



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. Montagni^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 km² and 487 km² in Sicily (Italy), 16,761 plots of size 30-m × 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.

© 2013 Elsevier B.V. All rights reserved.

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).

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 (Xanthopoulos 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.: Hyypä 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

(splits) that originate from each node, until it reaches the terminal node where it is assigned to a class. BRP splits or threshold values are selected using empirical risk minimization criteria which minimize classification errors evaluated against the training data. BRP methods commonly assume independent and identically distributed observations, but they make no assumptions about error distributions (Breiman et al., 1984). Once the final decision tree is constructed, it can be applied to all the elements of the population which, in traditional remote sensing applications, are the pixels of the input images.

Optimization with traditional CART approach is difficult because the selected decision tree may not lead to the absolute best classification. Because all partitioning strategies used for constructing decision trees are based on randomization, the same result is unlikely to be achieved with two independent runs. Nonetheless, trees are unbiased estimators (Dietterich, 2000) and hence, on average, produce correct predictions. Because decisions taken at one node of the tree cannot change the decisions taken at previous nodes, the combined results of the optimal decisions at each node may not result in the best classification (Lawrence et al., 2004). This problem may be addressed by boosting or bagging, both of which generate a group or ensemble of different decision trees. The classification results generated by the different trees are aggregated by majority or weighted majority to produce the final optimal result. Several methods for generating the tree ensembles may be considered (Lawrence et al., 2004).

Boosting is based on a first classification tree with subsequent trees generated by assigning greater weights to incorrectly classified training data (Bauer and Kohavi, 1999). Friedman (1999) presented a variant called gradient boosting.

Bagging is based on a bootstrapping approach for which multiple classification trees are developed by repeatedly selecting random subsets of the training data. Breiman (2001) called this classifier “Random Forests”™ (<http://www.stat.berkeley.edu/~breiman/RandomForests/cc.home.htm>).

Friedman (2002) proposed a hybrid between the boosting and bagging procedures called stochastic gradient boosting (SGB). Only a limited number of forestry applications of the SGB method have been reported, the most important of which relate to the use of predictors based on remotely sensed data. Good performance of SGB relative to traditional parametric methods were found for general land use/land cover remote sensing classification problems using multispectral or hyperspectral images for several study areas in the USA (Lawrence et al., 2004). In addition, Moisen et al. (2006) predicted presence/absence and basal area of 13 forest species in Utah (USA) using Landsat 7 ETM+ and ancillary information and found that SGB provided the most stable results when compared to generalized additive models. More recently, SGB was used to estimate aboveground forest biomass in Guinea-Bissau using radar data. Interesting results were obtained, but the mean error of the estimations was quite large (Carreiras et al., 2012).

1.3. Aim of the study

The current state of the art demonstrated that a few studies addressed the possibility to use ALS-derived metrics, associated or not with multispectral imagery, to map fuel types (Seielstad and Queen, 2003; Mutlu et al., 2008; García et al., 2011).

CART approaches using ALS for forestry applications are not known to have been reported, and only a few papers have reported applications of SGB to remote sensing problems (Lawrence et al., 2004; Moisen et al., 2006; Carreiras et al., 2012).

Clearly, more attempts are needed to investigate the potential of ALS data for use with forest fuel models and for mapping fuel types, particularly if based on the SGB method.

These considerations have at least partially motivated the present study whose aim is to evaluate SGB for a semi-automated classification of ALS raw, multi-echo data for classifying fuel types for two Mediterranean study sites in Sicily (Italy). Results obtained using SGB are compared to results obtained using a traditional CART and Random Forests methods.

The expected outcome should facilitate future implementations of ALS technology in forest fire prevention decision support systems, particularly with respect to selection of the optimal metrics and classification method for fuel types mapping.

2. Materials and methods

2.1. Study area

Two study areas of 487 km² and 165 km² were selected, one each in the Sicilian provinces of Palermo and Agrigento (Fig. 1).

Both study areas are characterized by typical Mediterranean climate with a bi-seasonal regime of hot/dry summers and cold/wet winters. Elevation ranges from 2 m to 1879 m above sea level (asl) in the largest area and from 198 to 1434 m asl in the smaller area. A mosaic of land cover types is found in both study areas, including agricultural lands, grasslands, shrublands and forest land.

A digital fuel map of Sicily was recently developed within the framework of the Forest Information System of Sicilian Region (<https://sif.regione.sicilia.it>). The main forest types were first mapped by manual on-screen delineation of high resolution (1 m × 1 m) digital orthophotos with a nominal scale of 1:10,000 and a minimum mapping unit of 0.1 ha. Fuel types were assessed in the field during the forest inventory for 935 plots distributed over the entire Sicilian region and were then used to classify the polygons of the forest types map. On the basis of an independent accuracy assessment based on a field verification carried out for producing metadata information by Regional authority, the overall accuracy of the resulting fuel types map was 0.9 (Hofmann et al., 2011). For this study, the fuel types map was assumed to be error free. The map depicts a classification of fuel complexes with respect to the NFFL fuel types (Burgan and Rothermel, 1984) for forest stands, open shrublands and grasslands of Sicily. Fuel types associated with the different forest cover types fall into grassland,

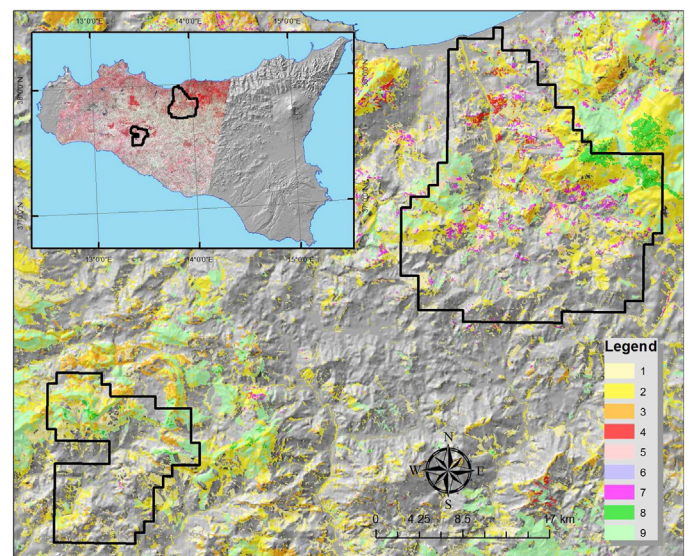


Fig. 1. Location of the study sites on the basis of the fuel types map (the system of nomenclature is in Table 1). Area 1 is smaller on south west, Area 2 is larger on north east. On the bottom left part the study areas over the IRS LISS-III imagery.

Table 1
Fuel characteristics of the fuel types mapped in Sicily.

