Int J Appl Earth Obs Geoinformation 63 (2017) 196–205

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

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

Identifying tropical dry forests extent and succession via the use of machine learning techniques

Wei Li[a,b], Sen Cao[b], Carlos Campos-Vargas[b], Arturo Sanchez-Azofeifa[b,*]

[a] College of Information and Communication Engineering, Harbin Engineering University, Harbin, 150001, China
[b] Alberta Centre for Earth Observation Sciences, Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, T6G 2E3, Canada

ARTICLE INFO

Keywords:
Secondary tropical dry forest
Hyperspectral MAPper dataset
Waveform LIDAR data
Machine learning
Feature-level fusion

ABSTRACT

Information on ecosystem services as a function of the successional stage for secondary tropical dry forests (TDFs) is scarce and limited. Secondary TDFs succession is defined as regrowth following a complete forest clearance for cattle growth or agriculture activities. In the context of large conservation initiatives, the identification of the extent, structure and composition of secondary TDFs can serve as key elements to estimate the effectiveness of such activities. As such, in this study we evaluate the use of a Hyperspectral MAPper (HyMap) dataset and a waveform LIDAR dataset for characterization of different levels of intra-secondary forests stages at the Santa Rosa National Park (SRNP) Environmental Monitoring Super Site located in Costa Rica. Specifically, a multi-task learning based machine learning classifier (MLC-MTL) is employed on the first shortwave infrared (SWIR1) of HyMap in order to identify the variability of aboveground biomass of secondary TDFs along a successional gradient. Our paper recognizes that the process of ecological succession is not deterministic but a combination of transitional forests types along a stochastic path that depends on ecological, edaphic, land use, and micro-meteorological conditions, and our results provide a new way to obtain the spatial distribution of three main types of TDFs successional stages.

1. Introduction

Tropical dry forests (TDFs) contain a wealth of unique biodiversity and are important habitats for millions of people (Janzen 1988). However, they are considered among the most endangered ecosystems on Earth (Hoekstra et al., 2005; Olson et al., 2001). Over the last 60-years, the extent of TDFs in Latin America has been systematically reduced as a result of the combination of anthropogenic factors, such as demographic growth, economic booms, and agricultural and colonization policies (Portillo-Quintero and Sanchez-Azofeifa, 2010). As today it is estimated that less than 40% of its total extension remains in Latin America. This extension, in the order of 500,000 km^2, is highly fragmented and under different degrees of ecological succession. Today, TDFs landscapes are a combination of different land-use and land-cover classes, being defined as “agro-landscapes” for the purpose of implementation conservation and management programs (Cao et al., 2015).

Mapping the ecological extent of secondary tropical dry forests has been mostly driven by assumptions that their extent is deterministic (e.g. forests and non-forests with no transitions from one forest class to another). In contrast, forest regeneration processes in tropical dry forests are not a deterministic phenomenon but a continuous stochastic phenomenon driven by wind and vertebrates combined by fire control efforts (Janzen 1988). From a pure ecological point of view, the forest vertical and horizontal structure, leaf area index (LAI), photosynthetically active radiation, seasonality and species composition help to drive differences between different secondary TDFs stages. These differences allow us to divide TDFs along an ecological gradient in which we can detect early, intermediate, and late succession or old growth succession (Arroyo-Mora et al., 2005; Cao et al., 2015; Chazdon 2008; Kalacska et al., 2004a). This classification into different ecological successional stages is essential to evaluate conservation policies (Sanchez-Azofeifa et al., 2013).

During the process of ecological transition from early stage to late stage, intermediate TDFs have the highest species diversity, larger variability on canopy openness, and significant differences in species composition (Kalascska et al., 2004b). Moreover, intermediate forests of different ages show significant differences on structure and composition among themselves. These differences are in general driven by species turn-over which in turn cause a very dynamic structure and forest species composition (Cao et al., 2015;Lucas et al., 2000). Such structural and compositional dynamics are generally ignored by

* Corresponding author.
E-mail address: arturo.sanchez@ualberta.ca (A. Sanchez-Azofeifa).

http://dx.doi.org/10.1016/j.jag.2017.08.003
Received 9 April 2017; Received in revised form 12 July 2017; Accepted 5 August 2017

