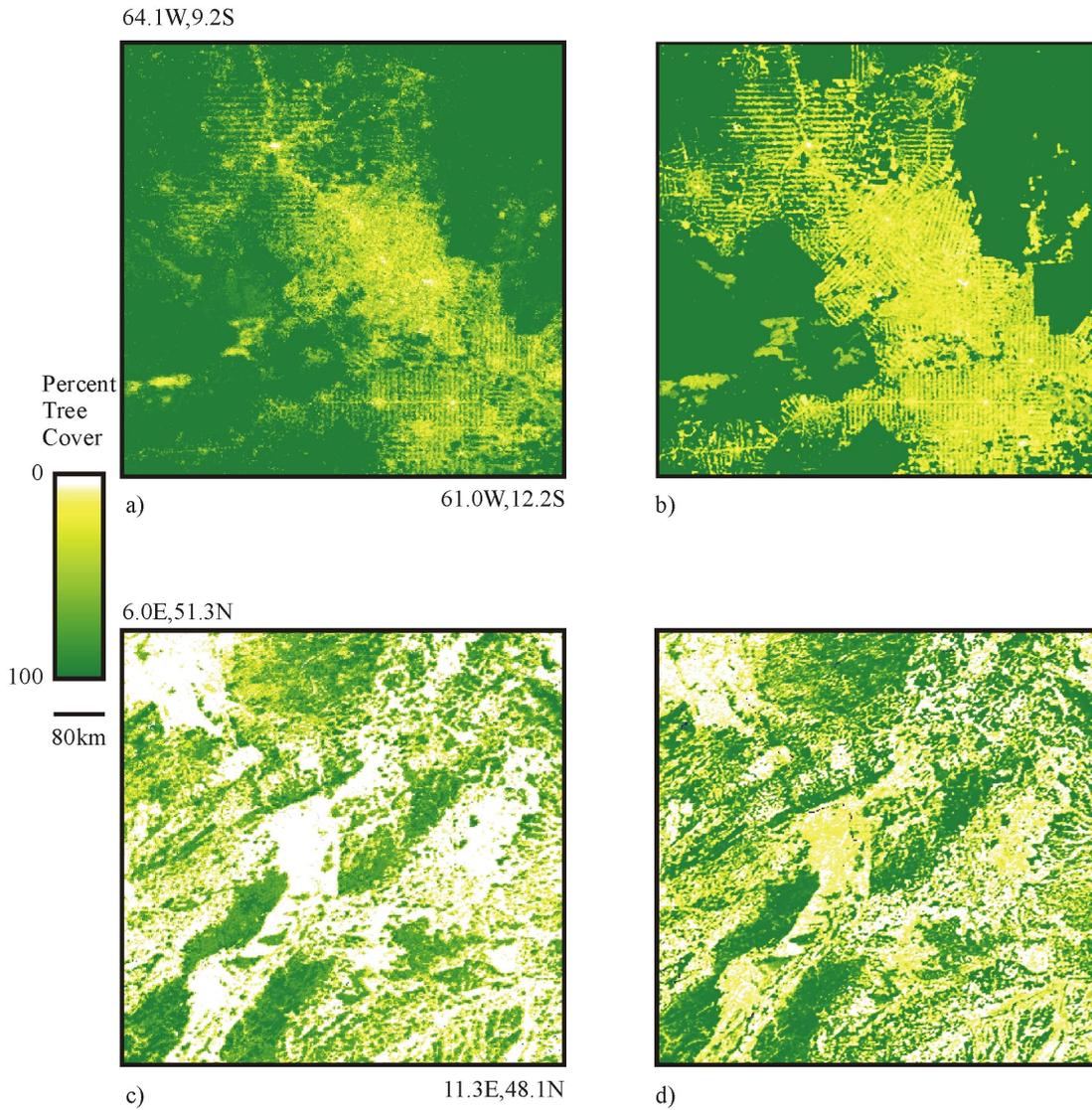


Figure 5. Two subsets comparing the 2000-01 500 meter MODIS tree cover map with a 1995-96 1km AVHRR tree cover map. a) is an area of Rondonia, Brazil from the AVHRR map, b) is the same area from the 2000-01 MODIS map, c) is an area along the French-German border from the AVHRR map, and d) the same area from the MODIS map.