NFFL fuel type no.	Main propagation element	Fuelbed structure	Associated forest types	Fuel load (t ha ⁻¹)
1	Grassland	Dry short grass (<1 m) Sparsely wooded grassland (tree cover <50%)	Quercus ilex Fagus sylvatica	1–2
2	Grassland	Scrubby land with mainly herbaceous vegetation Low cover woodland/shrubland (<50%) with grass understory Forest plantations with grass understory	Eucalyptus plantations	5–10
3	Grassland	Dry tall grassland (1 m)		4–6
4	Shrubland	Continuous tall shrubland (2 m)		25–35
5	Shrubland	Continuous low open shrubland (<1 m) Forest with low shrubland understory (<1 m)		5–8
6	Shrubland	Open shrubland (0.6–1.2 m) with cover >50% Low (<2 m) deciduous woodland with shrubland understory		10–15
7	Shrubland	Open high flammable shrubland (0.5–2 m) Forest with high flammable shrubland understory	Quercus suber Quercus pubescens Pinus halepensis, Pinus pinea, Pinus pinaster	10–15
8	Ground litter	Broadleaved dominated forest with compact litter Short-needle (<5 cm) coniferous plantations with undecomposed litter	Fagus Robinia Abies Cedrus Cupressus	10–12
9	Ground litter	Deciduous broadleaved dominated forest with not compact litter Thermophilous pinewoods	Quercus ilex, Quercus suber, Quercus petraeae and Quercus pubescens Quercus cerris, Castanea sativa Pinus halepensis, Pinus pinea, Pinus pinaster	7–9

shrubland or ground litter types depending on the vertical structure of fuelbeds and, accordingly, the fuelbed stratum identified as the main vector for surface fire propagation (grass/shrub understory, ground litter). A description of the nomenclature system is reported in Table 1.

2.2. ALS data

ALS data for both study areas were acquired during July and August 2010 by an Optech Pegasus system on board of a Piper Navajo I-BGFE with a scan angle of 20° at 100 kHz resulting in an average density of 1.5 points per square meter.

A Progressive Triangular Network (TIN) densification method, developed by Axelsson (1999, 2000) and implemented in Terrascan software (Soininen, 2010), was used to classify the point cloud into ground and non-ground returns. These filtered ground returns were used to generate a 1 m × 1 m pixel resolution Digital Terrain Model (DTM) using an Inverse Distance Weight (IDW) interpolation method (Andersen et al., 2005; Vepakomma et al., 2008). For this study, the parameters setting used to interpolate the ground points with IDW are those most commonly used for interpolation comparisons (Chaplot et al., 2006; Guo et al., 2010). Subsequently, we subtracted the DTM value from the height of each non-ground pulse to obtain normalized heights of the laser echoes.

2.3. Optical data

One IRS LISS-III multispectral image for June 3, 2010, was acquired for the study areas. The imagery consists of four spectral bands at 0.52–0.59, 0.62–0.68, 0.77–0.86, 1.55–1.70 μm covering green, red, near Infra Red (IR) and medium IR channels respectively. The image was geometrically corrected using ground control points acquired on the basis of high resolution orthophotos to achieve a positional accuracy less than the nominal pixel size (23.5 m).

2.4. Data preparation

The study areas were sampled using a tessellation-based stratified sampling scheme (Barabesi and Franceschi, 2011). The areas were divided into 100-m × 100-m squares, a point was randomly selected in each square, and a 30-m × 30-m plot with random orientation was centered at each point. This procedure produced 12,772 plots in Area 1 and 3989 plots in Area 2.

The plots were intersected with the fuel types map, and only those entirely contained in a single polygon of the fuel types map were retained for further investigation; the result was a total of 16,761 plots (Fig. 2).

The most dominant fuel types in the study areas were Fuel Type 2 (21%), Type 1 (12%) and Type 9 (12%) (Fig. 3).

Based on previously reported research (e.g. Naesset, 2002; Andersen et al., 2005; Evans et al., 2009; Falkowski et al., 2009; Tesfamichael et al., 2010; Estornell et al., 2011), 31 ALS-based metrics were calculated from the normalized height returns for each 30-m × 30-m plot. The differences among fuel types are mainly due to differences in vegetation structure. Because no prior studies were available on the use of ALS data in such condition, we use a large number of ALS-based metrics. Descriptions of the selected ALS-based metrics are reported in Table 2.

The majority are statistics characterizing the distributions of the “heights” of the ALS echoes within the plots. Because vertically stratified structures characterize Mediterranean vegetation, some metrics were derived by splitting the echoes into three, equally spaced, vertical levels and computing statistics for each level. The “Crown coverage” metric is calculated as the percentage of returns with heights greater than 1 m. The “Texture” metric characterizes ground level structure through the standard deviation of the heights between 0 m and 1 m. Mean values of the ALS-based metrics for the nine target fuel types are reported in Fig. 4.

For each plot the digital numbers from the four bands of the IRS imagery were also extracted.

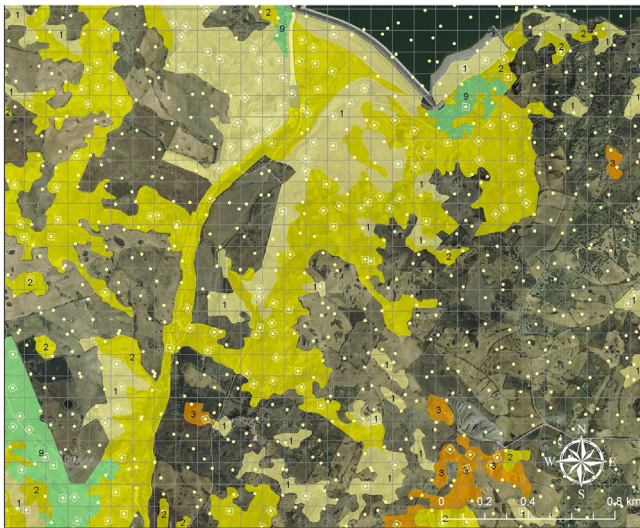


Fig. 2. Example of the sampling design adopted on the basis of the fuel type map (colors refer to the legend in Fig. 1 and labels to Table 1) and the digital orthophoto originally used for its delineation. Sampling points (in yellow) inside systematic squares of 100 m × 100 m (in gray). Around each sampling point one squared plot 30 m × 30 m (in white) is generated with random orientation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Observations for a stratified random 30%-sample of the 16,761 plots were extracted for each fuel type and used as a training dataset for developing CART models. The remaining observations were used as a validation dataset to calculate classification

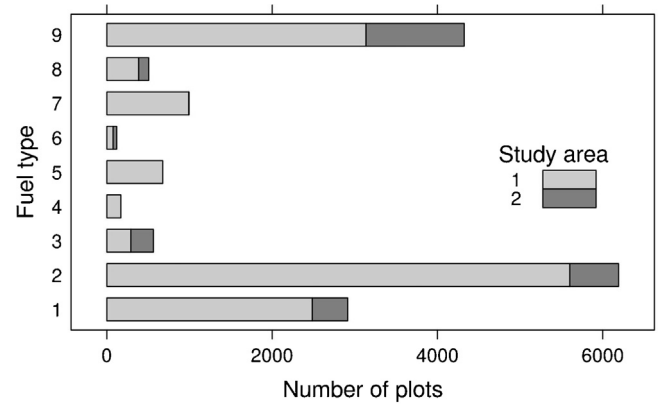


Fig. 3. Frequency distributions of fuel types in the study areas. Nomenclature of fuel types is reported in Table 1.

accuracies (Congalton and Green, 2008; Pontius and Millones, 2011).

