Remote Sensing of Environment 114 (2010) 2816–2832

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the Dry Chaco ecoregion of South America

Matthew L. Clark [a,*], T. Mitchell Aide [b], H. Ricardo Grau [c], George Riner [a]

[a] Center for Interdisciplinary Geospatial Analysis, Department of Geography and Global Studies, Sonoma State University, Rohnert Park, California, USA 94928
[b] Department of Biology, University of Puerto Rico, PO Box 23360, Río Piedras 00931-3360, Puerto Rico, USA
[c] CONICET, Instituto de Ecología Regional, Universidad Nacional de Tucumán, Casilla de Correo 34 (4107) Yerba Buena, Tucumán, Argentina

ARTICLE INFO

Article history:
Received 8 December 2009
Received in revised form 16 June 2010
Accepted 2 July 2010

Keywords:
Land cover and land use change
MODIS Enhanced Vegetation Index (EVI)
Time series analysis
Vegetation phenology
Random Forests
Google Earth interpretation
Dry Chaco ecoregion

ABSTRACT

Land use and land cover (LULC) maps from remote sensing are vital for monitoring, understanding and predicting the effects of complex human–nature interactions that span local, regional and global scales. We present a method to map annual LULC at a regional spatial scale with source data and processing techniques that permit scaling to broader spatial and temporal scales, while maintaining a consistent classification scheme and accuracy. Using the Dry Chaco ecoregion in Argentina, Bolivia and Paraguay as a test site, we derived a suite of predictor variables from 2001 to 2007 from the MODIS 250 m vegetation index product (MOD13Q1). These variables included: annual statistics of red, near infrared, and enhanced vegetation index (EVI), phenological metrics derived from EVI time series data, and slope and elevation. For reference data, we visually interpreted percent cover of eight classes at locations with high-resolution QuickBird imagery in Google Earth. An adjustable majority cover threshold was used to assign samples to a dominant class. When compared to field data, we found this imagery to have georeferencing error <5% the length of a MODIS pixel, while most class interpretation error was related to confusion between agriculture and herbaceous vegetation. We used the Random Forests classifier to identify the best sets of predictor variables and percent cover thresholds for discriminating our LULC classes. The best variable set included all predictor variables and a cover threshold of 80%. This optimal Random Forests was used to map LULC for each year between 2001 and 2007, followed by a per-pixel, 3-year temporal filter to remove disallowed LULC transitions. Our sequence of maps had an overall accuracy of 79.3%, producer accuracy from 51.4% (plantation) to 95.8% (woody vegetation), and user accuracy from 58.9% (herbaceous vegetation) to 100.0% (water). We attributed map class confusion to limited spectral information, sub-pixel spectral mixing, georeferencing error and human error in interpreting reference samples. We used our maps to assess woody vegetation change in the Dry Chaco from 2002 to 2006, which was characterized by rapid deforestation related to soybean and planted pasture expansion. This method can be easily applied to other regions or continents to produce spatially and temporally consistent information on annual LULC.

1. Introduction

Land use and land cover (LULC) maps are vital for monitoring, understanding and predicting the effects of complex human–nature interactions that span local, regional and global scales. For example, a spatial depiction of land conversion, such as deforestation for agriculture or pastures, or incremental changes, such as forest degradation and reforestation, are important for reducing uncertainty in carbon stocks and emissions, developing strategies for biodiversity protection, and understanding how globalization affects local and regional land use trends (Houghton, 2005; De Fries et al., 2007). Assessment of rapid land use changes, such as deforestation in the tropics (Archard et al., 2002), requires frequent measurements if it is to be incorporated into management and policy decisions. Furthermore, global issues span political and cultural boundaries, and so LULC maps need to be produced with spatially and temporally consistent information and accuracy. To meet these requirements, we need to develop cost-effective ways for automating the processing of satellite images and the production of LULC maps with high temporal resolution (Defries & Belward, 2000; Skole et al., 1997).

