

Document downloaded from:

<http://hdl.handle.net/10251/33190>

This paper must be cited as:

Estornell Cremades, J.; Ruiz Fernández, LÁ.; Velázquez Martí, B.; Hermosilla, T. (2012). Estimation of biomass and volume of shrub vegetation using LiDAR and spectral data in a Mediterranean environment. *Biomass and Bioenergy*. 46:710-721. doi:10.1016/j.biombioe.2012.06.023.



The final publication is available at

<http://dx.doi.org/10.1016/j.biombioe.2012.06.023>

Copyright Elsevier

**ESTIMATION OF BIOMASS AND VOLUME OF SHRUB VEGETATION  
USING LiDAR AND SPECTRAL DATA IN A MEDITERRANEAN  
ENVIRONMENT**

J Estornell<sup>a</sup>, L A Ruiz<sup>a</sup>, B Velázquez-Martí<sup>b</sup>, and T Hermosilla<sup>a</sup>

<sup>a</sup>Geo-Environmental Cartography and Remote Sensing Group (CGAT), Department of  
Cartographic Engineering, Geodesy and Photogrammetry. Universitat Politècnica de  
València. Camino de Vera s/n 46022 Valencia (Spain)

corresponding author: [jaescr@cgf.upv.es](mailto:jaescr@cgf.upv.es) Tel.:+34962849324; fax.: +34962849309

[laruiz@cgf.upv.es](mailto:laruiz@cgf.upv.es)

[txohergo@topo.upv.es](mailto:txohergo@topo.upv.es)

<sup>b</sup>Departamento de Ingeniería Rural y Agroalimentaria.

Universitat Politècnica de València. Camino de Vera s-n 46022 Valencia (Spain)

[borvemar@dmata.upv.es](mailto:borvemar@dmata.upv.es)

**Abstract**

Several studies have addressed the biomass and volume of trees using Airborne Light Detection and Ranging (LiDAR) data. However, little research has been conducted into shrub vegetation, which covers a high percentage of Mediterranean forest. We used LiDAR data and an airborne image to estimate biomass and volume of shrub vegetation. Field data were collected in 29 square plots of 100 m<sup>2</sup>. In each plot, the percentage of the surface covered was measured in the field. Shrub vegetation inside 3 stands for each plot was clear cut to calculate the biomass and volume of the 29 plots. To find the best fit between LiDAR-spectral data and field measurements, stepwise regressions were performed using previously selected variables. The highest accuracy was found when

variables derived from LiDAR data and the airborne image were combined ( $R^2$  values of 0.78 and 0.84 for biomass and volume, respectively). Biomass and volume were predicted using variables from height metrics of LiDAR data (median and standard deviation); density metrics (percentage of points whose height was between 0.50 m to 0.75 m, 0.75 m to 1 m, and higher than 1 m); and spectral data (standard deviation of green band, mean of the vegetation index NDVI). These results revealed the potential of LiDAR and spectral data to characterize shrub structure and make it possible to estimate and map the biomass and volume of this vegetation.

*Keywords:* LiDAR, spectral information, biomass, volume, shrub.

## **1. Introduction**

A high percentage of Mediterranean forest consists of dense, low shrub. The difficulty involved in shrub management and the lack of information about shrub behavior means that these areas are often left out of spatial planning projects. Nevertheless, these shrub ecosystems are important for the environment and landscape because they prevent soil erosion and desertification [1]; contribute to managing the wildlife habitat [2]; contribute to creating fuel-type maps for better accuracy in fire behavior modeling [3]; represent important CO<sub>2</sub> sinks; and help refill aquifers [4,5]. For management of forest areas, biomass is a key variable as it allows the function and structure of the ecosystem across the landscape to be evaluated [6-8]; and it is a potential source of renewable energy [9,10].

LiDAR has been used successfully in the last decade in forest applications [11-13]. In applications related to forest inventories, two approaches can be followed [14]: (i) estimation of the dasometric variables using regressions between field data and statistics

derived from LiDAR data in a plot or stand [15-18]; and (ii) calculation of the dendrometric variables needed to delineate the crown of a tree [10,19,20].

Little research has been conducted into shrub areas because of the inherent difficulty: this vegetation is low and occupies a continuous surface in which individual elements cannot be defined. Therefore, only the plot-stand approach can be applied. Moreover, their low height requires high accuracy levels in methodology and LiDAR data characteristics. Sometimes, buffers are used when vegetation heights measured in the field are compared to those calculated from LiDAR data to find the maximum correlations between both data [21]. A threshold is usually applied to LiDAR data in forestry applications to eliminate points close to the ground [22-24]. This value represents the maximum elevation to consider in shrub studies.