2.5. Analyses

The analyses focused on testing three non-parametric classification algorithms based on the CART approach. The first algorithm is a simple CART (Breiman et al., 1984); the second is based on a trees ensemble constructed using the Random Forests procedure (Breiman, 2001); and the third is a stochastic gradient boosting (SGB) model (Friedman, 2002). All three algorithms are implemented in the Salford Systems Predictive Modeling Suite (SPM) and

Table 2

Definition and main statistical parameters for the ALS-based metrics calculated in the study areas. Metric in the short name of the metric. 1st QU, Median, 3rd QU and Max (the first quintile, are the statistical parameter in the study sites.

	Metric	Mean	Min	1st Qu	Median	3rd Qu	Max	M.unit	Definition
1	Returns	–	10	157	492	1066	13,509	Count	Number of returns from the plot
2	Hmax	6.5	0.1	1.9	5.1	9.5	50.1	m	Above ground elevation of highest return
3	Hmean	2.2	0.0	0.4	1.2	3.2	17.9	m	Mean height of all returns
4	Hmin	0.0	0.0	0.0	0.0	0.0	0.3	m	Lowest height among all returns
5	CV	80.0	19.6	60.7	76.0	95.1	303.5	p.n.	CV of returns height
6	Texture	0.2	0.0	0.2	0.2	0.3	0.4	m	Standard deviation of non-ground returns [0 m < h <= 1 m];
7	Skewness	0.9	–2.4	0.2	0.8	1.4	8.1	p.n.	Skewness of returns height
8	Kurtosis	1.2	–1.9	–0.7	0.1	1.8	82.3	p.n.	Kurtosis of returns height
9	relHmean	0.3	0.0	0.2	0.3	0.4	0.8	p.n.	Relative mean height [(mean – min) / (max – min)]
10	CanopyC	34.3	0.5	10.3	30.5	56.4	91.8	p.n.	Non-ground percent of total returns
11	Pct10	0.5	0.0	0.1	0.1	0.3	13.6	m	Percentile 10 of heights distribution
12	Pct20	0.9	0.0	0.1	0.3	0.6	15.8	m	Percentile 20 of heights distribution
13	Pct30	1.3	0.0	0.2	0.4	1.3	17.1	m	Percentile 30 of heights distribution
14	Pct40	1.7	0.0	0.2	0.5	2.3	18.0	m	Percentile 40 of heights distribution
15	Pct50	2.1	0.0	0.3	0.7	3.0	18.6	m	Percentile 50 of heights distribution
16	Pct60	2.5	0.0	0.4	1.0	3.7	19.3	m	Percentile 60 of heights distribution
17	Pct70	2.9	0.0	0.5	1.4	4.5	25.8	m	Percentile 70 of heights distribution
18	Pct80	3.5	0.0	0.6	2.1	5.2	31.2	m	Percentile 80 of heights distribution
19	Pct90	4.2	0.1	0.9	2.9	6.3	37.0	m	Percentile 90 of heights distribution
20	Pct99	5.7	0.1	1.6	4.4	8.4	46.5	m	Percentile 99 of heights distribution
21	Density1	0.4	0.0	0.2	0.4	0.6	1.0	p.n.	Cumulative density of returns above Hmax/3
22	Density2	0.1	0.0	0.0	0.1	0.2	0.9	p.n.	Cum. dens. returns between 1 and 2 times Hmax/3
23	Rets1pct	58.9	1.6	41.7	62.1	77.6	99.3	p.n.	Percent of returns above Hmax*2/3
24	Hmean1	0.9	0.0	0.2	0.5	1.1	5.3	m	Mean height of returns above Hmax*2/3
25	CV1	69.2	0.2	55.7	66.1	79.8	333.4	p.n.	CV of heights above Hmax*2/3
26	Rets2pct	28.8	0.2	15.8	27.5	40.8	81.5	p.n.	Percent of returns between 1 and 2 times Hmax/3
27	Hmean2	3.2	0.1	0.9	2.5	4.6	31.4	m	Mean height of returns between 1 and 2 times Hmax/3
28	CV2	18.6	0.1	17.7	18.8	20.0	45.4	p.n.	CV of heights between 1 and 2 times Hmax/3
29	Rets3pct	11.7	0.2	4.2	8.6	16.1	84.5	p.n.	Percent of returns below Hmax/3
30	Hmean3	5.1	0.1	1.5	4.0	7.4	41.7	m	Mean height of returns below Hmax/3
31	CV3	10.2	0.0	9.0	10.1	11.4	28.2	p.n.	CV of heights below Hmax/3
32	IRS1	116.3	68	100	115	130	233	p.n.	Digital number from band 1 IRS LISS-III
33	IRS2	82.5	32	63	82	100	183	p.n.	Digital number from band 2 IRS LISS-III
34	IRS3	112	36	101	111	121	192	p.n.	Digital number from band 3 IRS LISS-III
35	IRS4	119.5	44	100	120	139	200	p.n.	Digital number from band 4 IRS LISS-III

Note: p.n. = pure number.

