MAPPING SUCCESSIONAL STAGES IN A WET TROPICAL FOREST USING LANDSAT ETM+ AND FOREST INVENTORY DATA

Fabio G. Gonçalves[1,2], Mikhail Yatskov[2], João Roberto dos Santos[3], Robert N. Treuhaft[1], Beverly E. Law[2]

[1] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA, {fabio.goncalves, robert.n.treuhaft}@jpl.nasa.gov

[2] Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA, {mikhail.yatskov, bev.law}@oregonstate.edu

[3] Instituto Nacional de Pesquisas Espaciais, Av. dos Astronautas, 1758, 12.227-010, São José dos Campos, SP, Brasil, jroberto@dsr.inpe.br

ABSTRACT

In this study, we test whether an existing classification technique based on the integration of Landsat ETM+ and forest inventory data enables detailed characterization of successional stages in a wet tropical forest site. The specific objectives were: (1) to map forest age classes across the La Selva Biological Station in Costa Rica; and (2) to quantify uncertainties in the proposed approach in relation to field data and existing vegetation maps. Although significant relationships between vegetation height entropy (a surrogate for forest age) and ETM+ data were detected, the classification scheme tested in this study was not suitable for characterizing spatial variation in age at La Selva, as evidenced by the error matrix and the low Kappa coefficient (12.9%). Factors affecting the performance of the classification at this particular study site include the smooth transition in vegetation structure between intermediate and advanced successional stages, and the low sensitivity of NDVI to variations in vertical structure at high biomass levels.

Keywords: Remote sensing, monitoring, NDVI, forest structure, entropy.

INTRODUCTION

Tropical forests contain about 50% of the Earth's plant biomass, although they represent only 17% of potential natural vegetation by area (MELILLO *et al.*, 1993). Land use and land cover (LULC) changes in these ecosystems can therefore significantly alter the atmospheric concentration of carbon dioxide (MALHI and GRACE, 2000).

Despite its importance, the role of tropical forests in the global carbon cycle is poorly understood (CLARK *et al.*, 2001). There is still considerable uncertainty in current estimates of carbon flux from LULC change processes, which reflects the lack of reliable data on the rates and extent of LULC changes.

Although ground-based surveys can provide detailed information for a particular location, this method is prohibitively expensive and time-consuming for large-scale analysis. As a result, considerable effort has been directed toward mapping forest cover change from remote sensing data.

In this context, our objective was to test whether an existing classification technique based on the integration of Landsat and forest inventory data enables detailed characterization of successional stages in a wet tropical forest site. The specific objectives of this work were: (1) to map forest age classes across the La Selva Biological Station in Costa Rica; and (2) to quantify uncertainty in the proposed approach in relation to field data and existing vegetation maps.

METHODS

Study site

La Selva is a wet tropical forest in Costa Rica, Central America, which has a wide variety of primary and secondary stands (Figure 1). The vegetation is characterized by the dominance of large trees and by a high abundance of woody lianas and epiphytes. Rainfall averages 4000 mm annually, with an average temperature of 26°C (for a detailed description, see MCDADE *et al.*, 1994).

