Relating LANDSAT ETM+ and forest inventory data for mapping successional stages in a tropical wet forest

Relacionando LANDSAT ETM+ e dados de inventário florestal para mapeamento de estádios sucessionais em uma floresta tropical úmida

Fabio G. Gonçalves[1]
Mikhail Yatskov[2]
João Roberto dos Santos[3]
Robert N. Treuhaft[4]
Beverly E. Law[5]

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 tropical wet 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 (0.129). Factors affecting the performance of the classification at this particular study site include the smooth transition in vegetation structure between intermediate and late successional stages, and the low sensitivity of NDVI to variations in vertical structure at high biomass levels.

Key words: remote sensing; monitoring; NDVI; forest structure; entropy.

1 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA; Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA; E-mail: fabio.goncalves@jpl.nasa.gov

2 Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA; E-mail: mikhail.yatskov@oregonstate.edu

3 Instituto Nacional de Pesquisas Espaciais, Avenida dos Astronautas, 1758, 12.227-010, São José dos Campos, São Paulo, Brasil; E-mail: jroberto@dsr.inpe.br

4 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA; E-mail: robert.n.treuhaft@jpl.nasa.gov

5 Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA; E-mail: bev.law@oregonstate.edu

Recebido para publicação em 02/08/2010 e aceito em 10/08/2010

Resumo

Nesse estudo, testamos se uma técnica de classificação existente, baseada na integração de imagens LANDSAT ETM+ e dados de inventário florestal, permite a caracterização detalhada de estádios sucessionais em uma área de floresta tropical úmida. Os objetivos específicos foram: (1) mapear classes de idade florestal na Estação Biológica de La Selva, na Costa Rica; e (2) quantificar as incertezas da abordagem proposta em relação aos dados de campo e mapas de vegetação existentes. Apesar de terem sido detectadas relações significativas entre dados ETM+ e medidas de entropia da altura da vegetação (um substituto para a idade florestal), o sistema de classificação testado nesse estudo não se demonstrou adequado para caracterizar a variação espacial em idade em La Selva, como evidenciado pela matriz de erro e o baixo coeficiente Kappa (0,129). Fatores que afetaram o desempenho da classificação nessa área de estudo em particular incluem a alta similaridade estrutural entre os estádios sucessionais intermediário e avançado, e a baixa sensibilidade do NDVI a variações na estrutura vertical em áreas com níveis elevados de biomassa.

Palavras-chave: sensoriamento remoto; monitoramento; NDVI; estrutura florestal; entropia.

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). Deforestation and forest degradation contribute ~17% of anthropogenic greenhouse gas emissions of carbon dioxide, making deforestation an obvious focus for monitoring to meet the United Nations Framework Convention on Climate Change (UNFCCC) requirements.

The annual change in forest area has an uncertainty of up to 100% in tropical forests. Uncertainties in emissions from deforestation and forest degradation are high for both annual values and trends, ranging from 25 to 100 percent, because of uncertainties in parameters used to translate deforested area into carbon dioxide emissions (NRC, 2010). Uncertainties in carbon sequestration are also high, in part due to the lack of reliable information on the status of forest regeneration, which can be a significant carbon sink depending on age and composition. Therefore, accurate discrimination of successional stages in tropical forests becomes critical for reducing uncertainties in terrestrial carbon fluxes.

Although ground-based surveys can provide detailed information on forest resources 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 (e.g. FOODY et al., 1996; LUCAS et al., 2000; KENNEDY et al., 2007).

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 tropical wet 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.

Materials and Methods

Study site

La Selva is a tropical wet forest in Costa Rica, Central America, which has a wide variety of primary (i.e. old-growth) and secondary (i.e. those regenerating after a major disturbance such as clear-cutting or fire) 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 4000mm annually, with an average temperature of 26°C (for a detailed description, see MCDADE et al., 1994).