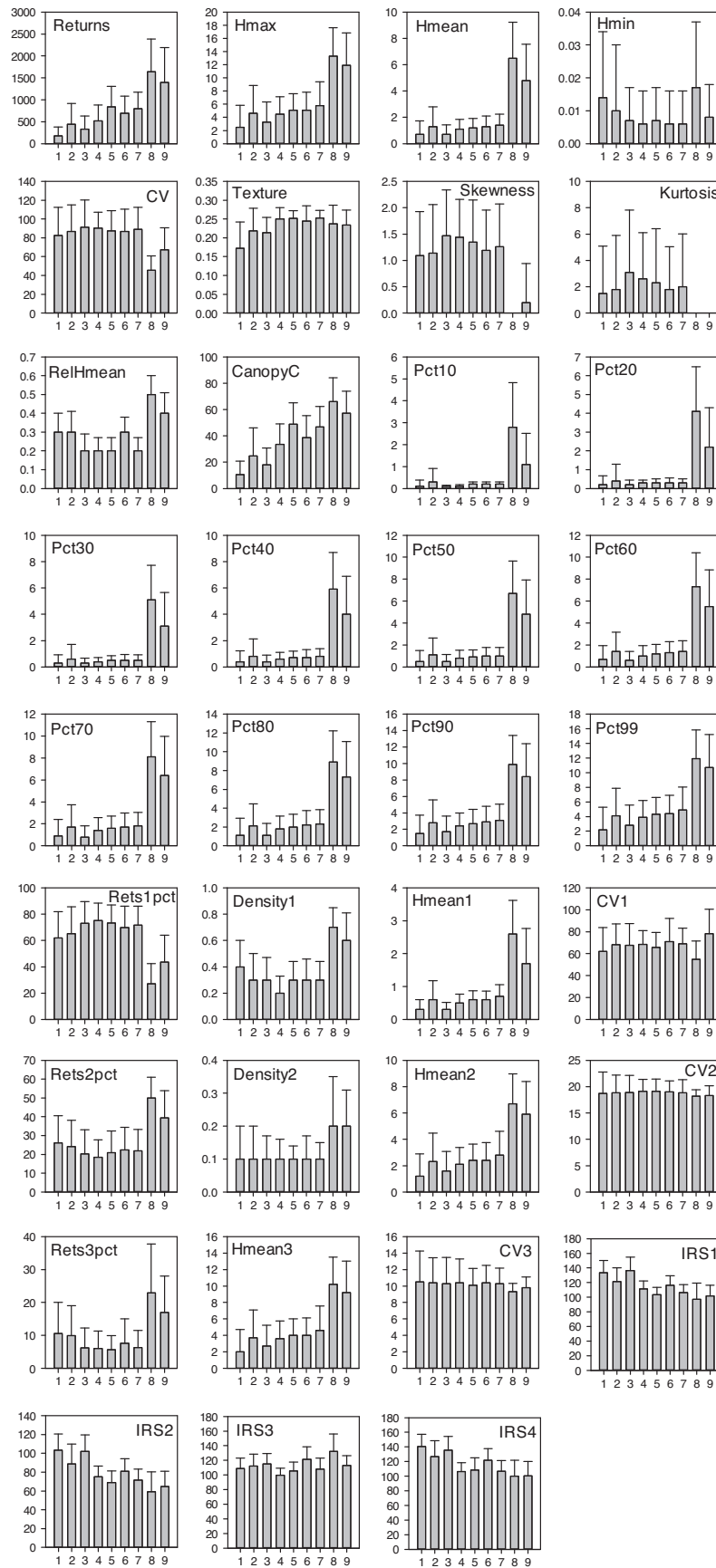


Fig. 4. Average values with standard deviations of the 31 ALS-based and 4 IRS predictors for the nine fuel types.

are commercially known as CART[®], RandomForest[®] and TreeNet[™], respectively (<http://www.salford-systems.com/spminfo.html>).

For each classification method, the importance of each predictor variable was assessed on the basis of the improvement in classification accuracy that could be attributed to the variable by comparing the results obtained with or without the inclusion of the variable. For each method, improvement values were scaled relative to the best performing predictor variable. The scores in percentages reflect the contribution each variable makes in classifying or predicting the response variable (Tolliver, 2009). In the cases of Random Forest and SGB, the values were averaged across all trees.

The models were constructed using the training datasets and then applied to predict the response variable for the validation data set. The resulting classifications were evaluated by cross-tabulating predictions and observations using a confusion matrix and calculating traditional accuracy indexes: overall accuracy, producer's and user's accuracies, and commission and omission errors (Congalton and Green, 2008). Additionally, following the procedure proposed by Pontius and Millones (2011), quantity and allocation disagreement were also calculated. The quantity disagreement is the amount of difference between predictions and observations that is due to the less than perfect match in the proportions of the classification categories (the nine fuel types, in our test). The allocation disagreement is the amount of difference between predictions and observations that is due to the less than optimal match in the spatial allocation of categories, given the proportions of the different categories in the predictions and observations set. The sum of quantity disagreement and allocation disagreement is defined as the total disagreement, that is the opposite of the overall accuracy.

2.5.1. CART

CART, as implemented in SPM, is a binary recursive partitioning procedure; it outputs a single decision tree for each run. On the basis of the nine classes of the response variable (Table 1) and the 35 predictors from the ALS and IRS training dataset (Table 2), the algorithm considers all possible ways to split the training dataset into two sub-sets (or nodes). CART ranks the splits on the basis of a popular impurity function called the "Gini criterion" (Breiman, 1996). For a given set (or node) the Gini impurity index is equal to 0 (no impurity) when all elements fall into the same target class, while maximum impurity is attained when all classes are equally represented (Merkle and Shaffer, 2011). The Gini criterion is based on impurity decrease, where the best split is identified by maximizing impurity decrease. CART recursively repeats this procedure for each child node until further splitting is impossible, either because the node is pure or it has attained the minimum number of observations. The "maximum tree" generated following this procedure was pruned by sequentially deleting non-influential splits to avoid overfitting. Pruning produces a set of smaller trees that are all nested within the original tree.

For each pruned tree, CART internally estimates the classification accuracy in terms of standard errors with the k -fold cross-validation technique using a data splitting of the validation dataset (Breiman et al., 1984). Detailed information on this procedure is available in Merkle and Shaffer (2011). Following the suggestion of Breiman et al. (1984), the smallest tree whose validation error was within one standard error of the smallest validation error in the pruned trees was selected as the final optimal tree.

2.5.2. Random Forests

A tree ensemble was generated for the training dataset using "Random Forests" as originally proposed by Breiman (2001). With Random Forests, a predetermined number of unpruned classification trees is constructed by a random, bootstrap selection from the original training data. Each node of each tree is split using a

random sample of the predictors and the best split among them is selected. Random Forests predictions are obtained by aggregating by majority the outputs produced by each tree of the ensemble.

The number of predictors to be used at each tree node and the number of trees retained in the final ensemble were the default values recommended by Breiman (in Liaw and Wiener, 2002), i.e. an ensemble of 500 trees each constructed by selecting the best of three randomly chosen predictors at each node.

2.5.3. Stochastic gradient boosting

With the boosting/bagging SGB procedure proposed by Friedman (2002), a series of small trees with six nodes each was constructed by sequentially fitting the prediction model developed using a randomly selected subsample of the training data. The sequential trees were constructed by introducing a set of probabilities of class membership for the response variable. Probabilities were calculated using the steepest gradient algorithm (Friedman, 1999) from the gradient of the loss function of the previous tree, with the gradient defined by deviance (twice the binomial negative log-likelihood) as a surrogate for misclassification rates (Lawrence et al., 2004).

For each iteration, a tree is constructed from a different random sub-sample of the dataset (selected without replacement) producing an incremental improvement in the model. The iteration process was stopped at 152 trees because a larger number of trees did not produce further improvement in the estimated accuracy of the model against the validation set.

3. Results

The distributions of fuel types were quite different for the two study areas (Fig. 3). Three fuel types were prominent: type 2, occupies approximately 40% of the plots in Area 1; type 9 dominates in Area 2; and type 1 has similar incidence, approximately 20%, in both areas; all other types have much smaller incidences.

The relevance of ALS-based metrics as fuel type predictors for the three CART variants are reported in Table 3.

The most relevant ALS metric in the SGB method was canopy cover, but four of the six most relevant predictors are from the IRS imagery (Table 3). Other relevant ALS metrics included percentiles of the height distribution, the mean height of all returns, and the number of returns (Table 3).