Most studies of shrub vegetation using LiDAR data have been based on height estimation and presence/absence of shrub vegetation [3,21]. As with trees, shrub height is underestimated using LiDAR data [25-27]. In relative terms, this underestimation is more important for shrub vegetation as it is lower in height [28-30]. However, less attention has been paid to biomass and volume estimation, which are key variables for forestry applications. One of the reasons for this is the difficulty of measuring the variables to be correlated with LiDAR data in the field. Without this data, the parameters extracted from LiDAR data cannot be correlated. Another reason is the lack of knowledge of this type of vegetation, unlike trees, where allometric equations are defined to estimate dendrometric variables from simple measurements. In studies based on mapping the presence or absence of shrub vegetation, good results were obtained when only LiDAR data was used [31]. In addition, the detection of this vegetation was improved when LiDAR data and spectral information were combined [2,32,33]. Other

studies successfully used this data combination to produce a classified vegetation map [26,34]. In forest applications, it was reported that when these data were all used together the prediction models improved significantly [35-38]. Therefore, it is a specific issue to analyze whether this combination of data can improve the prediction models for biomass and volume of shrub vegetation.

The aim of this paper is to study the possibility of estimating key variables for biomass and volume of shrub areas, which cover a high percentage of Mediterranean forest, by analyzing the combination of data which produces the highest accuracy in the predicted models. The regression equations can be applied to map the volume and biomass of the study area [39,40]. The results presented here illustrate the feasibility of estimating these variables.

## **2. Materials and methods**

### **2.1 Study area**

The 10 km<sup>2</sup> study area is located in Chiva (Valencia, Spain), defined by a rectangle whose UTM coordinates  $X_{\text{maximum}}$ ,  $Y_{\text{maximum}}$ ,  $X_{\text{minimum}}$ , and  $Y_{\text{minimum}}$ , are 689800 m, 4376028 m, 683800 m, and 4373000 m, respectively (Fig. 1). The area is located in zone 30N in the European Datum 1950 reference system. It is a mountainous area with a predominance of *Quercus coccifera* (Fig. 2), although other species can be found such as *Rosmarinus officinalis*, *Ulex parviflorus*, *Cistus albidus L.* and *Erica multiflora L.* These species are the most abundant in Mediterranean forests. The average occupation fraction of shrub vegetation is around 0.55. In contrast, there is a sub-area with trees occupying a surface fraction of 0.1. The altitude varies from 442 m to 1000 m, and the average slope fraction is 0.45.

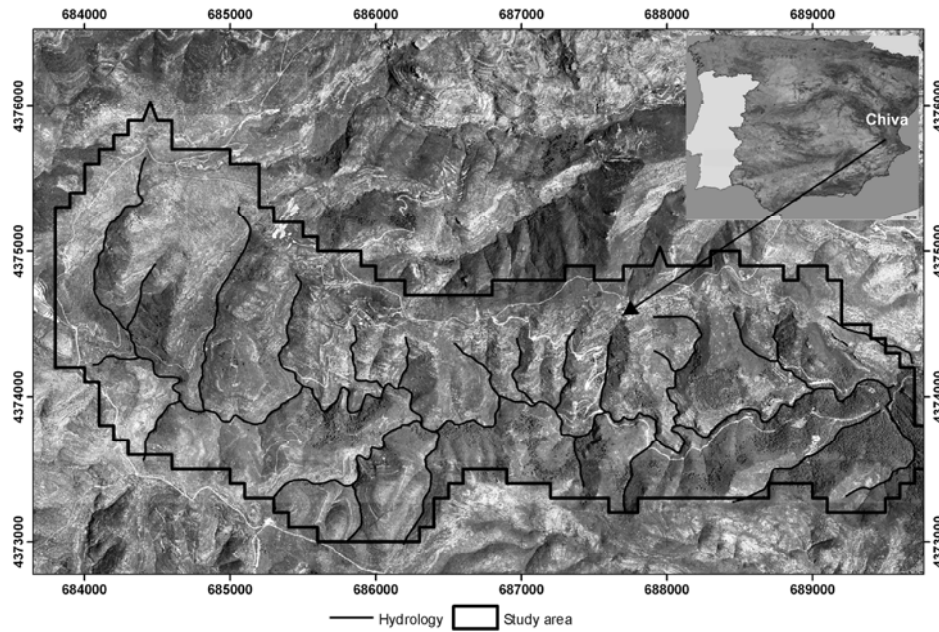


Fig. 1. Location of study area in Chiva (Spain). The black polygon represents area surveyed for LiDAR



Fig. 2. Study area photograph with dense presence of kermes oak (*Q. coccifera*).