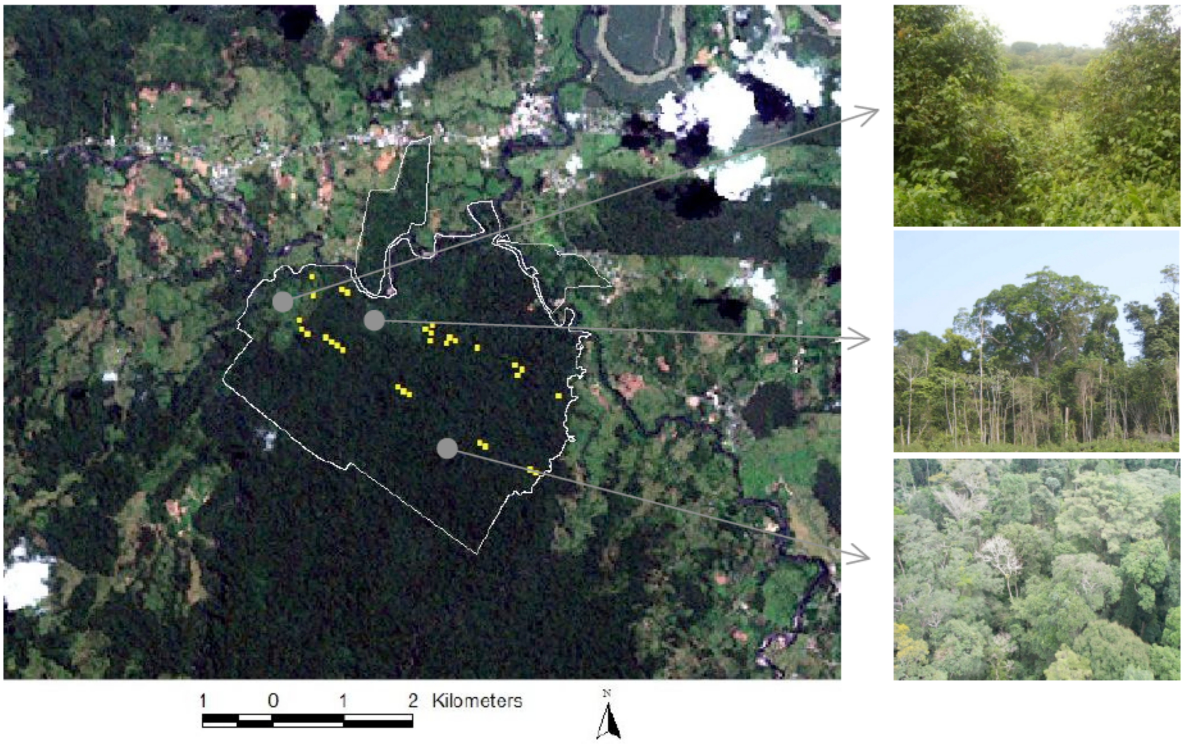

Figure 1. Landsat ETM+ image (R3-G2-B1 composite) of La Selva Biological Station in Costa Rica, outlined in white. The yellow dots are the locations of the 30 sample plots used in this work.

Data analysis

We used Landsat 7 Enhanced Thematic Mapper Plus (ETM+) data acquired over the study area in 2001, with average horizontal positioning error of less than one pixel (i.e. 30 m). Atmospheric correction was performed using the dark object subtraction technique (MORAN *et al.*, 1992), and the digital numbers were then converted to reflectance above the atmosphere using post-launch gains and offsets provided by the U.S. Geological Service (USGS).

In addition to the Landsat imagery, we used forest inventory data from thirty 0.1 ha stands placed in areas of both primary forest and secondary succession (TREUHAFT *et al.*, 2009), and a vegetation map updated from an IKONOS image acquired in 2000 (available at http://www.ots.ac.cr/).

The study area was stratified by age class based on the methodology of LU (2005), which consists of the following steps:

(1) Entropy (ENT) is calculated for each sample plot using tree height distribution derived from forest inventory data (p_i is the fraction of trees within height bin i);

$$ENT = -\sum_{i=1}^{j}\left[p_i \ln(p_i)\right]$$

(2) A model for estimating ENT from Landsat data is developed using linear regression analysis (potential explanatory variables: bands 1-5, band 7, and NDVI);

(3) This model is used to derive an ENT image for the entire study area;

(4) Successional stages are classified based on thresholds of ENT values derived from field data;

(5) The classification accuracy is evaluated using the error matrix approach (i.e., user's and producer's accuracy for each class, along with overall accuracy and the Kappa statistic).

RESULTS

The correlation coefficients between ENT and the Landsat data for the 30 plots were all significant with confidence greater than 95%, except for bands 4 and 5 (Table 1). All the significant correlations with individual bands were negative because ENT increases as the vegetation grows and becomes more complex, while the reflectance decreases. NDVI, on the other hand, is a spectral ratio that increases with vegetation density, showing therefore a positive correlation.

Table 1. Correlation coefficients between field ENT and Landsat 7 ETM+ data.

ETM+ data	Correlation w/ ENT
Band 1	-0.44*
Band 2	-0.59*
Band 3	-0.64*
Band 4	0.06
Band 5	-0.22
Band 7	-0.41*
NDVI	0.75*

(*) Marked correlations are significant at $p < 0.05$.

The final regression model (Figure 2), suggested by different variable selection techniques, included only NDVI and explained 57% of the variation in ENT. The scatterplot of residuals versus predicted values for this model (not shown here) suggested no evidence that the variance of the residuals increases with increasing values of NDVI. In addition, the normal probability plot indicated that the normality assumption is met.