CART and Random Forests, with overall accuracies of only 47% and 45%, did not produce sufficiently accurate classifications. SGB instead, with overall accuracy of 84% produced substantially more accurate classifications for the two study areas (Table 4).

In terms of omission errors, the three models differed substantially (CART 34%, Random Forests 44%, SGB 27%), but it is in terms of commission errors that SGB achieved much better results (11%) than CART and Random Forests (both approximately 60%).

The most evident advantage of SGB was acceptable classification accuracy, particularly from the producer's perspective, for fuel types with relatively small numbers of observations (from type 3 to type 8) (Fig. 5).

On the other hand, all the classifiers produced comparable results for the most widespread fuel types (types 1, 2, and 9) which, in the validation datasets, represent 82% of total number of observations.

The accuracy assessment analysis conducted using the methodology proposed by Pontius and Millones (2011) produced a difference between the CART and Random Forests results that was mainly influenced by a quantity disagreement (almost 70% of the total disagreement), and between CART and SGB that was mainly influenced by the allocation disagreement (almost 60% of the total disagreement).

Table 3
Relevance of the ALS-based metrics in the different models tested. Metrics are ranked top to down on the basis of the SGB relevance. The description of the metrics is in Table 2.

Metric	CART	Random Forest	SGB
CanopyC	93.5	52.6	100.0
IRS3	37.3	67.3	94.5
IRS1	20.3	24.5	89.2
Pct70	23.1	94.2	72.1
IRS4	52.7	16.3	66.5
IRS2	45.2	71.3	66.1
Returns	21.1	38.5	62.1
Pct80	4.5	61.2	50.1
Hmean1	3.8	58.9	47.1
Hmean	4.9	78.1	44.7
Pct50	3.8	100.0	39.3
Pct60	5.6	87.5	38.2
Pct20	6.9	78.9	35.3
CV	22.8	45.3	35.2
Hmin	12.7	3.6	35.1
Hmean3	21.7	15.9	33.1
Pct90	3.7	39.7	32.1
Density2			31.1
Kurtosis	12.8	23.8	29.9
relHmean			29.4
CV1	35.5	11.4	28.9
Rets2pct	11.4	11.3	28.7
Texture	28.9	28.4	28.4
Pct10	17.2	78.9	28.3
Pct99	13.1	21.8	28.2
CV3			28.2
Hmean2			28.1
Hmax			27.3
Pct30	100.0	87.5	26.7
Skewness			25.4
CV2	21.5	11.2	22.9
Pct40			22.7
Rets3pct			22.6
Density1			18.4
Rets1pct			15.1

Analysis of the two types of disagreements for the nine fuel types (Fig. 6) indicates that the main problem for CART and Random Forests is related to the strong quantity disagreement in fuel type 2 which is confused with almost all the other fuel types and produces omission errors greater than 80%.

A combination of ALS and IRS predictions was better than either source alone. In fact the same classification approaches using only the 31 ALS predictors produced overall accuracies as follows: CART 38%, Random Forest 37%, and SGB 73%. When using only the IRS LISS-III predictors the following overall accuracies: CART 35%, Random Forests 35%, and SGB 57%.

Table 4
Confusion matrix based on the classification by SGB.

Validation										Total	User accuracy	Commission error
Fuel type	1	2	3	4	5	6	7	8	9			
Classification												
1	1666	159	22	1	2	3	5	0	13	1871	89.1%	10.9%
2	346	3966	289	33	51	8	103	33	214	5043	78.6%	21.4%
3	14	5	80	0	0	0	2	0	0	101	79.0%	21.0%
4	0	0	0	80	0	0	0	0	0	80	100.0%	0.0%
5	2	26	0	0	377	3	10	0	17	434	86.7%	13.3%
6	0	0	0	0	0	68	1	0	0	69	98.7%	1.3%
7	2	56	4	2	18	1	534	0	14	631	84.7%	15.3%
8	1	6	0	0	0	0	0	274	9	290	94.6%	5.4%
9	37	213	5	5	32	2	51	66	2804	3215	87.2%	12.8%
Total	2069	4432	400	120	479	83	705	373	3071			
Producer accuracy	80.5%	89.5%	20.0%	66.7%	78.5%	81.2%	75.8%	73.3%	91.3%			
Omission error	19.5%	10.5%	80.0%	33.3%	21.5%	18.8%	24.2%	26.7%	8.7%			

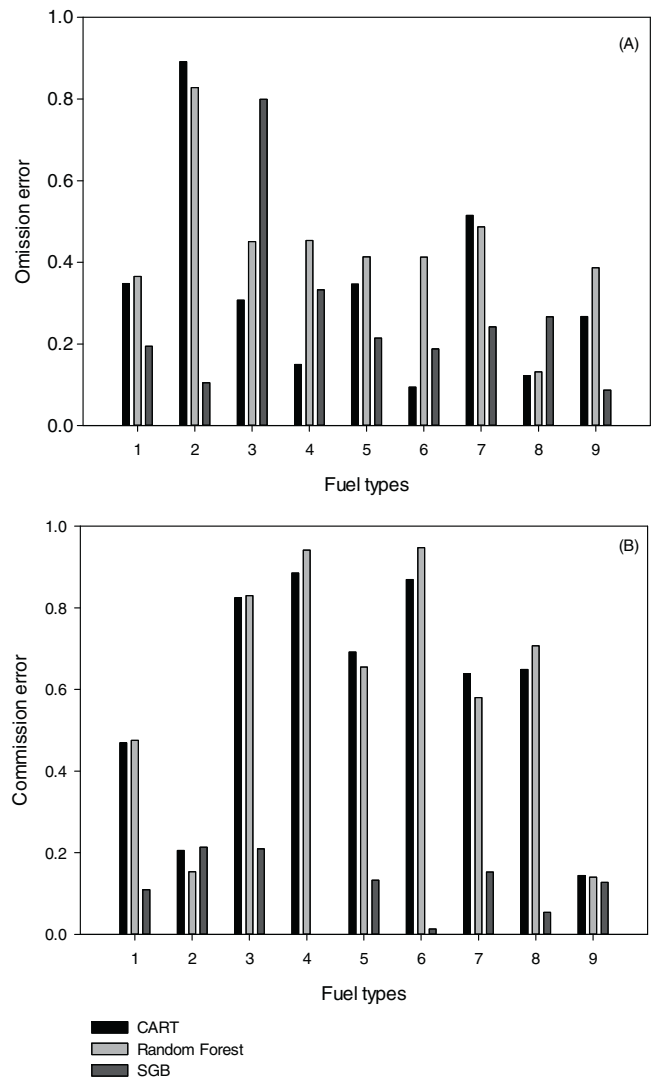


Fig. 5. Omission (A, above) and commission errors (B, below) for the three tested classification algorithms for the nine fuel types.