## 2.2 LiDAR and spectral data

The LiDAR data were acquired during a flight in December 2007, using an Optech ALTM 2050 system. The technical parameters were: flight height – 700 m above

ground; pulse frequency – 50 kHz; scan frequency – 47 Hz; and scan angle –  $\pm 18^\circ$ ; pulse density points –  $4 \text{ m}^{-2}$ . However, given that 10 overlapping flight lines were registered, some areas had a higher point density. For this reason, the average point density of the study area considering all the returns was  $8 \text{ m}^{-2}$ . The number of echoes was 2. A fraction around 99 % of the LiDAR data belonged to the first pulse. This could be caused by the small differences between the canopy of vegetation and the ground.

A spectral airborne image registered by the Ultracam D, made by Vexcel Imaging GMBH, was used with a cell size of 0.5 m. This image was acquired in July 2006, and it contained three spectral bands: infrared, red, and green.

### 2.3 Field data

To assess the digital terrain model (DTM) calculated, 1379 ground-surveyed checkpoints throughout the study area were measured using a RTK-GPS system (Leica System 1200). To obtain field values of biomass and volume to be correlated to the statistics derived from the LiDAR data, two classes of analysis were performed. Firstly, a dendrometric analysis was carried out to determine the actual volume and biomass of a whole plant from measurable variables, such as the base diameter, height, and weight of the plant. Next, a dasometric analysis was performed to calculate the biomass and volume occupied by vegetation in a plot. This analysis used measurements taken from individual shrubs, and the percentage occupation of each shrub species in a plot.

For the dendrometric analysis (calculation of volume of whole plants), samples of 100-150 plants of each species were analyzed from 29 square plots of  $100 \text{ m}^2$ , located in different bioclimatic layers (elevation), slope, and aspect. The plots were distributed randomly among the cells of a grid defined in the studied area. The size of the cell was 1

km<sup>2</sup> and at least one plot with vegetation in each cell was defined. In order to determine the real volume of a whole plant from measurable variables, such as the base diameter of the plant and its height, a global form factor was defined [5]. This factor is calculated as a quotient of the real volume of the plant and the volume of a geometric model taken as reference (Eq. 1).

$$f = \frac{\text{Real volume of the plant}}{\text{Model volume of the plant}} \quad (1)$$

The real volume of each plant was obtained by submerging it in water and determining the volume displaced. As a model volume, a solid of revolution (cylinder, paraboloid, cone or neiloid) can be used. The cylinder was the solid of revolution best suited to the species studied [5]. Therefore, after calculating the *form factor*,  $f$ , and with the field measurements of the plants, such as the stem diameter and height, we were able to calculate the volume of each plant using Eq. 2 (below), where  $V_i$  is the real volume of the whole plant,  $d$  is the base diameter of the main stem,  $h$  the height of the individual plant measured for each plant in the sample group, and  $f$  the form factor. To calculate dry biomass, we considered the weight of each individual both after cutting and after applying a drying process. Once the materials were dry, their dry biomass was determined. The wet biomass of a plant was its weight measured after cutting only.

$$V_i = \frac{\pi \cdot d^2}{4} \cdot h \cdot f \quad (2)$$

The dasometric analysis (calculation of the real volume of stands and plots) was carried out on the same 29 plots cited above. In each plot, the percentage of the surface covered by each species was measured in the field. The vegetation inside 3 stands was clear cut.



Each stand was circular and one meter in diameter. In each clear cutting, we counted the number of shrub plants inside the stand, identified the species, and measured the weight, length, and base diameter of stems for each plant. The biomass and real volume occupied by each species in each cleared stand was then calculated. The real volume of each plant was the result of measuring the diameter and height, and using Eq 2. After counting the number of plants of each species in a stand, we calculated their percentage and the sum of the volume in a stand. Knowing the area of the stand, we were then able to calculate the volume of each species per square meter in the stand by applying the percentages of occupation calculated previously. This result was multiplied by the area occupied by each species in a plot, thus giving the total volume of vegetation in a plot. For biomass, 87 stands were clear cut determining the wet weight of biomass by means of a dynamometer. Then, some samples were put into a plastic container to take them to laboratory without moisture content losses. Determination of moisture content was done according to the norm UNE-EN 14774-2. Samples were placed on metal trays and located in a stove with controlled temperature ( $105 \pm 2$ ) °C. The drying time did not exceed 24 h in order to avoid possible unnecessary loss of volatile substances. For sample preparation and result registration an electronic balance was used with the precision of 0.001 g. To obtain the dry biomass in each stand the wet weight was corrected removing the weight of water content. The statistics of field biomass and volume are summarized in Table 1.

Table 1. Descriptive statistics of field data for the 29 plots.

Statistic	Wet Biomass (kg)	Dry Biomass (kg)	Volume (m <sup>3</sup> )
Mean	659	421	0.48
Minimum	245	166	0.16
Maximum	1233	821	1.08
Standard deviation	280	185	0.22

## 2.4 DTM calculation

To compute the DTM from LiDAR data, it is necessary to apply algorithms to eliminate points belonging to any object above the ground surface such as vegetation or buildings. To achieve this, we generated an IDL program based on iterative processes [22,41]. As input data type, we used rasterized point data with a cell size of 1 m<sup>2</sup> and selected the lowest LiDAR point for each cell. Further information can be found in [42] The DTM iterative algorithm had five steps:

**Step 1:** the study area was divided into windows with an initial window size ( $v1$ ). In each analysis window, the lowest elevation point was selected. With these points, an initial DTM (DTM1) was calculated by applying the Delaunay triangulation method.