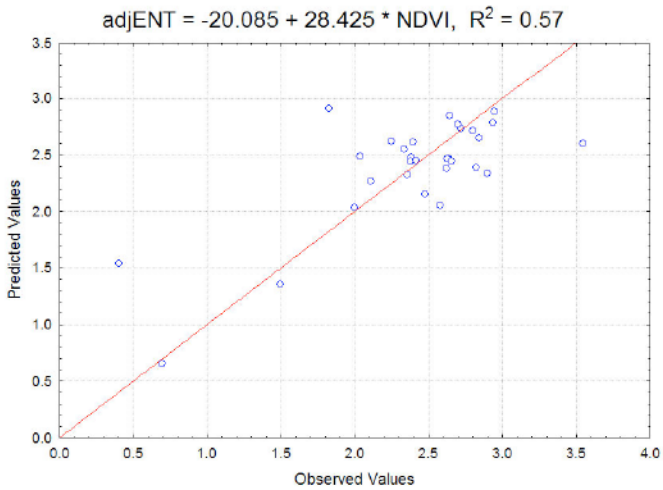

Figure 2. resulting regression model (field ENT as a function of NDVI), and scatterplot of predicted versus observed values.

Using this model, it was possible to predict ENT for the entire study area (Figure 3). Based on the thresholds of ENT derived from the field data (initial succession: <1.6, advanced: 1.6-2.4, intermediate: >2.4), the ENT image was color density sliced yielding the vegetation map shown in Figure 4a.

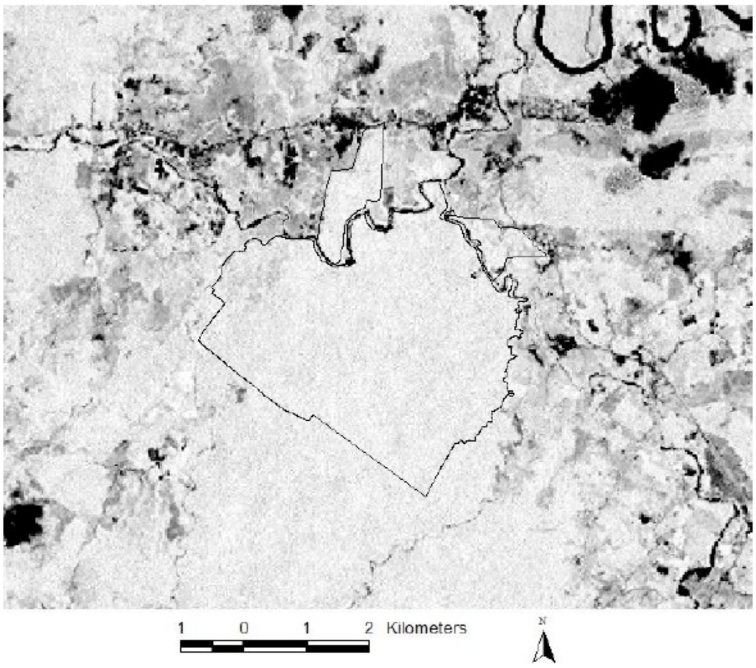

Figure 3. Estimated ENT image for La Selva Biological Station. The brighter the area, the higher the ENT.

The overall accuracy of the classification was 57.7%, closely matching the variability in ENT that is accounted for by the resulting regression model. Nevertheless, the Kappa coefficient, which is a statistical measure of agreement that takes into account the agreement occurring by chance, was 12.9%. The user's accuracy – the probability that a sample from the classification actually matches what it is from the reference data – for initial, intermediate, and advanced stages was 41.7%, 28.0%, and 60.8%, respectively, while the producer's accuracies – the probability that a reference sample will be correctly mapped – were 48.2%, 1.4%, and 88.4% (Table 2).