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. 30m). 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 a vegetation map derived from visual classification of an IKONOS image acquired

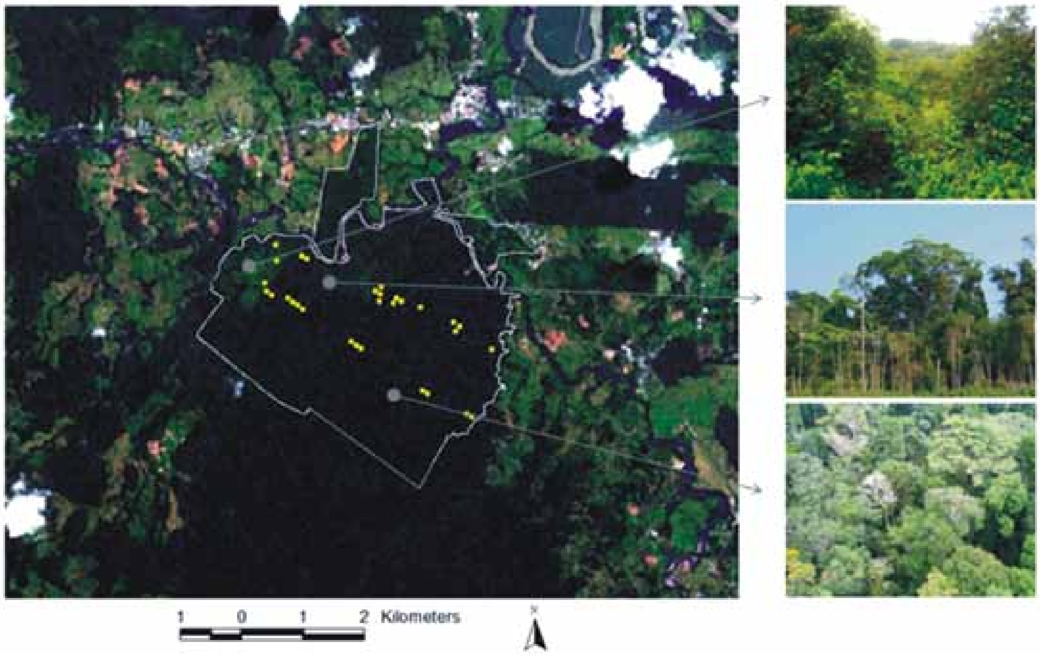

Figure 1. LANDSAT ETM+ image (R3-G2-B1 composite) of La Selva Biological Station in Costa Rica, outlined in white. The yellow dots are the locations of the 30 sample plots used in this work

in 2000 (available at http://www.ots.ac.cr/), and forest inventory data from thirty 0.1 ha stands measured in 2006 (TREUHAFT et al., 2009). Of those, 20 were placed in areas of primary forest (6 of which were selectively logged), and 10 in areas of secondary forest. The sites were chosen along two flight lines of the aircraft carrying the radar used for the measurements described by Treuhaft et al. (2009), potentially covering a wider range in vertical structure and aboveground biomass than one would expect from a simple random sample.

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

a) Entropy (ENT) is calculated for each sample plot using the tree height distribution derived from forest inventory data (Eq. 1: p_i is the fraction of trees within height bin i, and j is the bin corresponding to the tallest trees);

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

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

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

d) Successional stages are classified based on thresholds of ENT values derived from field data. For the purposes of this analysis, three classes were considered: early (secondary forests ≤ 18 years), intermediate (secondary forests between 18 and 40 years + selectively-logged forests), and late (primary forests) successional stages.

e) 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 the individual 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. Linear 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.

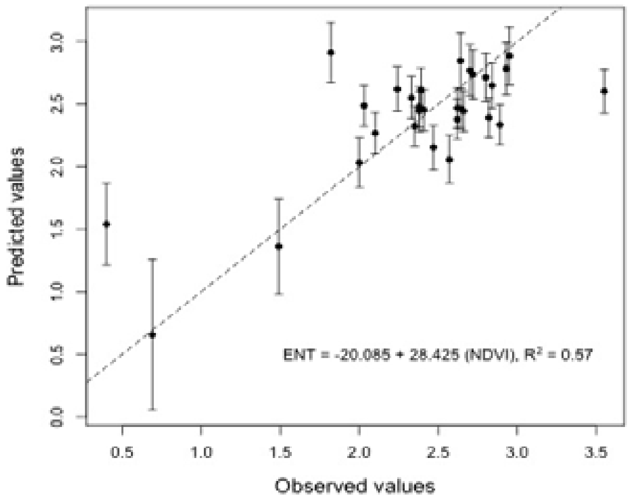

Figure 2. Scatterplot of predicted versus observed entropy values, and resulting regression model (field entropy as a function of NDVI). The vertical bars denote 95% confidence intervals of the regression, shown along with the 1:1 line

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 (early succession: <1.6, intermediate: >2.4, late: 1.6-2.4), the ENT image was color density sliced yielding the vegetation map shown in figure 4a.