26.0E,10.4S

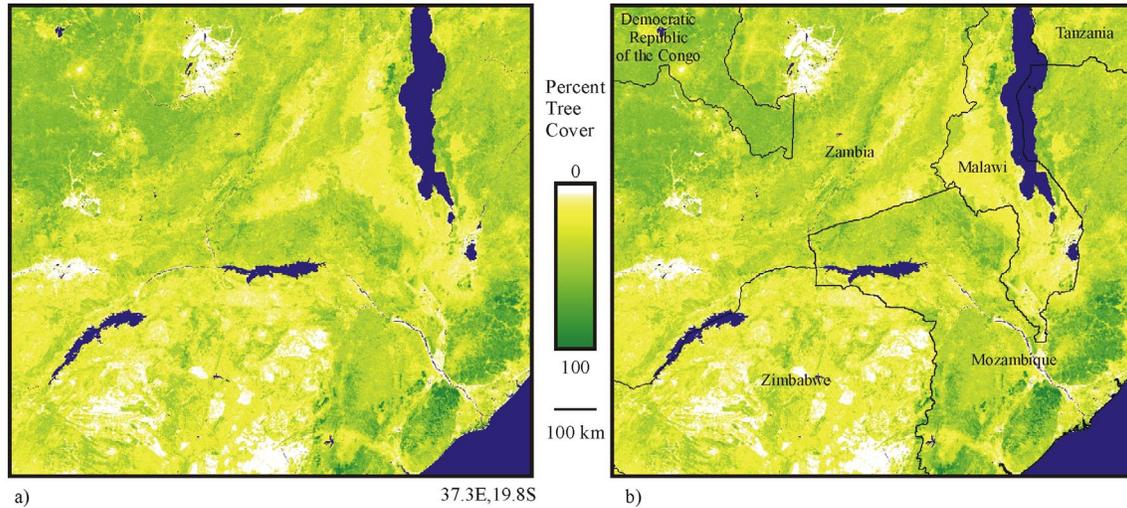


Figure 6. Transnational boundary differences in percent tree cover. The highest population density in this subset is found in Malawi, which is shown to have greater clearing of tree cover than adjacent countries. Mozambique is less disturbed as evidenced by Tete province jutting into the more intensively used landscapes of Zimbabwe, Zambia and Malawi. The arm of Congo extending into Zambia's copper belt is less disturbed than the heavily developed lands across the border.

The map will be updated regularly and used to monitor change in global tree cover.

Figure 8 shows the MODIS percent tree cover map with an overlay of a change study using AVHRR data for 1982-1999 (Hansen et al. 2003). MODIS data from the 250 and 500 meter bands should capture change in forest cover more accurately. Optimum change study intervals, whether annual or 5 to 10 year epochs, will be sought.

Improvements to the methodology, such as the inclusion of MODIS thermal bands, will be implemented as soon as is feasible.