There is a strong tradition of using data from medium resolution sensors (10–60 m)–especially Landsat–for mapping LULC change at local to national scales (Alves & Skole, 1996; Steininger et al., 2001; Roberts et al., 2002; Zak et al., 2004; Boletta et al., 2006; Killeen et al., 2007; Gasparri & Grau 2009; Huang et al., 2009). This level of spatial resolution is generally sufficient for detecting fine-scale land use patterns. However, data costs, small image extent, cloud cover, haze,

* Corresponding author. Tel.: +1 707 664 2558.
E-mail address: mateolclark@gmail.com (M.L. Clark).

0034-4257/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2010.07.001

and infrequent measurements can make data from medium resolution sensors impractical for regional and global mapping (Asner, 2001; Hansen et al., 2008).

Satellites such as MODIS, SPOT-Vegetation, and MERIS offer multispectral measurements with lower spatial resolution (250 to 1000 m), relatively large scenes, and near-daily coverage that allow multiple observations in a year despite cloud coverage. Multi-temporal and multispectral analysis of these data can be used to produce LULC maps and other land cover descriptors, such as the timing, length and frequency of vegetation growing seasons. Several global land cover maps have been produced from low resolution satellites: 1.1-km AVHRR (IGBP DISCover , Loveland et al., 2000; UMD GLCC, Hansen et al., 2000), 1-km SPOT-Vegetation (GLC2000, Bartholomé & Belward, 2005), 500 m and 1000 m MODIS (MOD12Q1, Friedl et al., 2002; MCD12Q1, Friedl et al., 2010), and 300 m MERIS (Globcover, Bicheron et al., 2008). These map products generally focus on separating natural vegetation types for global carbon assessment and differ by source images, spatial scale, reference data, classification techniques and class rules, making comparison problematic (Herold et al., 2008). Evergreen broadleaf trees and areas without vegetation (snow, ice, barren) tend to be well classified, but accuracy is poor with large pixels that mix spectral and temporal signals from trees, shrubs and herbaceous vegetation (Herold et al., 2008). Most global maps provide "baseline" information from a single time period (mostly circa 2000), thus precluding analysis of LULC change using one product with consistent error. The 500 m MCD12Q1 MODIS product offers annual LULC maps from 2001 to 2007, with plans to continue into the future, but these products have just been released and have not been thoroughly assessed for class accuracy and change detection (Friedl et al., 2010).

In summary, there is a lack of LULC map products at regional to continental scales with spatially and temporally consistent information content and error, which prevents analyses of coupled natural and human systems across political boundaries and through time. To satisfy this need, we develop a scalable method for mapping annual LULC at these spatial scales. We use the Dry Chaco ecoregion in South America as a case study; however, the main impetus for this study is a larger project focused on recent LULC change in Latin America and the Caribbean. Our method is novel in that it integrates: 1) reference data interpreted from high-resolution imagery sampled in space and time within an Internet-based tool (Google Earth), 2) a flexible classification scheme based on percent cover thresholds, 3) predictor variables that respond to phenological variation in MODIS vegetation index time series data, 4) annual maps produced at 250 m scale, and 5) a Random Forests classifier that is robust in the face of heterogeneous classes and reference data error.

2. Study area: Dry Chaco ecoregion

To develop and test our method, we worked in the Dry Chaco ecoregion that spans Argentina, Bolivia and Paraguay (Olson et al., 2001). This is the second largest ecoregion in Latin America, covering 790,000 km^2 between 17°32′26″S and 33°52′7″S latitude and 67°43′12″W and 57°59′26″W longitude (Fig. 1), and includes the largest continuous neotropical dry forest (Eva et al., 2004). The ecoregion is characterized by a monsoonal climate with a strong seasonality (dry winters, rainy summers), but average temperature and in particular rainfall vary significantly across the area. Annual mean temperature varies from 18°C in the southern part of the ecoregion to 21°C in the north, and precipitation varies from 500 mm/year in the center to 1000 mm/year in the eastern and western extremes (Minetti, 1999). The vegetation is dominated by dry forest trees and shrubs, but natural grasslands occur in areas with sandy soils and frequent fires. Most of the ecoregion is flat, with elevation rising in the wetter, western side. Elevation above mean sea level for the ecoregion calculated from a digital elevation model (see Section 2.7) had a range of 56 to 3577 m and average of 326 ± 289 m. Lowland areas, having elevation below 700 m, cover 91% of the ecoregion.