**Step 2:** A smaller analysis window ( $v2$ ) was selected to find new minimum heights from the input data.

**Step 3:** We compared the points selected in step 2 with DTM1 (calculated in step 1), and selected those that were lower than a defined height threshold ( $u1$ ). Points with differences larger than this threshold were rejected. A new DTM (DTM2) was then determined with the selected points.

**Step 4:** A window size ( $v3$ ), smaller than  $v2$ , was selected and the minimum height in each window was selected.

**Step 5:** As in step 3, points of minimum height with a difference compared to DTM2 greater than a second threshold ( $u2$ ) were eliminated. The final DTM was calculated with the remaining points.

## 2.5 Statistics derived from LiDAR and spectral data

To estimate the biomass and volume of shrub vegetation, statistics derived from all LiDAR point cloud and spectral data were obtained and used as explanatory variables in the regression models. For LiDAR data, the bare-earth surface elevation was first subtracted from each LiDAR point by using the DTM calculated. Then, for each plot the following statistics (LiDAR height metrics) were obtained: maximum height, mean, standard deviation, coefficient of variation, kurtosis, skewness, interquartile distance, percentile values (5th, 10th, 20th, 25th, 30th, 40th, 50th, 60th, 70th, 75th, 80th, 90th, 95th).

Given that variables derived from the point distribution in height (density metrics) have been used successfully in some forestry applications [16,17,37], we calculated the percentage of points belonging to the following height intervals to be considered in the biomass and volume models of shrub vegetation: 0.25 m to 0.5 m, 0.5 m to 0.75 m, 0.75 to 1 m, and the fraction of points higher than 1 m.

For the airborne image, the following statistics were calculated: minimum, maximum, mean, and standard deviation of the cell values within the plots for the bands green (G), red (R), infrared (IR), and NDVI.

The large set of potential predictor variables calculated in the above section (24 variables calculated from LiDAR point cloud and 16 variables calculated from spectral information) would make it hard to explain their physical response. Therefore, we needed to reduce the set, and also avoid the selection of highly correlated predictor

variables. Thus, we calculated a correlation matrix and grouped variables using cluster analysis.

## 2.6 Estimation of biomass and volume

To estimate biomass (wet and dry) and volume, we performed a stepwise regression analysis with the variables selected from the above section. We used two types of data in this study, and tested all the possible combinations of these. Firstly, only variables derived from LiDAR data were selected. In a second step, we used only spectral variables derived from the airborne image. Finally, the biomass and volume were estimated combining LiDAR data and the statistics derived from the spectral variables. The models obtained were compared by calculating RMSE and  $R^2$ .

The models with the highest coefficients of determination and the lowest RMSE were selected for mapping the biomass of the study area. Initially, the study area was divided into cells of 100 m<sup>2</sup>, the same size as the plots, in which the value of each explanatory variable was calculated. Then, we applied the predicted model to estimate biomass and volume in each cell.

## 3. Results and discussion

### 3.1 DTM

In this study, a DTM was calculated to subtract the bare-earth surface elevation from each LiDAR point elevation. The lowest RMSE was for analysis window sizes of 10 m, 5 m and 2.5 m ( $v1$ ,  $v2$  and  $v3$ ) and height thresholds equal to, or greater than, 1.5 m. These parameters produced a DTM with a mean signed error of 0.02 m, a standard deviation of 0.19 m, and an RMSE of 0.19 m.

### 3.2 Pre-selection of LiDAR and spectral variables

There was a high correlation between the mean height and the percentile values, the lowest being for the 5th and 10th percentiles (Table 2). The values of maximum, kurtosis, skewness, and coefficient of variation showed the lowest correlation of all the variables. Standard deviation also showed a low correlation with these values and with percentiles 5 through to 50. These results were confirmed in the dendrogram (Fig. 3) obtained from the cluster analysis. Fig. 3a reveals how four groups of variables can be established in the first steps of clustering. Firstly, maximum, kurtosis, skewness and coefficient of variation defined independent groups. Secondly, two additional groups were defined. In the first group the following variables were included: 60th, 70th, 75th, 80th, 90th, 95th, mean, standard deviation, and interquartile distance. The second group was defined by the followed variables: 5th, 10th, 20th, 25th, 30th, 40th, and 50th. According to these results, we considered maximum, kurtosis, skewness, and coefficient of variation as explanatory variables, and we chose two variables from the two additional groups which showed a low correlation among the rest of the variables. All possible combinations were tested.