4. Discussion

Regularly updated fuel types maps for large areas are essential to support fire prevention actions and as input layers for fire propagation modeling, fire management, prevention, and firefighting. This

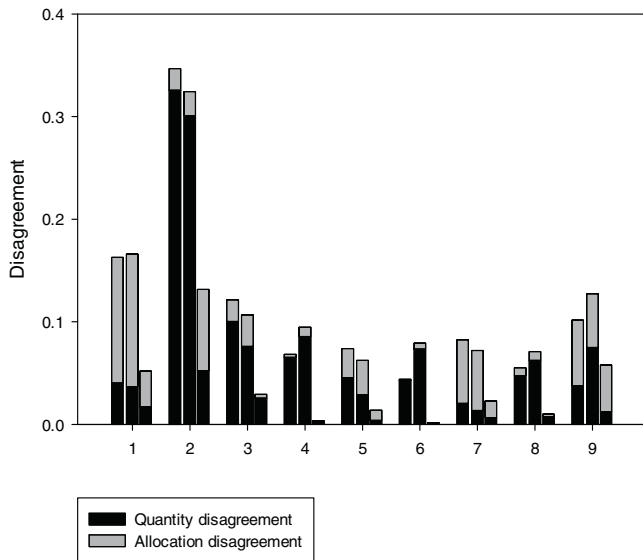


Fig. 6. Quantity disagreement and allocation disagreement for the three classification methods (from left to right: CART, Random Forests and SGB) for the nine fuel types.

is particularly true for Mediterranean areas where fire regimes are extremely severe with hundreds of thousands of hectares of forests destroyed each year. Climate change scenarios for these areas indicate increases in summer temperatures, decreases in small rainfall events, and increases in intense rainfall events leading to a potential decrease of the time between two successive fire events (Mouillot et al., 2002).

Fuels types are difficult to describe and map due to their complexity and variability (Mutlu et al., 2008). The results of this study suggest that remotely sensed data may be an essential information source for this purpose because automated and semi-automated classification systems limit time-consuming and expensive manual activities such as visual photointerpretation and fieldwork.

The utility of ALS data for characterizing forest ecosystems, both for mapping forest classes or forest types and for estimating quantitative variables such as forest biomass, has been confirmed for a large number of studies (Zhao et al., 2011; Corona et al., 2012). Many of these studies focused on estimating variables useful for fire management (Arroyo et al., 2008), but very few specifically addressed fuel types mapping. Most of the studies relied on the combined use of ALS with optical data (Mutlu et al., 2008; Koetz et al., 2008; García et al., 2011). Of importance, ALS data alone were frequently regarded as insufficient for achieving satisfactory classification accuracies (Koetz et al., 2008).

For this study, we demonstrated the utility of ALS-based metrics for predicting fuel types, although the inclusion of multispectral optical data from IRS LISS-III consistently increased the classification accuracy. A primary result is that a combination of ALS metrics and spectral information provided sufficient information for reliable assessment of fuel types.

The SGB method greatly improved the classification performance over simple and bagged CART models, confirming the potential findings reported by Lawrence et al. (2004) for application of this method for land use/land cover classifications using optical data, especially when some classes are relatively rare.

Friedman (2002) acknowledged that the reason for the positive effect obtained with bagging/boosting in the predictive accuracy of CART must be determined on a case-by-case basis. The SGB bootstrapping random procedure has the capability of reducing the variance of the response variable in the training sample; this positive effect is particularly important for more accurate classifications

of relatively rare classes. Our results confirm this hypothesis. In fact, SGB was much more effective than CART and Random Forests for classifying less abundant fuel types such as types 3–8 in our dataset which represented only 15% of the overall number of the observations.

As stressed by Lawrence et al. (2004), the ability of SGB to achieve better results cannot be generalized and is closely related to the input data, an observation that also applies to other machine learning approaches such as neural networks and SVM. For the three approaches evaluated in this study, CART produces the most transparent output, resulting in a decision tree that can be easily interpreted.

The combined use of multispectral data and ALS metrics could further improve this approach, but identification of the optimal classification method should always be based on empirical tests. Data mining approaches are, in fact, all guided by input data; hence generalization of the results is possible only in weak, mainly qualitative terms.

We further speculate that the wider availability of ALS data might foster slight changes in the approach for classifying fuel types that could both simplify their (semi) automated recognition on the basis of laser echoes and their effectiveness for characterizing the spatial distribution of vegetation structure as it pertains to fuel properties.

5. Conclusions

Three primary conclusions may be drawn from the study. First, ALS metrics in combination with optical multispectral data can be used to support supervised discrimination among different fuel types. Second, the combination of ALS metrics and multi-spectral data produced greater accuracies than either data source used separately. Third, among the non-parametric procedures we evaluated, satisfactory accuracy (84%) was obtained for only the stochastic gradient boosting method, whereas traditional CART and Random Forests produced much smaller accuracies. This result may be attributed to the capability of SGB to better classify rarer fuel types that in the dataset.

Based on these conclusions, we offer four recommendations. First, because these methods are all data driven, we recommend considerable caution when generalizing the results to different conditions.

Second, for mapping fuel types, we recommend use of a combination of ALS and optical data rather than data from either source alone. Third, we expect that the use of remotely sensed data, together with advanced (semi) automated classification methods, will contribute to greater availability of forest information useful for preventing and fighting wild fires and reducing their social, economic, and environmental effects. Therefore, we recommend further development of this approach to fuel types as a means of enhancing the effectiveness of operational use of fire-behavior models for large areas. Fourth, ALS data can be used to estimate canopy and ground fuel characteristics using height measurements (Andersen et al., 2005) which, in turn, can be used to construct wall-to-wall maps of fuel properties. Therefore, for future applications, we recommend production of thematic layers related to fuel characteristics rather than construction of more traditional fuel types maps based on aggregated system of nomenclatures.

Acknowledgements

This work was carried out within the activities of the EFI Project Center MOUNTFOR, it was supported by the ITALID project of the Dipartimento di Bioscienze e Territorio of the University of Molise. The work was also partially carried out under the project FUME “Forest fires under climate, social and economic changes in Europe,

the Mediterranean and other fire-affected areas of the World” (EC VII FP IP ENV.1.3.1.1: grant agreement no. 243888), and partially under the C.FORSAT project “Modeling the carbon sink in Italian forest ecosystems using ancillary data, remote sensing data and productivity models” (national coordinator: Prof. G. Chirici), funded by the FIRB2008 program of the Italian Ministry of University and Research (grant RBFR08LM04). The contribution of Alessandro Montagni to this manuscript was granted by the “Fondazione Cassa di Risparmio di Trento e Rovereto” within the project “Using remotely sensed LiDAR data to construct and assess forest attribute maps and related spatial products”.

We are grateful to Compagnia Generale Riprese Aeree/BLOM Italia for the ALS data, to Regione Sicilia for the use of the fuel types map, to Salford Systems (<http://www.salford-systems.com>) for the support in the use of the Salford Predictive Modeler, and to Martin Isenburg (<http://rapidlasso.com>) for supporting in the use of LASTOOLS.