154.5W,68.8N

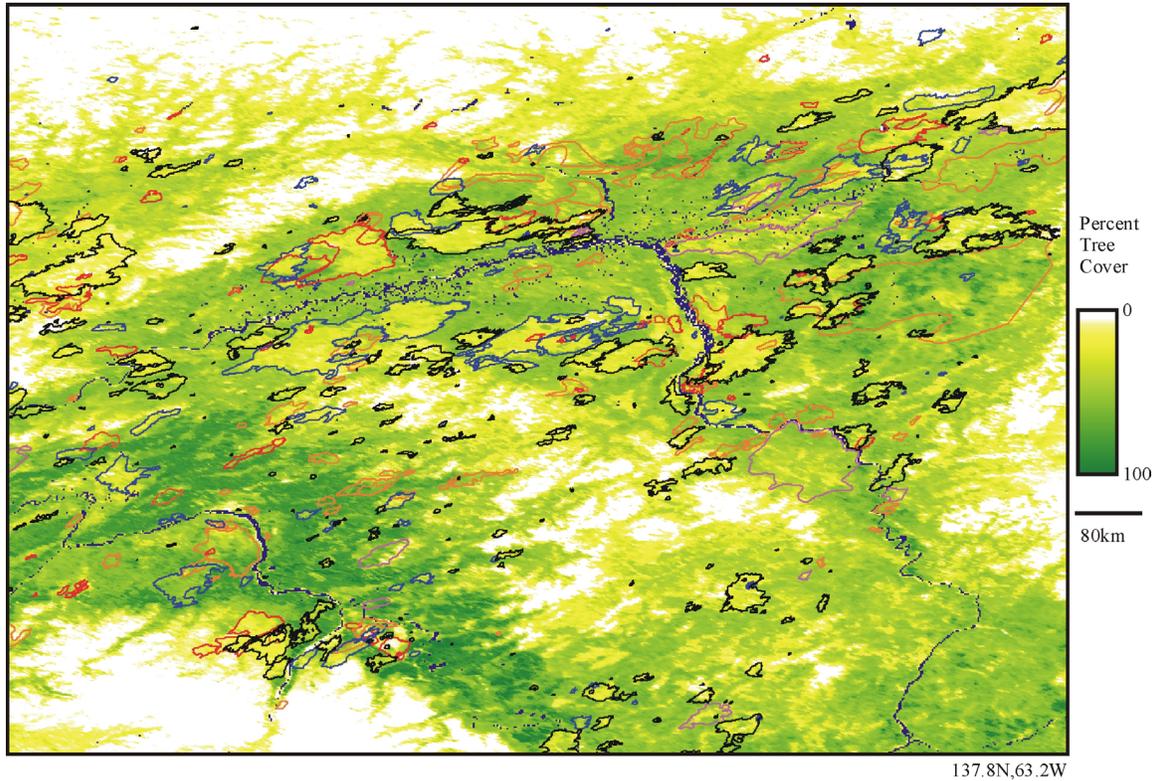


Figure 7. MODIS data with burn scar overlay. Black vectors represent burn scars from 1990-2000, blue 1980-1989, red, 1970-1979, magenta 1960-1969 and orange 1950-1959. More recent scars are fairly well delineated in the percent tree cover map. Data is from Murphey et al. (2000).

Validation

A multi-resolution mapping approach in conjunction with field data is being used at a number of sites to develop validation data for the percent tree cover map. The exercise includes using field data along with IKONOS and Enhanced Thematic Mapper Plus (ETM+) data to create validation test areas the size of an ETM+ image. Performing this exercise in a wide variety of biomes will help to create a test bed against which successive iterations of the tree cover product can be validated. The method has been initially tested for a Western Province, Zambia woodland site (Hansen et al. 2002) and is now being used in other areas.