Table 2. Correlation matrix of variables derived of LiDAR height metrics

Statistic	Max	Mean	Std	cv	Inter	Skew	Kur	P05	P10	P20	P25	P30	P40	P50	P60	P70	P75	P80	P90	P95
Max	1.00	0.71	0.76	-0.54	0.61	-0.39	0.51	0.70	0.62	0.58	0.56	0.55	0.56	0.59	0.63	0.57	0.70	0.71	0.74	0.77
Mean	0.71	1.00	0.83	-0.38	0.76	-0.27	-0.10	0.79	0.84	0.88	0.90	0.91	0.94	0.96	0.98	0.99	0.98	0.97	0.92	0.89
Std	0.76	0.83	1.00	0.17	0.95	0.09	-0.05	0.51	0.46	0.49	0.51	0.52	0.59	0.65	0.74	0.84	0.89	0.92	0.97	0.98
Inter	0.61	0.76	0.95	0.16	1.00	-0.08	-0.25	0.35	0.32	0.37	0.41	0.43	0.52	0.61	0.71	0.82	0.86	0.89	0.93	0.92
Skew	0.39	-0.27	0.09	0.61	-0.08	1.00	0.78	-0.02	-0.23	-0.37	-0.41	-0.41	-0.44	-0.43	-0.40	-0.32	-0.27	-0.23	-0.11	0.00
Kur	0.51	-0.10	-0.05	0.10	-0.25	0.78	1.00	0.23	0.12	0.00	-0.03	-0.05	-0.11	-0.15	-0.17	-0.18	-0.18	-0.17	-0.14	-0.07
P05	0.56	0.80	0.81	-0.68	0.41	-0.39	-0.18	0.85	0.96	1.00	1.00	1.00	0.98	0.96	0.92	0.86	0.82	0.77	0.68	0.63
P10	0.70	0.88	0.89	-0.28	0.86	-0.27	-0.18	0.70	0.74	0.79	0.82	0.83	0.88	0.92	0.96	0.99	1.00	1.00	0.96	0.93
P20	0.70	0.79	0.81	-0.53	0.35	-0.02	0.23	1.00	0.95	0.97	0.85	0.83	0.81	0.77	0.75	0.72	0.70	0.67	0.62	0.61
P30	0.62	0.84	0.46	-0.66	0.32	-0.23	0.12	0.95	1.00	0.97	0.96	0.95	0.92	0.88	0.84	0.77	0.74	0.70	0.61	0.58
P40	0.56	0.88	0.49	-0.69	0.37	-0.37	0.00	0.87	0.97	1.00	1.00	0.99	0.97	0.94	0.90	0.83	0.79	0.75	0.65	0.60
P50	0.55	0.91	0.52	-0.68	0.43	-0.41	-0.05	0.83	0.95	0.99	1.00	1.00	0.99	0.97	0.93	0.87	0.83	0.79	0.68	0.63
P60	0.56	0.94	0.59	-0.64	0.52	-0.44	-0.11	0.81	0.92	0.97	0.98	0.99	1.00	0.99	0.97	0.92	0.88	0.84	0.74	0.69
P70	0.59	0.86	0.65	-0.58	0.61	-0.43	-0.15	0.77	0.88	0.94	0.95	0.97	0.99	1.00	0.99	0.95	0.92	0.89	0.80	0.74
P80	0.63	0.98	0.74	-0.49	0.71	-0.40	-0.17	0.75	0.84	0.90	0.92	0.93	0.97	0.99	1.00	0.98	0.96	0.94	0.87	0.81
P90	0.67	0.99	0.84	-0.35	0.82	-0.32	-0.18	0.72	0.77	0.83	0.86	0.87	0.92	0.95	0.98	1.00	0.99	0.98	0.94	0.90
P95	0.71	0.97	0.92	-0.20	0.89	-0.23	-0.17	0.67	0.70	0.75	0.77	0.79	0.84	0.89	0.94	0.98	1.00	1.00	0.98	0.95
CV	0.74	0.92	0.97	-0.05	0.93	-0.11	-0.14	0.62	0.61	0.65	0.68	0.68	0.74	0.80	0.87	0.94	0.96	0.98	1.00	0.99
	0.77	0.89	0.98	0.04	0.92	0.00	-0.07	0.61	0.58	0.60	0.63	0.63	0.69	0.74	0.81	0.90	0.93	0.95	0.99	1.00
	-0.04	-0.38	0.17	1.00	0.16	0.61	0.10	-0.53	-0.66	-0.69	-0.68	-0.68	-0.64	-0.58	-0.49	-0.35	-0.28	-0.20	-0.05	0.04

As with the LiDAR data, we defined three groups of variables for spectral information (Fig. 3b). The first group included: minimum values of the IR, R, and G, and mean value of IR; the second group: maximum values and standard deviation of IR, R, and G bands, mean of the R and B, and standard deviation of the NDVI. The third group

included: the minimum, maximum, and mean of the NDVI band. High correlation among the variables of each group was observed in the correlation matrix (Table 3). We selected one variable from each group. In addition to the spectral and height variables derived from the LiDAR point cloud and airborne image, we considered the variables related to the canopy point density: 0.25 m to 0.5 m, 0.5 m to 0.75 m, 0.75 m to 1 m, and the percentage of points higher than 1 m.