References

- Ahokas, E., Kaartinen, H., Hyypää, J., 2003. A quality assessment of airborne laser scanner data. In: *Proceeding of the IAPRS*, vol. XXXIV-3/W13, Dresden, Deutschland, p. 7.
- Andersen, H.-E., McGaughey, R.J., Reutebuch, S.E., 2005. Estimating forest canopy fuel parameters using LiDAR data. *Remote Sensing of Environment* 94, 441–449.
- Anderson, H.E., 1982. Aids to determining fuel models for estimating fire behavior. In: *Rep. No. GTR INT-122*. USDA, Forest Service, Ogden, UT.
- Arroyo, L.A., Pascual, C., Manzanera, J.A., 2008. Fire models and methods to map fuel types: the role of remote sensing. *Forest Ecology and Management* 256, 1239–1252.
- Axelsson, P.E., 2000. DEM generation from laser scanner data using adaptive models. *International Archives of Photogrammetry and Remote Sensing* 33, 110–117.
- Axelsson, P.E., 1999. Processing of laser scanner data – algorithms and applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 54, 138–147.
- Barabesi, L., Franceschi, S., 2011. Sampling properties of spatial total estimators under tessellation stratified designs. *Environmetrics* 22, 271–278.
- Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Machine Learning* 36, 105–142.
- Benediktsson, J.A., Swain, P.H., Ersoy, O.K., 1990. Neural network approaches versus statistical-methods in classification of multisource remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing* 28, 540–552.
- Breiman, L., 1996. Technical note: some properties of splitting criteria. *Machine Learning* 24, 41–47.
- Breiman, L., 2001. Random Forests. *Machine Learning* 45 (1), 5–32, <http://dx.doi.org/10.1023/A:1010933404324>.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Chapman and Hall/Wadsworth, Inc., New York.
- Brivio, P.A., Petrucci, B., Boschetti, M., Carrara, P., Pepe, M., Rampini, A., Stroppiana, D., Zaffaron, P., 2009. A multi-year geographic database of fire affected areas derived from satellite images in the National Parks of Italy. *Italian Journal of Remote Sensing* 41, 65–78, <http://dx.doi.org/10.5721/ijRS20094125>.
- Burgan, R., Rothermel, R.C., 1984. BEHAVE: fire behaviour prediction and fuel modeling system-fuel subsystem. In: *Rep. No. GTR INT-167*. Intermountain Forest and Range Experiment Station, Ogden, UT.
- Carreiras, J.o.M.B., Vasconcelos, M.J., Lucas, R.M., 2012. Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sensing of Environment* 121, 426–442.
- Chaplot, V., Darboux, F., Bourennane, H., Leguédou, S., Silvera, N., Phachomphon, K., 2006. Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density. *Geomorphology* 77, 126–141.
- Chauve, A., Durrieu, S., Bretar, F., Pierrot-Deseilligny, M., Puech, W., 2007. Processing full-waveform LiDAR data to extract forest parameters and digital terrain model: validation in an alpine coniferous forest. In: *Proceedings of ForestSAT 2007*, Montpellier, France.
- Chuvieco, E., Cocero, D., Riaño, D., Martín, P., Martínez-Vega, J., Riva, J.d.I., Pérez, F., 2004. Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. *Remote Sensing of Environment* 92, 322–331.
- Chuvieco, E., Riaño, D., van Wagtenonk, J., Morsdorf, F., 2003. Fuel loads and fuel types. In: Chuvieco, E. (Ed.), *Wildland Fire Danger Estimation and Mapping*. The Role of Remote Sensing Data. World Scientific Publishing Co. Ltd., Singapore, pp. 1–32.
- Congalton, R.G., Green, K., 2008. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed. CRC Press, Taylor and Francis group.
- Corona, P., Cartisano, R., Salvati, R., Chirici, G., Floris, A., Di Martino, P., Marchetti, M., Scrinzi, G., Clementel, F., Travaglini, D., Torresan, C., 2012. Airborne laser scanning to support forest resource management under alpine, temperate and Mediterranean environments in Italy. *European Journal of Remote Sensing* 45, 27.
- Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Machine Learning* 40, 139–157.
- EC Forest fires in Europe 2010, 2011. Report 11. EUR 24910 EN. European Commission, Joint Research Centre, Institute for Environment and Sustainability, ISBN 978-92-79-20920-8, <http://dx.doi.org/10.2788/46294>, 98 pp.
- Erdody, T.L., Moskal, L.M., 2010. Fusion of LiDAR and imagery for estimating forest canopy fuels. *Remote Sensing of Environment* 114, 725–737.
- Estornell, J., Ruiz, L.A.B.V.-M., Fernández-Sarría, A., 2011. Estimation of shrub biomass by airborne LiDAR data in small forest stands. *Forest Ecology and Management* 262, 1697–1703.
- Evans, J.S., Hudak, A.T., Faux, R., Smith, A.M., 2009. Discrete return LiDAR in natural resources: recommendations for project planning. *Data processing, and deliverables*. *Remote Sensing* 1, 776–794.
- Falkowski, M.J., Evans, J.S., Martinuzzi, S., Gessler, P.E., Hudak, A.T., 2009. Characterizing forest succession with LiDAR data: an evaluation for the inland Northwest, USA. *Remote Sensing of Environment* 113, 946–956.
- Forest Europe, State of Europe's Forests 2011, 2011. Status and Trends in Sustainable Forest Management in Europe. FOREST EUROPE, UNECE and FAO, ISBN 978-82-92980-05-7.
- Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment* 61, 399–409.
- Friedman, J.H., 1999. Greedy function approximation: a gradient boosting machine. In: *Technical Report*. Department of Statistics, Stanford University.
- Friedman, J.H., 2002. Stochastic gradient boosting. *Computational Statistics and Data Analysis* 38, 367–378.
- García, M., Riaño, D., Chuvieco, E., Salas, J., Danson, F.M., 2011. Multispectral and LiDAR data fusion for fuel type mapping using Support Vector Machine and decision rules. *Remote Sensing of Environment* 115, 1369–1379.
- Guo, Q., Li, W., Yu, H., Alvarez, O., 2010. Effects of topographic variability and LiDAR sampling density on several DEM interpolation methods. *Photogrammetric Engineering and Remote Sensing* 76, 12.
- Harding, D., Lefsky, M., Parker, G., Blair, J., 2001. Laser altimetry canopy height profiles – methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment* 76, 283–297.
- Hardy, C., 2005. Wildland fire hazard and risk: problems, definitions, and context. *Forest Ecology and Management* 211, 73–82.
- Hofmann, A., Cibella, R., Bertani, R., Miozzo, M., Fantoni, I., Luppi, S. (Eds.), 2011. *Strumenti conoscitivi per la gestione delle risorse forestali in Sicilia*. Sistema Informativo Forestale Regionale. Regione Siciliana, Palermo, ISBN: 978-88-905700-0-1.
- Hug, C., Ullrich, A., Grimm, A., 2004. LiteMapper-5600 – a waveform-digitizing LiDAR terrain and vegetation mapping system. In: *Proceedings of the ISPRS Workshop on Laser Scanners for Forest and Landscape Assessment*, Freiberg, Germany, vol. 36, pp. 24–29.
- Hyypää, J., Schardt, M., Haggrén, H., Koch, B., Lohr, U., Scherrer, H., Paananen, R., Luukkonen, H., Ziegler, M., Hyypää, H., Pyysalo, U., Friedländer, H., Uuttera, J., Wagner, S., Inkinen, M., Wimmer, A., Kukko, A., Ahokas, A., Karjalainen, M., 2001. HIGH-SCAN: the first European-wide attempt to derive single-tree information from laser scanner data. *The Photogrammetric Journal of Finland* 17, 58–68.
- Keane, E., Burgan, R., van Wagtenonk, J., 2001. Mapping wildland fuels for fire management across multiple scales: integrating remote sensing, GIS, and biophysical modeling. *International Journal of Wildland Fire* 10, 301–319.
- Koetz, B., Morsdorf, F., van der Linden, S., Curt, T., Allgöwer, B., 2008. Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. *Forest Ecology and Management* 256, 263–271.
- Lawrence, R., Bunn, A., Powell, S., Zambon, M., 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment* 90, 331–336.
- Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., Harding, D., 1999. LiDAR remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment* 70, 339–361.
- Liaw, A., Wiener, M., 2002. Classification and regression by RandomForest. *R News* 2 (3), 18–22.
- Magnussen, S., Eggermont, P., Lariccia, V., 1999. Recovering tree heights from airborne laser scanner data. *Forest Science* 45, 407–422.
- Maltamo, M., Erikainen, K., Packalen, P., Hyypää, J., 2006. Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry* 79, 217–229.
- Mavsar, R., Varela, E., Corona, P., Barbati, A., Marsh, G., Chapter 3. Economic, legal and social aspects of post-fire management 2012. In: *Moreira, F., Arianoutsou, M., Corona, P., De las Heras, J. (Eds.), Post-Fire Management and Restoration of Southern European Forests*. Springer, New York/Dordrecht/Heidelberg/London, pp. 45–78, http://dx.doi.org/10.1007/978-94-007-2208-8_3, ISBN: 978-94-007-2207-1.
- Merkle, E.C., Shaffer, V.A., 2011. Binary recursive partitioning: background, methods, and application to psychology. *British Journal of Mathematical and Statistical Psychology* 64, 161–181, <http://dx.doi.org/10.1348/000711010X503129>.
- Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E., Edwards Jr., T.C., 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling* 199, 176–187.
- Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., Barbati, A., Corona, P., Vaz, P., Xanthopoulos, G., Mouillot, F., Bilgili, E., 2011. Landscape-wildfire interactions in southern Europe: implications for landscape management. *Journal of Environmental Management* 92, 2389–2402.

