International Journal of Applied Earth Observation and Geoinformation 23 (2013) 372–383

Contents lists available at SciVerse ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Enhanced land use/cover classification of heterogeneous tropical landscapes using support vector machines and textural homogeneity

Jaime Paneque-Gálvez[a,b,*], Jean-François Mas[b], Gerard Moré[c], Jordi Cristóbal[d], Martí Orta-Martínez[a], Ana Catarina Luz[a], Maximilien Guèze[a], Manuel J. Macía[e], Victoria Reyes-García[f]

[a] Institut de Ciència i Tecnologia Ambientals (ICTA), Universitat Autònoma de Barcelona (UAB), 08193 Bellatera, Barcelona, Spain
[b] Centro de Investigaciones en Geografía Ambiental (CIGA), Universidad Nacional Autónoma de México (UNAM), Antigua Carretera a Pátzcuaro No 8701, Col. Ex-Hacienda de San José de La Huerta, 58190 Morelia, Michoacán, Mexico
[c] Centre de Recerca Ecològica i Aplicacions Forestals (CREAF), Universitat Autònoma de Barcelona (UAB), 08193 Bellatera, Barcelona, Spain
[d] Departament de Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Barcelona, Spain
[e] Departamento de Biología, Unidad de Botánica, Universidad Autónoma de Madrid (UAM), Calle Darwin 2, 28049 Madrid, Spain
[f] ICREA and Institut de Ciència i Tecnologia Ambientals (ICTA), Universitat Autònoma de Barcelona (UAB), 08193 Bellatera, Barcelona, Spain

ARTICLE INFO

Article history:
Received 28 November 2011
Accepted 15 October 2012

Keywords:
Remote sensing
Thematic classification comparison
SVM
k-Nearest neighbor
Hybrid classification
Texture
Bolivian Amazon

ABSTRACT

Land use/cover classification is a key research field in remote sensing and land change science as thematic maps derived from remotely sensed data have become the basis for analyzing many socio-ecological issues. However, land use/cover classification remains a difficult task and it is especially challenging in heterogeneous tropical landscapes where nonetheless such maps are of great importance. The present study aims at establishing an efficient classification approach to accurately map all broad land use/cover classes in a large, heterogeneous tropical area, as a basis for further studies (e.g., land use/cover change, deforestation and forest degradation). Specifically, we first compare the performance of parametric (maximum likelihood), non-parametric (k-nearest neighbor and four different support vector machines – SVM), and hybrid (unsupervised–supervised) classifiers, using hard and soft (fuzzy) accuracy assessments. We then assess, using the maximum likelihood algorithm, what textural indices from the gray-level co-occurrence matrix lead to greater classification improvements at the spatial resolution of Landsat imagery (30 m), and rank them accordingly. Finally, we use the textural index that provides the most accurate classification results to evaluate whether its usefulness varies significantly with the classifier used. We classified imagery corresponding to dry and wet seasons and found that SVM classifiers outperformed all the rest. We also found that the use of some textural indices, but particularly homogeneity and entropy, can significantly improve classifications. We focused on the use of the homogeneity index, which has so far been neglected in land use/cover classification efforts, and found that this index along with reflectance bands significantly increased the overall accuracy of all the classifiers, but particularly of SVM. We observed that improvements in producer's and user's accuracies through the inclusion of homogeneity were different depending on land use/cover classes. Early-growth/degraded forests, pastures, grasslands and savanna were the classes most improved, especially with the SVM radial basis function and SVM sigmoid classifiers, though with both classifiers all land use/cover classes were mapped with producer's and user's accuracies of ~90%. Our classification approach seems very well suited to accurately map land use/cover of heterogeneous landscapes, thus having great potential to contribute to climate change mitigation schemes, conservation initiatives, and the design of management plans and rural development policies.

1. Introduction

Accurate land use/cover (LUC) maps derived from remotely sensed data form the basis for quantifying and monitoring the spatio-temporal patterns of LUC change. In most tropical regions, LUC change is taking place at unprecedented rates (Gibbs et al., 2010) and, therefore, accurate LUC maps are key for assessing its implications for climate change, biodiversity conservation, and peoples' livelihoods. Nevertheless, LUC classification remains a

* Corresponding author at: Institut de Ciència i Tecnologia Ambientals (ICTA), Universitat Autònoma de Barcelona (UAB), 08193 Bellatera, Barcelona, Spain. Tel.: +34 93 586 85 48; fax: +34 93 586 80 08.
E-mail address: jpanequegalvez@gmail.com (J. Paneque-Gálvez).

0303-2434/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jag.2012.10.007

challenging task in heterogeneous tropical areas for several reasons. A major problem lies in the difficulty of acquiring cloud-free multispectral imagery, which may be partly overcome through the use of radar imagery (Freitas et al., 2008), but their interpretation is not straightforward in tropical areas (Almeida-Filho et al., 2007). Another major drawback is related to the limitations (in terms of cost, time, and accessibility) for carrying out fieldwork to collect sufficient information on LUC classes, which hampers the training and validation stages of supervised and hybrid LUC classification approaches. Other constraints for accurate LUC mapping in tropical regions are the usual lack of aerial photography, of previous LUC maps, and of ancillary data (e.g., digital elevation models – DEMs, geological maps) that may be used to improve classification results.