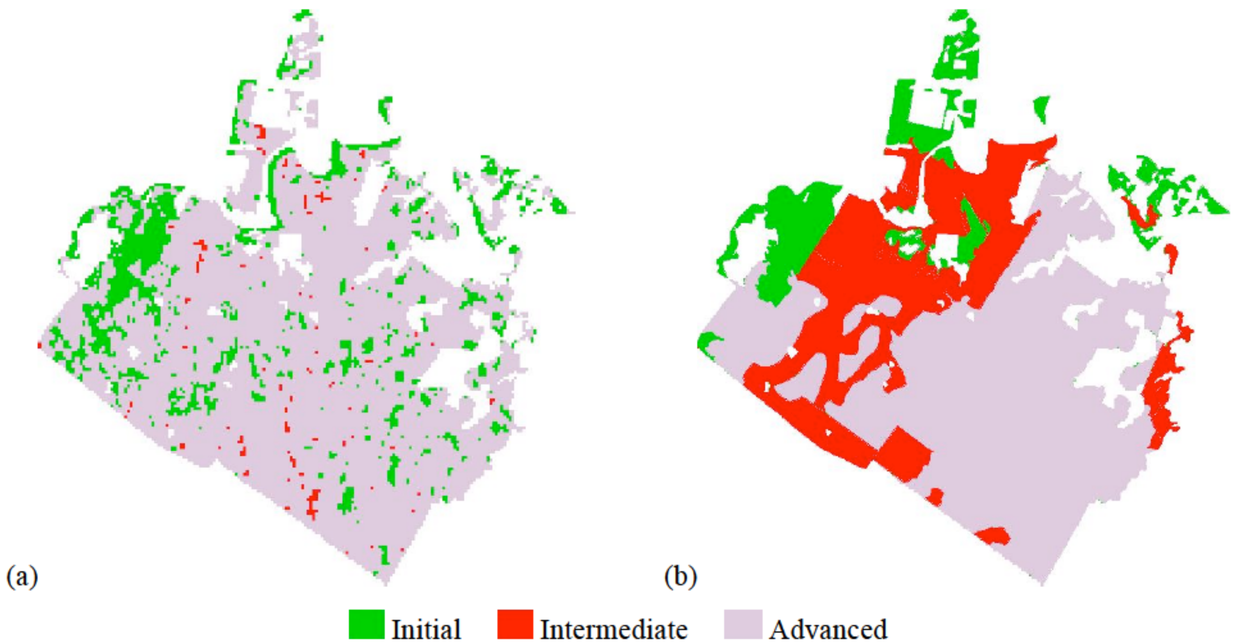

Figure 4. (a) Vegetation classification map derived from the analysis, and (b) reference map used to assess the classification accuracy.

Table 2. Error matrix summarizing the relationship between the classification and the reference map.

		Reference data (pixel counts)			
	Class	*Initial*	*Intermediate*	*Advanced*	Total
Classification (pixel counts)	*Initial*	96	38	96	230
	Intermediate	2	7	16	25
	Advanced	101	450	856	1407
	Total	199	495	968	1662

DISCUSSION

Although significant relationships between ETM+ data and entropy values were detected, the classification scheme tested in this study was not suitable for characterizing spatial variation in age at La Selva, as evidenced by the error matrix and the low Kappa coefficient. This is in contrast with a previous study conducted in the Brazilian Amazon (LU, 2005), which reported an overall accuracy of 80.4%. It should be noted that entropy values depend on the choice of vertical bin size, which might have been different between the two studies. Other methodological departures include the use of different age classes and the inclusion of NDVI as a potential explanatory variable.

One of the major factors affecting the performance of the classification at this particular study site was the smooth transition in vegetation structure between intermediate and advanced successional stages. In fact, the field-based entropy in advanced stands was on average lower than in intermediate stands, contrary to expectations. One possible explanation for this finding is that old growth stands in this ecosystem tend to be dominated by a single species (*Pentaclethra macroloba*), which would cause the vertical structure to become more homogeneous. The sensitivity of NDVI to variations in aboveground biomass is known to saturate at relatively low levels, which may have contributed to the low spectral separability observed between these two classes.

Finally, we note that the reference map itself may contain generalizations and errors, which were not taken into account in our analysis. Therefore, it is possible that part of the initial successional stands detected in our classification reflects real small scale disturbances such as treefalls that are not represented in the reference map. Future directions suggest use of 3-dimensional remote sensing techniques of InSAR and lidar for measurements of vertical structure of forests (e.g. TREUHAFT *et al.*, 2009), which have proven to be less prone to saturation effects and therefore more sensitive to spatial variation in age than NDVI.

ACKNOWLEDGEMENTS

F.G.G. gratefully acknowledges fellowship support from CAPES/Fulbright. The research described in this paper was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

CLARK, D.A. *et al.* Net primary production in tropical forests: an evaluation and synthesis of existing field data. *Ecological Applications*, v. 11, p. 371-384, 2001.

LU, D., Integration of vegetation inventory data and Landsat TM image for vegetation classification in the western Brazilian Amazon. *Forest Ecology and Management*, v. 213, p. 369-383, 2005.

MALHI, Y.; GRACE, J. Tropical forests and atmospheric carbon dioxide. *Trends in Ecology and Evolution*, v. 15, p. 332-337, 2000.

MCDADE L.A. *et al.* (eds.). *La Selva: ecology and natural history of a neotropical rain forest*. Chicago: The University of Chicago Press, 1994.

MELILLO, J.M. *et al.* Global climate change and terrestrial net primary production. *Nature*, v. 363, p. 234-240, 1993.

MORAN, M. *et al.* Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. *Remote Sensing of Environment*, v. 41, p. 169-184, 1992.

TREUHAFT, R.N. *et al.* Vegetation profiles in tropical forests from multibaseline interferometric synthetic aperture radar, field, and lidar measurements. *Journal of Geophysical Research*, v. 114, D 23110, 16 pp., 2009.