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 0.129. Although this result suggests that the agreement between the classification and the reference map exceeds chance levels (i.e. Kappa > 0), Kappa values < 0.4 are generally not considered to be satisfactory (LANDIS and KOCH, 1977; FLEISS, 1981). The user's accuracy – the probability that a sample from the classification actually matches what it is from the reference data – for early, intermediate, and late 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).

Discussion and Conclusions

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.

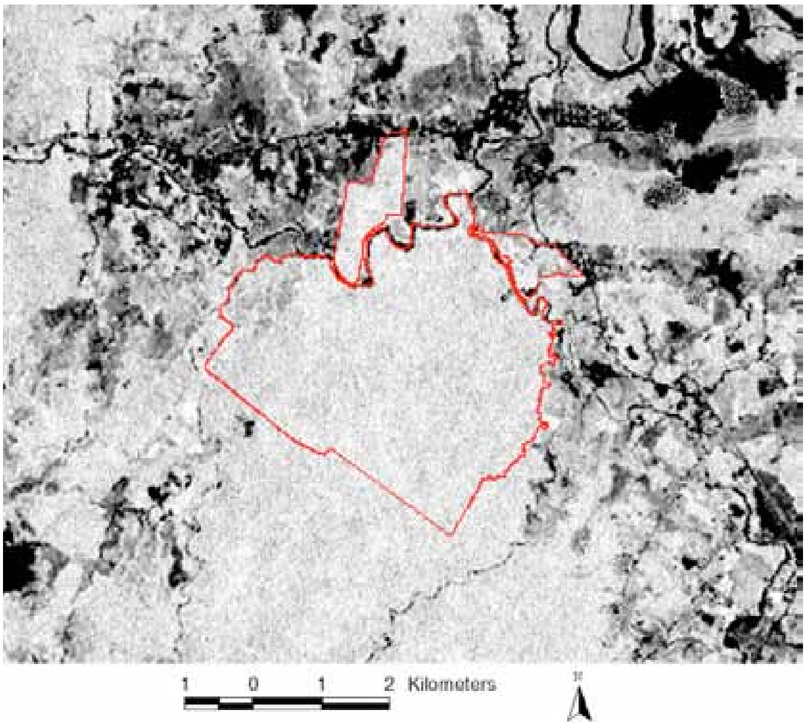

Figure 3. Estimated ENT image for La Selva Biological Station (outlined in red). The brighter the area, the higher the ENT