To address the problems related to LUC classification in tropical areas, a fundamental issue is the selection of the classifier. The use of machine learning algorithms have gained momentum in recent years and some assessments of their relative performance compared to other classifiers have been conducted in the Amazon region (Lu et al., 2004; Carreiras et al., 2006). Specifically, support vector machines (SVM) have been shown to attain high accuracies in LUC mapping and outperform other algorithms (Huang et al., 2002; Foody and Mathur, 2004a; Pal and Mather, 2005; Kavzoglu and Colkesen, 2009; Mountrakis et al., 2011; Szuster et al., 2011). SVM have two significant advantages for LUC mapping. First, since SVM classifiers seek to separate LUC classes by finding a plane in the multidimensional feature space that maximizes their separation, rather than by characterizing such classes with statistics, they do not need a large training set but just the training samples that are support vectors (Foody and Mathur, 2004b). Thus, for SVM classifiers Foody and Mathur (2006) suggested to use small training sets composed of purposely selected mixed pixels containing the support vectors, as this approach does not compromise classification accuracies and may save considerable time. Second, SVM algorithms are independent of data dimensionality (Dixon and Candade, 2008), which is a key feature when using many spectral bands such in hyperspectral imagery or when ancillary data are included in the classification process; conversely, for classifiers that depend on dimensionality (e.g., artificial neural networks), training sets must exponentially increase in size to maintain classifier performance (Dixon and Candade, 2008).

Another fundamental issue to enhance LUC classification is the adequate selection of input variables, which some authors suggest may have the same impact as the selection of the classifier (Heinl et al., 2009). Nevertheless, we argue that the combination of an allegedly superior classifier such as SVM with appropriate ancillary data should improve results, as observed by Watanachaturaporn et al. (2008) using multisource classification with SVM. Different textural measures are a potential source of ancillary data and their benefits for LUC classification have been highlighted in studies using different techniques and classifiers (e.g., Berberoglu et al., 2000; Chica-Olmo and Abarca-Hernández, 2000). Specifically for forest classification, the inclusion of textural data has proven useful for mapping forest age, forest types, detecting forest cover change, and characterizing canopy structure (e.g., Palubinskas et al., 1995; Franklin et al., 2000; Zhang et al., 2004; Kayitakire et al., 2006; Malhi and Román-Cuesta, 2008), particularly with high spatial resolution imagery (Ota et al., 2011). A significant advantage of using texture to enhance image classification in tropical regions, where other ancillary data sources may not exist, is that textural data can be extracted from the image itself. Thus, for example, the gray-level co-occurrence matrix method can be used to extract textural indices that can be included as data bands in the classification process (Gong et al., 1992).

Given the current limitations for classifying LUC in complex tropical areas, in this paper our main goal is to establish a robust classification approach to accurately map all the broad LUC classes considered in a heterogeneous tropical area. Specifically, we first test if LUC classification results obtained with SVM classifiers compare favorably with other parametric, non-parametric, and hybrid classifiers; we then assess what textural indices from the gray-level co-occurrence matrix lead to greater classification improvements at the spatial resolution of Landsat imagery (30 m), and rank them accordingly; and finally, we evaluate if the usefulness of a textural index for LUC classification varies significantly in relation to the classifier utilized.

2. Study area, field surveys, and map legend definition

2.1. Study area

The study area is located in the department of Beni, Bolivia (Fig. 1). We selected this large area because its landscapes are highly heterogeneous as a transition across three biogeographic areas: (1) montane tropical forests covering the foothills (over 400 m) of the Andes to the west, (2) lowland tropical forests to the south and center of the study area, and (3) wet savanna areas to the north and east. Lowland forests are located below 400 m and contain some deciduous species owing to a marked seasonality (dry and wet seasons) (Guèze et al., in press). In wet savannas vegetation is controlled by small variations in ground elevation and relief, which in turn are shaped by river dynamics and periodic flooding. Savanna areas consist of swamps, marshes and lagoons with aquatic vegetation in the lowest areas; semi-natural grasslands and pastures in areas less prone to be flooded; and scrublands and patches of forests on mounds that do not get seasonally flooded. The vegetation formations of the study area are also shaped by the land use type and intensity of its different inhabitants, who range from Andean indigenous peoples in montane forests, to local peasants, cattle ranchers, and different native and colonist indigenous peoples in lowland forests and savanna areas.

2.2. Field surveys and map legend definition

Two field surveys were undertaken across the study area to collect LUC data. The first focused on forested areas (old-growth, early-growth, and degraded forests), water, bare soil, and infrastructure/urban categories, and was carried out in June–August 2009 (dry season). The second one took place in April–May 2010 (end of the wet season) and focused on the large savanna areas that are present across the study area, which mix with patches of pastures, semi-natural grasslands and scrublands. Planning the acquisition of ground-truth data was done upon preliminary analyses of the most recent Landsat-5 Thematic Mapper (TM) scenes (April 2009). LUC data were acquired with handheld GPS units, with typical mean positioning errors of 2–4 m in open areas and 4–6 m in forested areas. Additionally, to assist in the processes of geometric correction and geometric accuracy assessment, we collected GPS points at road crossings and other human-made features on the ground, and GPS tracks along the major roads and rivers across the study area.

The definition of broad LUC classes was carried out prior to the field surveys and was based on previous knowledge of the area and initial remotely sensed data exploration, which consisted in carrying out several unsupervised ISODATA classifications on the most recent Landsat imagery we had, and checking the classification obtained by Killeen et al. (2007) for our study area. Nevertheless, the definition of LUC classes was modified according to our field observations and thorough examination of the spectral signatures extracted from our field data. Eight broad LUC classes were finally considered (Table 1).