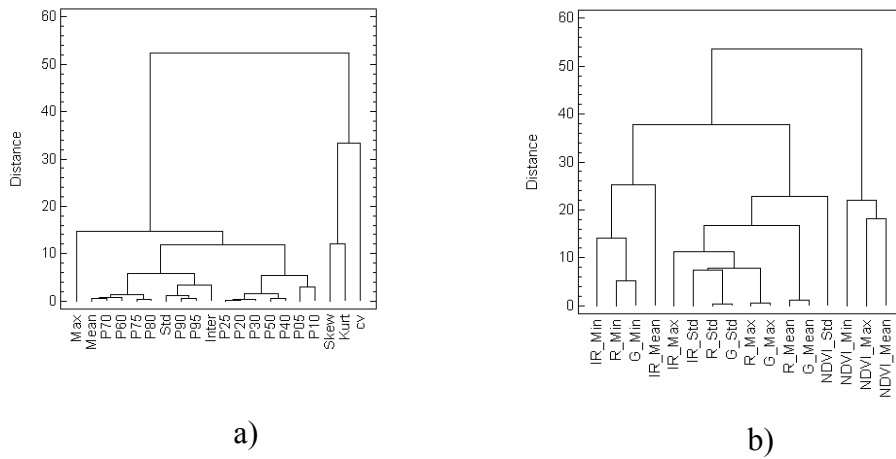


Fig 3. Dendrogram of variables derived from LiDAR height metrics (a) and of the spectral image (b).

Table 3. Correlation matrix of spectral variables

Statistic	R_Min	R_Max	R_Mean	R_Std	R_Min	R_Max	R_Mean	R_Std	G_Min	G_Max	G_Mean	G_Std	NDVI_Min	NDVI_Max	NDVI_Mean	NDVI_Std
R_Min	1.00	-0.07	0.53	-0.69	0.63	-0.26	-0.12	-0.38	0.82	-0.25	-0.07	-0.38	0.33	-0.22	0.35	-0.37
R_Max	-0.07	1.00	0.44	0.60	0.17	0.85	0.36	0.70	0.02	0.86	0.34	0.72	-0.65	-0.20	-0.10	0.50
R_Mean	0.53	0.44	1.00	0.09	0.44	0.30	0.36	0.42	0.46	0.32	0.47	0.42	-0.14	-0.07	0.19	0.34
R_Std	-0.69	0.60	0.09	1.00	-0.19	0.76	0.63	0.86	-0.49	0.76	0.61	0.88	-0.66	-0.17	-0.54	0.49
R_Min	0.63	0.17	0.44	-0.19	1.00	0.27	0.47	0.13	0.91	0.27	0.46	0.14	-0.29	-0.78	-0.30	-0.12
R_Max	-0.26	0.85	0.30	0.76	0.27	1.00	0.68	0.86	0.00	0.99	0.65	0.87	-0.88	-0.43	-0.51	0.51
R_Mean	-0.12	0.36	0.36	0.63	0.47	0.68	1.00	0.71	0.20	0.67	0.98	0.73	-0.70	-0.65	-0.84	0.22
R_Std	-0.38	0.70	0.42	0.86	0.13	0.86	0.71	1.00	-0.14	0.85	0.71	0.99	-0.77	-0.25	-0.44	0.72
G_Min	0.82	0.02	0.46	-0.49	0.91	0.00	0.20	-0.14	1.00	0.02	0.22	-0.14	-0.03	-0.56	-0.02	-0.22
G_Max	-0.25	0.86	0.32	0.76	0.27	0.99	0.67	0.85	0.02	1.00	0.65	0.87	-0.86	-0.42	-0.49	0.50
G_Mean	-0.07	0.34	0.47	0.61	0.46	0.65	0.98	0.71	0.22	0.65	1.00	0.74	-0.66	-0.58	-0.76	0.25
G_Std	-0.38	0.72	0.42	0.88	0.14	0.87	0.73	0.99	-0.14	0.87	0.74	1.00	-0.77	-0.27	-0.47	0.67
NDVI_Min	0.33	-0.65	-0.14	-0.66	-0.29	-0.88	-0.70	-0.77	-0.03	-0.86	-0.66	-0.77	1.00	0.42	0.64	-0.54
NDVI_Max	-0.22	-0.20	-0.07	-0.17	-0.78	-0.43	-0.65	-0.25	-0.56	-0.42	-0.58	-0.27	0.42	1.00	0.68	0.25
NDVI_Mean	0.35	-0.10	0.19	-0.54	-0.30	-0.51	-0.84	-0.44	-0.02	-0.49	-0.76	-0.47	0.64	0.68	1.00	0.02
NDVI_Std	-0.37	0.50	0.34	0.49	-0.12	0.51	0.22	0.72	-0.22	0.50	0.25	0.67	-0.54	0.25	0.02	1.00