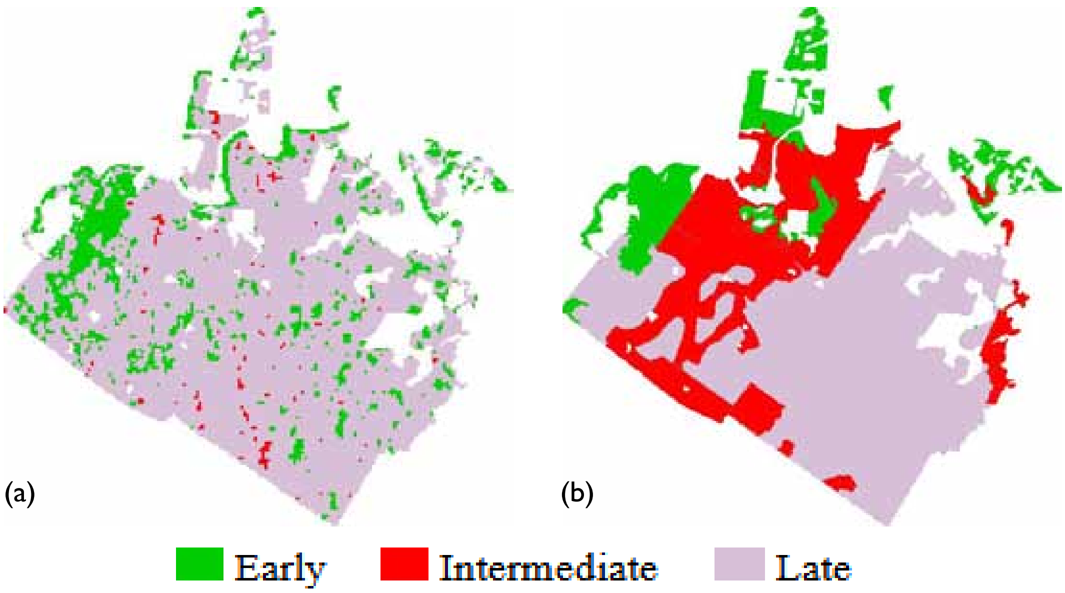

Figure 4. (a) Vegetation classification map derived from the analysis, and (b) reference map used to assess the classification accuracy

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

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

Other methodological departures include the use of different age classes and a smaller number of young stands in our analysis, 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 late successional stages. In fact, the field-based entropy in late 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. This is in contrast to moist tropical forests and temperate evergreen needleleaf forests with high biomass, which tend to develop more structural complexity in old forests. 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. NDVI has typically been used to quantify variation in leaf area index (LAI), where the sensitivity also saturates at low LAI (e.g. LAW and WARING, 1994).

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 early successional stands detected in our classification reflects real small scale disturbances such as treefalls that are not represented in the reference map. We should also note that the data used in this study (field and Remote Sensing) were collected at spatial scales that varied by as much as two orders of magnitude, which might have introduced additional errors in the analysis.

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

FLEISS, J. L. **Statistical methods for rates and proportions**. 2 ed. New York: John Wiley, 1981.

FOODY, G. M.; PALUBINSKAS, G.; LUCAS, R. M.; CURRAN, P.; HONZAK, M. Identifying terrestrial carbon sinks: classification of successional stages in regenerating tropical forest from Landsat TM data. **Remote Sensing of Environment**, v. 55, n.3, p. 205-216, 1996.

KENNEDY, R. E.; COHEN, W. B.; SCHROEDER, T. A. Trajectory-based change detection for automated characterization of forest disturbance dynamics. **Remote Sensing of Environment**, v. 110, p. 370-386, 2007.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, p. 159-174, 1977.

LAW, B. E.; WARING, R. H. Remote sensing of leaf area index and radiation intercepted by understory vegetation. **Ecological Applications**, v. 4, p. 272-279, 1994.

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.

LUCAS, R. M.; HONZAK, M.; CURRAN, P. J.; FOODY, G. M.; MILNE, R.; BROWN, T.; AMARAL, S. Mapping the regional extent of tropical forest regeneration stages in the Brazilian Legal Amazon using NOAA AVHRR data. **International Journal of Remote Sensing**, v. 21, p. 2855-2881, 2000.

MCDADE, L.A.; BAWA, K. S.; HESPENHEIDE, H. A.; HARSTSHORN, G. S. (eds.). **La Selva**: ecology and natural history of a neotropical rain forest. Chicago: The University of Chicago Press, 1994.

MELILLO, J. M.; MCGUIRE, A. D.; KICKLIGHTER, D. W.; MOORE, B.; VOROSMARTY, C. J.; SCHLOSS, A. L. Global climate change and terrestrial net primary production. **Nature**, v. 363, p. 234-240, 1993.

MORAN, M. S.; JACKSON, R. D.; SLATER, P. N.; TEILLET, P. M. Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. **Remote Sensing of Environment**, v. 41, n. 2-3, p. 169-184, 1992.

NATIONAL RESEACH COUNCIL – NRC. Verifying greenhouse gas emissions: methods to support international climate agreements. Washington DC: National Academies Press, 132 pp., 2010.

TREUHAFT, R. N.; CHAPMAN, B. D.; SANTOS, J. R.; GONÇALVES, F. G.; DUTRA, L. V.; GRAÇA, P. M. L. A.; DRAKE, J. B. Vegetation profiles in tropical forests from multibaseline interferometric synthetic aperture radar, field, and lidar measurements. **Journal of Geophysical Research**, v. 114, D 23110, 16 p., 2009.