International Journal of Applied Earth Observation and Geoinformation 25 (2013) 87–97

Contents lists available at SciVerse ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

Stochastic gradient boosting classification trees for forest fuel types mapping through airborne laser scanning and IRS LISS-III imagery

G. Chirici[a,*], R. Scotti[b], A. Montaghi[c], A. Barbati[d], R. Cartisano[d], G. Lopez[a], M. Marchetti[a], R.E. McRoberts[e], H. Olsson[c], P. Corona[d]

[a] Dipartimento di Bioscienze e Territorio, Università degli Studi del Molise, C.da F.te Lappone, snc, 86090 Pesche, Isernia, Italy
[b] Dipartimento di Agraria (Nuoro Forestry School), Università degli Studi di Sassari, Via Cristoforo Colombo, 1, 08100 Nuoro, Italy
[c] Department of Forest Resource Management, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden
[d] Dipartimento per la Innovazione nei Sistemi Biologici, Agroalimentari e Forestali (DIBAF), Università degli Studi della Tuscia, via San Camillo de Lellis, snc, 01100 Viterbo, Italy
[e] Northern Research Station, U.S. Forest Service, Saint Paul, MN, USA

ARTICLE INFO

Article history:
Received 20 December 2012
Accepted 29 April 2013

Keywords:
Airborne laser scanning
IRS LISS-III imagery
Forest fuel type mapping
Forest fires
Mediterranean forests
Classification and regression trees

ABSTRACT

This paper presents an application of Airborne Laser Scanning (ALS) data in conjunction with an IRS LISS-III image for mapping forest fuel types. For two study areas of 165 km^2 and 487 km^2 in Sicily (Italy), 16,761 plots of size 30-m $\times$ 30-m were distributed using a tessellation-based stratified sampling scheme. ALS metrics and spectral signatures from IRS extracted for each plot were used as predictors to classify forest fuel types observed and identified by photointerpretation and fieldwork. Following use of traditional parametric methods that produced unsatisfactory results, three non-parametric classification approaches were tested: (i) classification and regression tree (CART), (ii) the CART bagging method called Random Forests, and (iii) the CART bagging/boosting stochastic gradient boosting (SGB) approach. This contribution summarizes previous experiences using ALS data for estimating forest variables useful for fire management in general and for fuel type mapping, in particular. It summarizes characteristics of classification and regression trees, presents the pre-processing operation, the classification algorithms, and the achieved results. The results demonstrated superiority of the SGB method with overall accuracy of 84%. The most relevant ALS metric was canopy cover, defined as the percent of non-ground returns. Other relevant metrics included the spectral information from IRS and several other ALS metrics such as percentiles of the height distribution, the mean height of all returns, and the number of returns.

1. Introduction

1.1. Forest fuel type mapping

Fire is one of the most important factors affecting the dynamics of Mediterranean forests, and its intensity and recurrence may substantially affect successional pathways, even preventing post-fire auto-successional trends (Moreira et al., 2011; Brivio et al., 2009). In the period 2006–2009, fires affected approximately 1,366,000 ha of European forests (Forest Europe, 2011). The European Forest Fire Information System of the European Commission (EC, 2011) indicates that the most affected areas, excluding the Russian Federation, are the five southern member states: Portugal, Spain, France, Italy and Greece. In these five countries, fires burnt a total area of 14,620,968 ha during the period 1980–2010. Annual totals varied from 158,621 ha in 2008 to almost one million ha in 1985.

Fires reduce flows of forest ecosystem services (Mavsar et al., 2012) such as wood and non-wood products, game for hunting, soil protection, recreational and social value, carbon sequestration, biodiversity protection.

The amount of fuel loading, as well as forest composition and forest structure, affect the characteristics and frequency of forest fires (Schmidt et al., 2002). Thus, an accurate description of the forest in terms of fuel conditions is essential to support fire management and to predict fire risk (Chuvieco et al., 2004).

Describing fuel conditions is an extremely complex task because multiple variables must be considered. The total amount of biomass in different forest components (herbs, shrubs, trees), as well as fuel structural characteristics (surface-area-to volume ratio, fuel density, fuel loading, height and stratification of fuel strata), fuel chemical composition and moisture content all affect fire behavior. To simplify the description of a forest area in terms of these characteristics, vegetation types with similar fuel conditions are grouped

* Corresponding author. Tel.: +39 0874404138.
E-mail address: gherardo.chirici@unimol.it (G. Chirici).

0303-2434/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jag.2013.04.006

together to form "fuel types". Forest areas classified as the same fuel type have similar fire hazard and/or fire propagation behavior (Hardy, 2005; Pyne et al., 1996). Thus, mapping fuel types is critical for fire prevention planning, from broad area estimation of fire hazard to local-scale simulation of fire propagation behavior.

Several fuel type classification systems are used worldwide (Arroyo et al., 2008). One of the most widely used is the system of 13 standard fuel types developed by the Northern Forest Fire Laboratory (NFFL), Rocky Mountain Research Station, U.S. Forest Service (Burgan and Rothermel, 1984). This system is used for input to the most common fire simulation models (Behave, Farsite). The Prometheus system (Chuvieco et al., 2003) is an adaptation of the NFFL classification to Mediterranean conditions and includes seven fuel types based on the type and height of the propagation element which are further divided into three major groups: grasses, shrubs and ground litter.

Fuel type mapping is critical for spatial modeling of forest fire behavior, particularly for characterizing surface and crown fuels (Arroyo et al., 2008; Chuvieco et al., 2003; Keane et al., 2001) and fire hazards for different forest types (Xanthopoulus et al., 2012). Accordingly, the impact of alternative fuel treatments such as silvicultural interventions and prescribed fire on fuel properties and the effects of the treatments on fire propagation can be better understood.