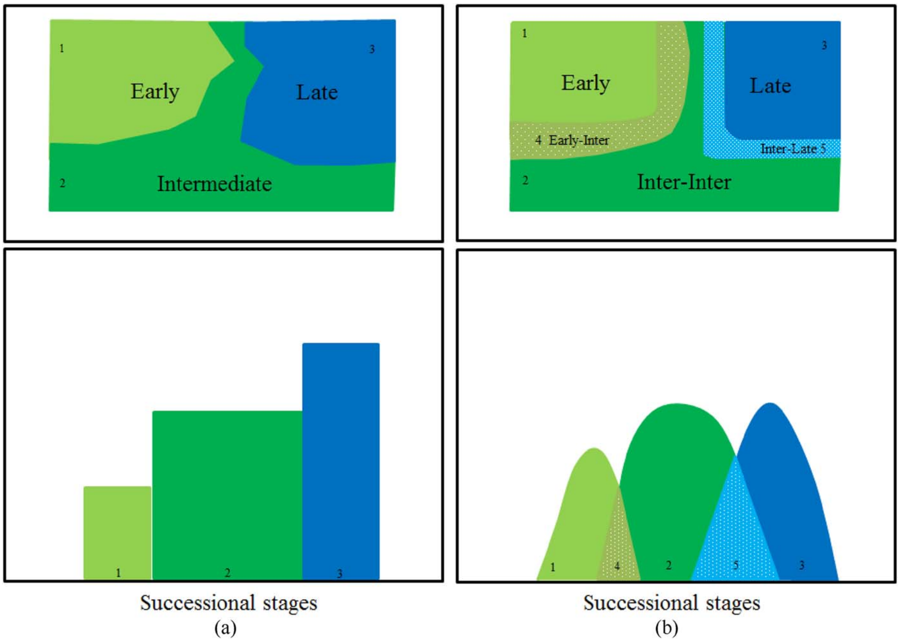

Fig. 1. The traditional and our principles of mapping secondary TDFs. (a) Traditional methods assume that different successional stages have deterministic extents. There are no transitions between successions. (b) Our method considers the forests regeneration as a continuous stochastic phenomenon, where there are transitions between successions.

secondary mapping efforts that either map TDFs as primary or secondary, or deterministically divide secondary forests into early, intermediate and late successions without considering the subtle differences that are present on intermediate forests as they evolve from an early to a late succession (Fig. 1).

Remote sensing techniques have the potential to quantify secondary TDFs succession. Early studies have mainly focused on linking vegetation indices or spectral reflectance with specific forest structural parameters (e.g., canopy openness, tree height, and LAI) or biodiversity parameters (e.g., the Holdridge Complexity Index and the Shannon Diversity Index) (Kalacska et al., 2004a; Kalacska et al., 2004b). For example, Kalacska et al. (2007a,b) used hyperspectral remote sensing in the wet and dry seasons and in the transition from wet to dry seasons, to study the relationship between succession and a series of forest structural variables that define the Holdridge Complexity Index; an ecological variable used to separate forests along a successional gradient. Their study highlighted the importance of the dry season in TDFs succession mapping, and showed the pronounced differences in shortwave infrared (SWIR1; ~1420 nm to 1920 nm) (Kalacska and Sanchez-Azofeifa 2008). Latest work has involved more advanced remote sensing products and algorithms such as LIDAR (Castillo et al., 2012; Castillo et al., 2011), multi-angle remote sensing (Millan et al., 2014; Millan et al., 2015), and hyperspectral spectral mixture analysis (Cao et al., 2015) with the aim of providing detailed and accurate information on ecological succession processes to estimate ecosystem services. However, these efforts have proved that the use of single optical data is not enough for the succession mapping in TDFs, and additional variables should be considered.

Over the years, different classifiers have been developed to improve the classification accuracy in forests area using hyperspectral images. Three levels of features are typically employed: spectral value features, gradient features, and texture features (Haralick and Shanmugam 1973; Kalacska et al., 2007a; Ohmann and Gregory 2002; Yang and Zhang 2010). Spectral value features can be connected to the biomass of the TDFs during the dry season (Kalacska and Sanchez-Azofeifa, 2008; Kalacska et al., 2004b). Gradient features can be used to quantify relations between vegetation and the environment for a sample of field plot locations (Ohmann and Gregory 2002), and can reflect biomass changes among different stages of forest regeneration (Huisman and Weissing 1994; Pflugmacher et al., 2014). Texture features have the potential to delineate different forest types, horizontal forest structure, biomass, and land cover categories (Bastin et al., 2014; Drake et al., 2003; Ghimire et al., 2010; Wang et al., 2012). Therefore, it is beneficial to explore them altogether using machine learning classifiers since these techniques can efficiently learn from these features and incorporate prior knowledge into the identification process. The multi-task learning based machine learning classifier (MLC-MTL) is one of the machine learning classifiers that can fuse different features of a hyperspectral image to acquire accurate mapping results (Liu et al., 2009; Yi et al., 2011). MLC-MTL firstly builds a feature dictionary which contains spectral, gradient, and texture features of training samples. Then, it fits Gaussian mixture models to the feature dictionary of training samples and derives the coefficients of the Gaussian mixture models (Yu et al., 2012). Finally, the mapping result can be acquired according to the location of the coefficients corresponding to the samples of the feature dictionary (Yi et al., 2011).

In this paper, we aim to evaluate the possibility of deconstructing secondary TDFs forests along their successional path based on advanced machine-learning techniques. Our approach uses hyperspectral remote sensing in the SWIR1 spectrum, plus information generated from the LIDAR vegetation imaging system (LVIS) to explore transitional zones between different types of intermediate secondary dry forests. LVIS data are first processed to derive metrics of forest structure in three dimensions (Castillo et al., 2012; Castillo et al., 2011; Kalacska and Sanchez-Azofeifa 2008), which in turn are integrated with hyperspectral remote sensing information using our proposed MLC-MTL to divide the intermediate TDFs into three classes: early-intermediate (10–20 years old), intermediate–intermediate (20–40 years old), and intermediate-late (> 40 years old).

2. Study area and data

2.1. Study area

Our study area is located at the Santa Rosa National Park (SRNP) Environmental Monitoring Super Site, Guanacaste, Costa Rica (Fig. 2). The study area covers an area of approximately 66.2 km^2 and it was a part of a large Hacienda-system dedicated to cattle ranching for almost 200 years before becoming a national park. Current land cover at the SRNP is a combination of pasture areas used as fire breaks and a mosaic of secondary TDFs divided into three successional stage categories: (1) early successional forests, consisting of pasture with shrubs, short