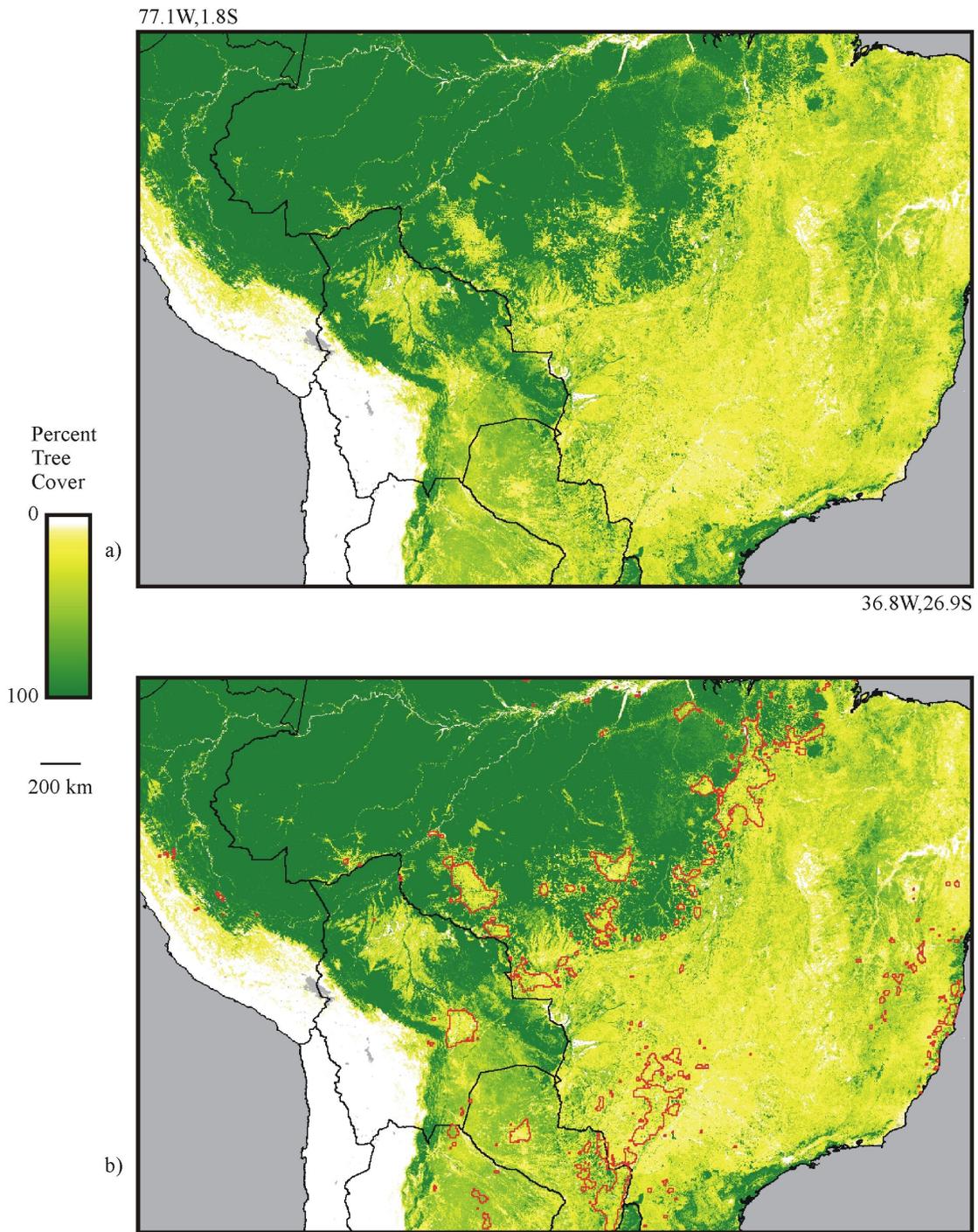


Figure 8. A portion of South America tree cover with deforestation hotspot overlay. The change areas are from a 19 year study of 8km AVHRR data (Hansen et al. 2003).