Traditionally, fuel types have been mapped by field survey, a costly and time-consuming activity. In searches for less costly alternatives, researchers have spent the last decades investigating the potential role of remote sensing for fuel type mapping.

Optical images are not sensitive to below-canopy forest attributes (Keane et al., 2001), and therefore have a limited ability to detect vertical fuel distribution, understory characteristics, and the surface height of crown fuels (Arroyo et al., 2008). Airborne Laser Scanning (ALS), on the other hand, is a promising technology for generating reliable representations of horizontal and vertical forest structure. Its capacity for scanning wide areas and producing precise vertical and horizontal estimates of forest attributes (Lefsky et al., 1999; Naesset and Økland, 2002; Ahokas et al., 2003) is highly valued.

A large number of studies, conducted mostly in temperate and boreal forests of Europe and North America, indicate the potential of discrete return ALS, in combination with optical remotely sensed images, for estimating tree or forest variables as components of fuel models (e.g. Riaño et al., 2003; Andersen et al., 2005; Mutlu et al., 2008; Erdody and Moskal, 2010). For example, tree height can be readily estimated from the ALS point cloud (e.g. Magnussen et al., 1999; Naesset and Økland, 2002; Popescu et al., 2002). Also, crown base height, a critical fuel property used to estimate surface and crown fires, can be estimated using automated, individual tree recognition methods that extract single tree locations and delineate full crowns (e.g.: Hyyppä et al., 2001; Persson et al., 2002; Morsdorf et al., 2004). Crown variables such as size (diameter, area, or volume), biomass or bulk density (foliage biomass divided by the crown volume) can be empirically estimated from ALS data with parametric methods such as allometric models (Riaño et al., 2004; Popescu, 2007) or non-parametric data mining methods such as Support Vector Machines (SVM) (Zhao et al., 2011) or classification and regression trees (CART) (Yu et al., 2010).

For predicting the characteristics of ground fuels (grass, shrubs, small trees), especially for low vegetation canopy heights (Riaño et al., 2007) ALS may be of limited value. The extraction of information on understory vegetation such as percentage cover and height of the shrub layer has been investigated in only a few of studies (e.g. Harding et al., 2001; Riaño et al., 2003; Maltamo et al., 2006). Promising approaches based on the decomposition of full waveform raw ALS data have also been proposed, but these applications require more in-depth investigation as soon as data became more available (Hug et al., 2004; Persson et al., 2005; Chauve et al., 2007).

In exploratory work, Seielstad and Queen (2003) empirically demonstrated that the profiles of laser returns, as a function of height above the estimated median ground surface, could be used to characterize fuel types 8 (closed timber litter) and 10 (timber) in the Anderson (1982) nomenclature system. Mutlu et al. (2008), Koetz et al. (2008) and García et al. (2011) investigated the use of a combination of ALS and multispectral optical imagery for mapping fuel types. Mutlu et al. (2008) generated a ALS multiband and combined it with multispectral Quick Bird bands to achieve a per-pixel discrimination of Anderson's fuel types using traditional supervised parametric classifiers. A similar approach was used by Koetz et al. (2008) to classify general land cover classes with SVM using a combined dataset with images from an AISA/Eagle imaging spectrometer and six gridded ALS variables. García et al. (2011) used SVM to classify a multi-dimensional dataset constructed from four raw multispectral bands, nine additional indexes from an Airborne Thematic Mapper (ATM), and eight indexes from ALS data for 10 land and forest cover types. Forest types were then further classified via the Prometheus fuel models system (Chuvieco et al., 2003) with a decision tree using ALS metrics.

Our study was conducted in two large study areas in Sicily (Italy) for two reasons: (i) to cover southern Mediterranean biogeographical areas where environmental conditions are characterized by extreme drought and sparse forest vegetation but have yet to be covered in previous studies, and (ii) because in these areas raw ALS data are available together with an official fuel type map recently released by the local forest administration of Regione Sicilia.

Because our first tentative tests to predict fuel types on the basis of ALS and IRS data based on traditional parametric (maximum likelihood) and non-parametric (*k*-Nearest Neighbors) methods were unsuccessful, we decided to test CART methods that were recently applied for similar purposes by García et al. (2011).

1.2. Classification and regression trees

Traditional parametric classification approaches are based on distributional assumptions that are often not satisfied for remotely sensed variables (Muñoz and Felicísimo, 2004; Lawrence et al., 2004). Alternative approaches including machine learning techniques such as neural networks (Benediktsson et al., 1990) and decision trees (Friedl and Brodley, 1997) that do not rely on assumed distributions merit consideration.

Binary recursive partitioning (BRP) (Merkle and Shaffer, 2011) or Classification Tree Analysis, implemented in the well-known CART data mining tool (www.salford-systems.com), is a supervised approach that recursively splits the training data into progressively more pure subsets by defining thresholds within the multidimensional predictors ranges. For remote sensing applications, predictors are usually multispectral data from optical imagery or other statistics acquired from active sensors such as radar or ALS, while target classes are based on land use/land cover nomenclature systems.

Procedures based on CART are potentially useful for classification problems because they incorporate nonlinear effects without requiring additional variables or variable transformations (Breiman et al., 1984). In addition, outliers have limited impacts on results, and there is no theoretical constraint concerning the number and type of variables to be considered as potential predictors; in particular, predictor collinearity does not adversely affect numerical calculations (Strobl et al., 2008).

BRP output is usually presented as an inverted tree structure called a "decision tree". Starting from the top (the root of the tree), and depending on the values of the predictor variables, a single observation progresses down through the different branches