### 3.3 Estimation of Biomass

The results of the stepwise regressions for estimating dry biomass are shown in Table 4 for each group of the explanatory variables selected. Namely: statistics derived only

from LiDAR data; metrics derived from the spectral variables; metrics derived from merging both LiDAR data and spectral variables. The model which gave the lowest  $R^2$  value (0.34) was the one which used only spectral data, which indicated the low explanatory power of spectral data for predicting shrub biomass. This finding is in line with previous findings on the low ability of this data to describe canopy height variations and the vertical distribution of canopy elements [29,43]. In contrast, spectral images can be used to map forest species composition [26]. We achieved better results for models in which LiDAR metrics were used, as evidenced by the explained variance ( $R^2 = 0.67$ ). Previous research has demonstrated the ability of LiDAR data to predict biomass for several forest types: Douglas-fir, western hemlock, dry ponderosa pine, and birch and spruce forest [18]. Good results were also obtained in forests whose main species were black pine, Spanish juniper, and Holm oak [38].

Table 4. Estimation of dry biomass in plots

Input data	Model	$R^2$	RMSE (kg)
LiDAR	$B_d = -656.5 + 3067.4 H\_Std + 3102.9* H\_Median - 47.1 p_1 - 18.7 p_{0.75-1} - 29.2 p_{0.5-0.75}$	0.67	118.2
Airborne image	$B_d = 783.5 + 935.5 NDVI\_Max - 3.2 IR\_Max$	0.34	157.1
LiDAR and airborne image	$B_d = -371.5 + 3200.4 H\_Std + 2919.1 H\_Median - 47.5 p_1 - 17.8 p_{0.75-1} - 32.9 p_{0.5-0.75} - 7.4 G\_Std$	0.79	96.6

$p < 0.05$

$B_d$  dry biomass in kg in 100 m<sup>2</sup> plots; Independent variables derived from LiDAR data by plots: Median of the heights ( $H\_Median$ ) and Standard deviation of the heights ( $H\_Std$ ); variables derived from the point distribution in height by plots (density metrics): percentage of points in a plot whose height is above 1 m ( $p_1$ ), percentage of points in a plot whose height is between 0.75 m to 1 m ( $p_{0.75-1}$ ); percentage of points in a plot whose height is between 0.50 m to 0.75 m ( $p_{0.5-0.75}$ ); Variables derived from the airborne image by plots: Standard deviation of the green band ( $G\_Std$ ), maximum value of the infrared band ( $IR\_Max$ ), maximum value of the NDVI image ( $NDVI\_Max$ ).

The best model was obtained when we combined the metrics derived from the LiDAR data and the airborne image, which gave  $R^2$  and RMSE values of 0.79 and 96.55 kg, respectively. LiDAR predicted versus field-measured biomass showed a good linear relationship close to the 1:1 line (Fig. 4). Although LiDAR metrics explained the majority of the variability in this model, biomass prediction fraction improved by 1 % when spectral data was added.

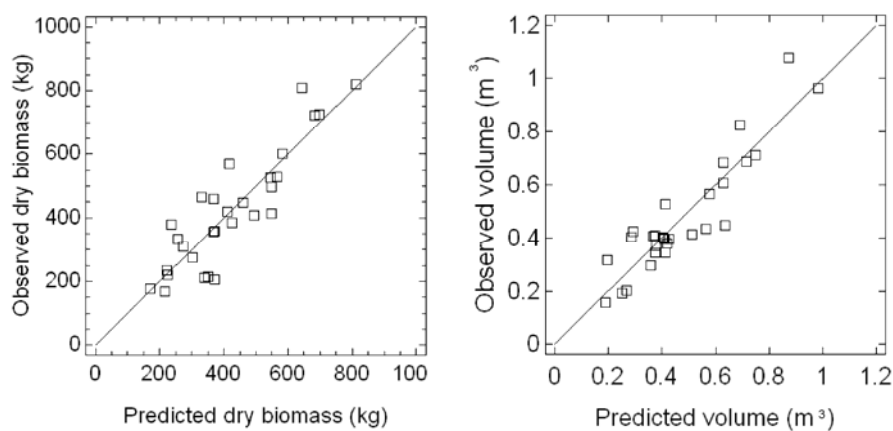
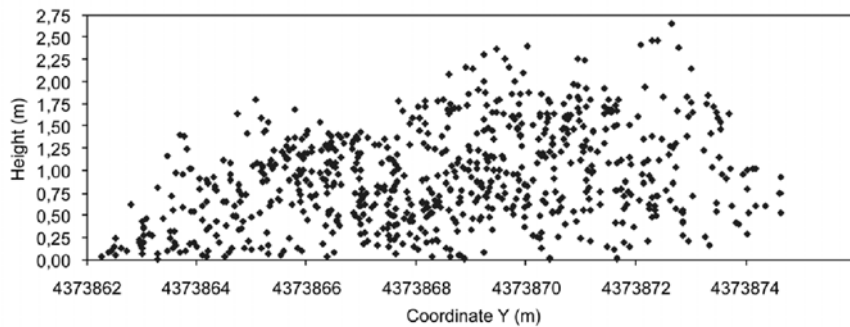