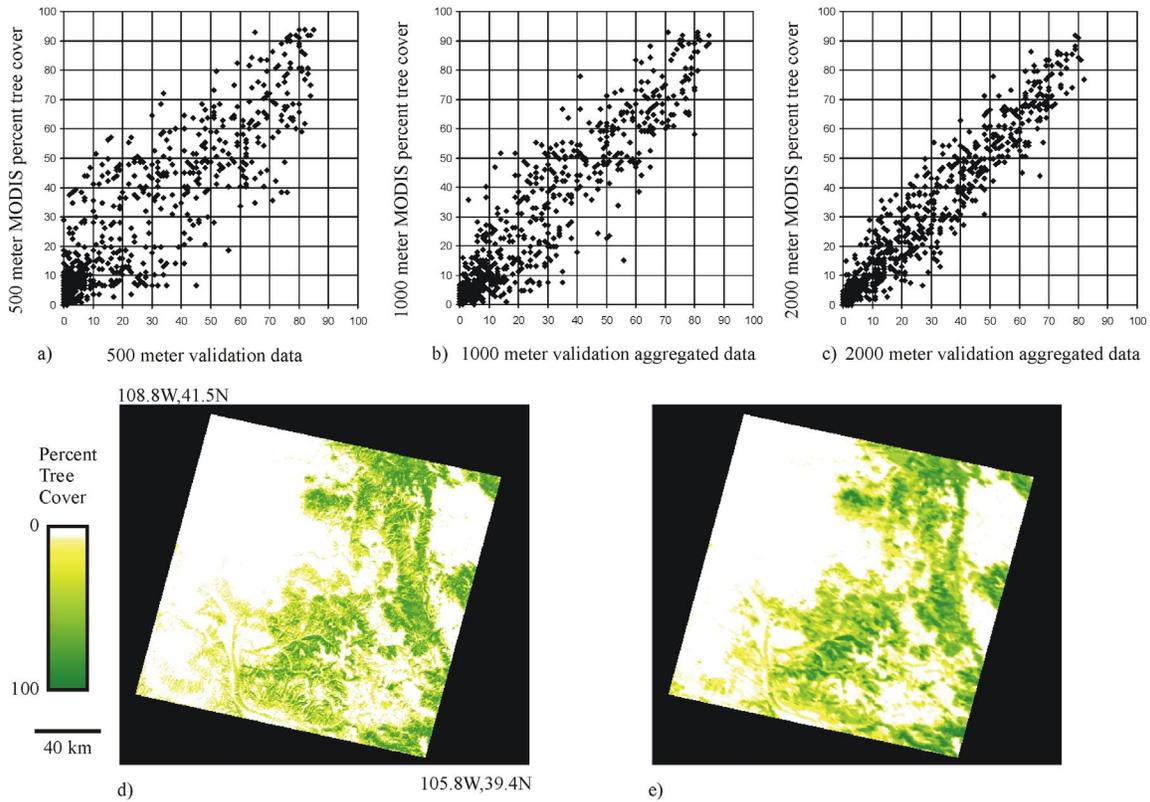


Figure 9. Validation data from Colorado, USA, WRS 035/032. a) Plot of 500 meter MODIS estimated percent tree cover versus 500 meter validation data derived from IKONOS/Enhanced Thematic Mapper Plus/field data study, $y=0.99$, $R^2=0.81$. b) Data averaged to a 1km meter resolution, $y=1.05x$, $R^2=0.89$. c) Data averaged to a 2km spatial resolution, $y=1.06x$, $R^2=0.94$. d) Validation data percent tree cover. e) 500 meter MODIS estimated percent tree cover.

Figure 9 shows results from a Colorado, USA test area. Averaging the product dramatically improves the validation measures. The greater scatter at 500 meter spatial resolution is probably an artifact of resampling in the MODIS data. As with all global data processing, a nearest neighbor scheme is used to reduce processing time. This approach leads to a geometric degradation of the signal as the process is repeated throughout the compositing process. Averaging the product to a 1km spatial resolution appears to ameliorate some of these effects.

Conclusion

The first layer of the MODIS Vegetation Continuous Field product, percent tree canopy cover, has been generated and is available for use (see below). The map reveals the improved spatial/spectral characteristics in the MODIS data compared to heritage AVHRR data. This should lead to a wider variety of applications which employ the MODIS-derived maps. Visible bands in the MODIS data provided the most discrimination along with NDVI and AVHRR brightness temperatures. This points out the need to add thermal information to the MODIS land data stream. Currently in production are other vegetation layers, including percent herbaceous/shrub, bare ground and tree leaf type and leaf longevity. Upon completion, these maps should enhance the current understanding of global land cover distributions and provide a basis for monitoring land cover change globally.