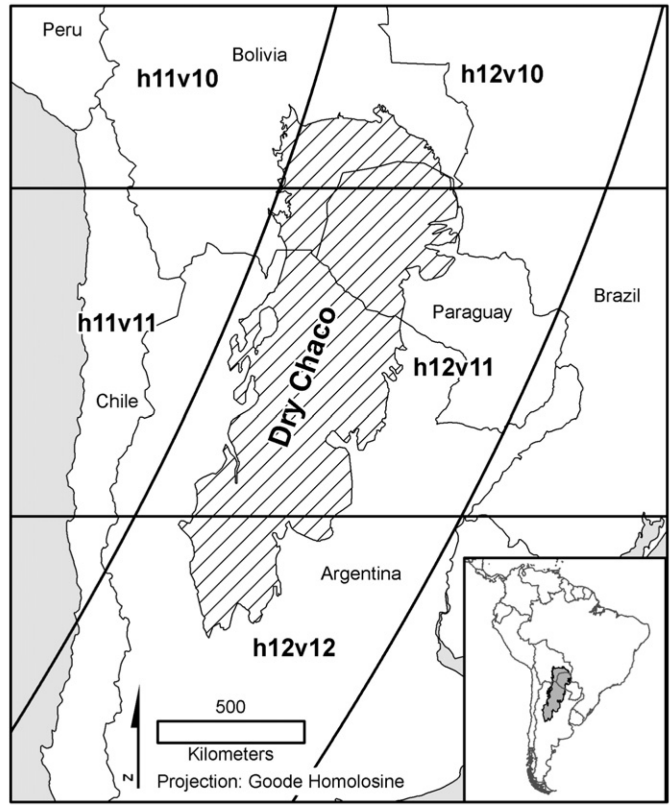

Fig. 1. The Dry Chaco ecoregion study site. MODIS tile extents are shown as the grid, and tile numbers are given with horizontal and vertical reference numbers (e.g., h11v10).

Historically, much of the ecoregion has been severely degraded by extensive cattle ranching and timber and charcoal extraction. In addition, agriculture has occurred in the foothills of the Andes and in irrigated valleys for more than a century. Citrus plantations (mostly lemon, some grapefruit and oranges) and sugar cane are important crops near the Andes. In the northern humid zones there are some banana plantations, while olives are cultivated in the drier southwest. The region also has some scattered tree plantations (pine, eucalyptus and poplar) and minor fruit orchards including blueberries, peaches, figs, and walnuts. During the last 30 years, the conversion of forest to agriculture has accelerated, mostly driven by growing global food demand (Grau et al., 2005). In Argentina, the majority of this new deforestation has occurred in the wetter eastern and western parts of the ecoregion, where millions of hectares of Chaco forest have been replaced with soybeans and pastures (Gasparri & Grau, 2009).

As a consequence of its large area and rapid land use change, the ecoregion has the largest carbon stock and the largest source of emissions from deforestation in Latin America outside the Amazon basin (Gasparri et al., 2008). The ecoregion is part of the Tropical/Subtropical Dry Broadleaf Forest biome, which globally has a high percentage of area converted and relatively low protection (Hoekstra et al., 2005), and it has the largest continuous habitat for large mammals (e.g., jaguars, peccaries) outside the Amazon basin, making it important for regional conservation (Altritcher and Boaglio, 2004; Altritcher et al., 2006; Redford et al., 1990). Given its large extent and rapid changes, developing remote sensing methods to monitor LULC is a priority for the Dry Chaco. Existing studies have used Landsat images, covering parts of the ecoregion in Bolivia (Killeen et al., 2007), Paraguay (Huang et al., 2009) and Argentina (Zak et al., 2004; Boletta et al., 2006; Gasparri & Grau 2009). Methods varied from automated per-pixel classification to visual interpretation; however, despite the lack of consistent products, all studies reported accelerated deforestation since the 1990s.