- Morsdorf, F., Meier, E., Kötz, B., Itten, K., Allgöwer, B., 2004. LiDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment* 92, 353–362.
- Mouillot, F., Rambal, S., Joffre, R., 2002. Simulating climate change impacts on fire frequency and vegetation dynamics in a Mediterranean-type ecosystem. *Global Change Biology* 8, 423–437.
- Muñoz, J., Felicísimo, A.M., 2004. Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science* 15, 285–292.
- Mutlu, M., Popescu, S., Stripling, C., Spencer, T., 2008. Mapping surface fuel models using LiDAR and multispectral data fusion for fire behavior. *Remote Sensing of Environment* 112, 274–285.
- Naeset, E., Økland, T., 2002. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment* 79, 105–115.
- Persson, A., Holmgren, J., Söderman, U., 2002. Detection and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing* 68, 925–932.
- Persson, A., Söderman, U., Topel, J., Ahlberg, S., 2005. Visualization and analysis of full-waveform airborne laser scanner data. In: *Proceedings of the ISPRS Workshop on Laser Scanning*, vol. 36, Enschede, The Netherlands, pp. 103–108.
- Pontius, R.G., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32, 4407–4429.
- Popescu, S.C., Wynne, R.H., Nelson, R.H., 2002. Estimating plot-level tree heights with LiDAR: local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture* 37, 71–95.
- Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne LiDAR. *Biomass and Bioenergy* 31, 646–655.
- Pyne, S.J., Andrews, P.L., Laven, R.D., 1996. *Introduction to Wildland Fire*. John Wiley and Sons, Inc.
- Riaño, D., Meier, E., Allgöwer, B., Chuvieco, E., Ustin, S., 2003. Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment* 86, 177–186.
- Riaño, D., Chuvieco, E., Condés, S., Gonzalez-Matezans, J., Ustin, S., 2004. Generation of crown bulk density for *Pinus sylvestris* L. from LiDAR. *Remote Sensing of Environment* 92, 345–352.
- Riaño, D., Chuvieco, E., Ustin, S., Salas, F., Rodriguez-Perez, J., Ribeiro, L., Viegas, D., Moreno, J., Fernandez, H., 2007. Estimation of shrub height for fuel type mapping combining airborne LiDAR and simultaneous colour infrared images. *International Journal of Wildland Fire* 16, 341–348.
- Seielstad, C., Queen, L., 2003. Using airborne laser altimetry to determine fuel models for estimating fire behaviour. *Journal of Forestry* 101, 10–15.
- Schmidt, K.M., Menakis, J.P., Hardy, C.C., Hann, W.J., Bunnell, D.L., 2002. Development of coarse-scale spatial data for wildland fire and fuel management. In: *General Technical Report RMRS-GTR-87*. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, pp. 41.
- Soininen, A., 2010. *TerraScan User's Guide*. Terra Solid Ltd.
- Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9, 307. <http://dx.doi.org/10.1186/1471-2105-9-307>.
- Tolliver, D., 2009. What is the variable importance measure? Available on line at <http://www.salford-systems.com/blog/company/what-is-the-variable-importance-measure.html> (accessed 17.04.12).
- Tesfamichael, S.G., Ahmed, F.B., Van Aardt, J.A.N., 2010. Investigating the impact of discrete-return LiDAR point density on estimations of mean and dominant plot-level tree height in *Eucalyptus grandis* plantations. *International Journal of Remote Sensing* 31, 2925–2940.
- Vepakomma, U., St-Onge, B., Kneeshaw, D., 2008. Spatially explicit characterization of boreal forest gap dynamics using multi-temporal LiDAR data. *Remote Sensing of Environment* 112, 2326–2340.
- Xanthopoulos, G., Calfapietra, C., Fernandes, P., 2012. Fire hazard and flammability of European Forest Types. In: *Moreira, F., Arianoutsou, M., Corona, P., De las Heras, J. (Eds.), Post-Fire Management and Restoration of Southern European Forests*, vol. 24. Springer-Managing Forest Ecosystems, Netherlands, pp. 93–120.
- Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., Viitala, R., 2010. Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS Journal of Photogrammetry and Remote Sensing* 66, 28–37.
- Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M., 2011. Characterizing forest canopy structure with LiDAR composite metrics and machine learning. *Remote Sensing of Environment* 115, 1978–1996.