Data availability

The MODIS product in Integerized Sinusoidal tile format for canopy cover is available from the EROS Data Center at <http://edcimswww.cr.usgs.gov/pub/imswelcome> and per continent in the Pathfinder Goode's projection at <http://glcf.umiacs.umd.edu/>.

Appendix 1. Data structure of ISIN tile format

The MOD44B version available from the EROS Data Center is delivered as set of HDF-EOS files divided into the standard MODIS tiles in the Integerized Sinusoidal projection. Each tile covers a spatial extent of 10 degrees of latitude by 10 degrees of longitude at the

equator. This is an equal area projection and each tile contains 2400 samples by 2400 rows. Data quantization is 8-bit with the following valid values:

0-100 = percent tree canopy cover values
255 = interrupted space
254 = unprocessed data
253 = water
252 = land fill values in unprocessed data (0 percent tree cover)
251 = bad data in processed land area (0 percent tree cover)

References

Bonan, G. B., Levis, S., Kergoat, L., Oleson, K. W., 2002, Landscapes as patches of plant functional types: An intergrating concept for climate and ecosystem models, *Global Biogeochemical Cycles*, 16, 1360-1384.

DeFries, R. S., Houghton, R. A., Hansen, M. C., Field, C. B., Skole, D. and Townshend, J., 2002, Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980's and 1990's, *Proceedings of the National Academy of Sciences*, 99, 14256-14261.

DeFries, R. S., Hansen, M. C., Townshend, J. R. G., and Janetos, A. C., 2000, A new global 1-km dataset for percentage tree cover derived from remote sensing. *Global Change Biology* 6, 247-254.

DeFries, R. S., Townshend, J. R. G., and Hansen, M. C., 1999, Continuous fields of vegetation characteristics at the global scale at 1-km resolution, *Journal of Geophysical Research*, 104, 16,911-16,923.

DeFries, R. S., et al., 1995, Mapping the land surface for global atmosphere-biosphere models: Towards continuous distributions of vegetation's functional properties, *Journal of Geophysical Research*, 100, 20,867-20,882.

Hansen, M. C., and DeFries, R. S., 2003, Detecting long term global forest change using continuous fields of tree cover maps from 8km AVHRR data for the years 1982-1999, submitted, *Ecosystems*.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Marufu, L., Sohlberg, R., 2002, Development of a MODIS tree cover validation data set for Western Province, Zambia, *Remote Sensing of Environment*, 83, 320-335.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R. A., Dimiceli, C., and Carroll, M., 2002, Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS data, *Remote Sensing of Environment*, 83, 303-319.

Hansen, M. C., DeFries, R.S., Townshend, J.R.G., and Sohlberg, R., 2000, Global land cover classification at 1km spatial resolution using a classification tree approach *International Journal of Remote Sensing* 21, 1331-1364.

Matthews, E., 2001, Understanding the FRA, World Resources Institute Forest Briefing No. 1, (World Resources Institute: Washington, D. C.)

Murphy, P. J., Mudd, J. P., Stocks, B. J., Kasischke, E. S., Barry, D., Alexander, M. E., and French, N. H. F., 2000, Historical fire records in the North American boreal forest, in E. Kasischke and B. Stocks, eds. *Fire Climate Change, and Carbon Cycling in the Boreal Forest*, (New York: Springer-Verlag), 274-288.

Venables, W. N. and Ripley, B. D., 1994, *Modern Applied Statistics with S-Plus*, (New York: Springer-Verlag).

Wan, Z., Zhang, Y., Zhang, Q., and Li, Z., 2002, Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data, *Remote Sensing of Environment*, 83, 163-180.

Zhu, Z. and Waller, E., 2001, Global forest cover mapping for the United Nations Food and Agriculture Organization Forest Resources Assessment 2000 Program, UNFAO, Rome, Italy.