Fig. 4. Scatterplots of predicted biomass versus observed biomass (a) and predicted volume versus observed volume (b).

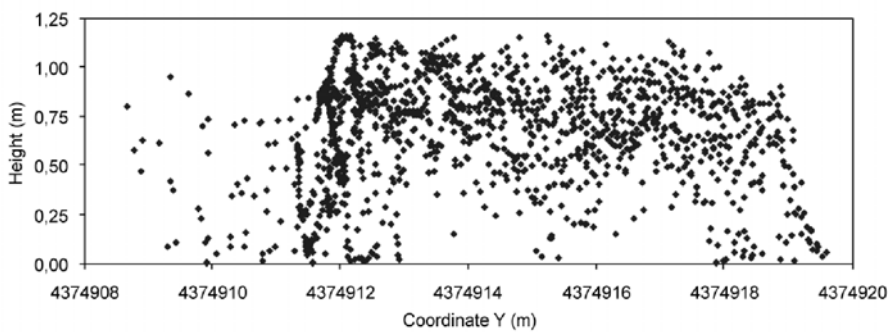
An improvement in the prediction of forest variables when both data are combined was also found in other tree studies [35, 44, 45]. This stepwise model included six significant variables, the most important of which in terms of the contribution in the explicative variation of biomass was standard deviation of heights ( $H\_Std$ ). This variable characterizes the canopy structure [38] and describes the variability of the vegetation heights in a plot. In our study we observed that the higher the standard deviation, the more biomass was estimated. Here, we must emphasise the importance of the density metrics for biomass prediction. This finding was in line with other forest



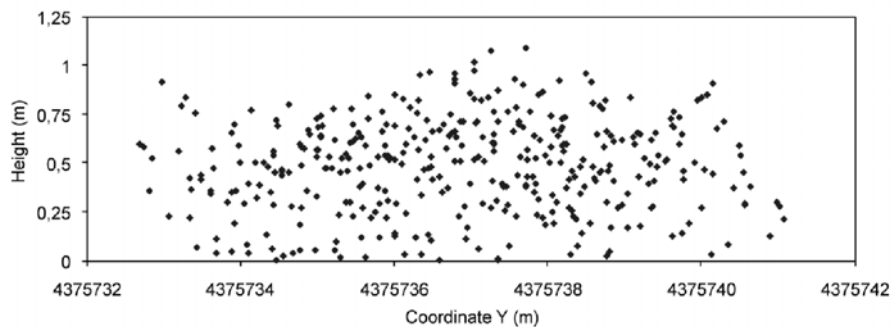
studies in which variables derived from density metrics were used for predicting biomass variation [16,18,37]. We observed that when density metrics were removed from the stepwise model, the value of  $R^2$  decreased to 0.57. Fig. 5 shows the distribution of the points for the plots with the maximum percentage of heights greater than 1 m (5a); with the maximum percentage of heights in the interval 0.75 m to 1 m (5b); and with the maximum percentage of heights in the interval 0.5 m to 0.75 m (5c). As expected, the biomass of the first plot was larger than the rest of the plots. The plot with the lowest percentage of points in the intervals 1 m to 2.75 m and 0.75 m to 1 m showed the lowest biomass.



a)



b)



c)

Fig. 5. Plot with the highest percentage of points with height greater than 1 m (a); plot with the highest percentage of points in the interval 0.75 m to 1 m (b); plot with the highest percentage of points in the interval 0.50 m to 0.75 m (c).

Only one variable derived from spectral data was used: standard deviation of the green band ( $G\_Std$ ). The results of this were in line with [44]. Fig. 6 shows the relationship between the value of the  $G\_Std$  and the biomass in two plots with maximum and minimum presence of vegetation. The upper plot showed a high presence of vegetation. This was indicated by the red colour of the plot and the narrow range of low digital values for the histogram of the green band. In contrast, the bottom plot showed a minimum presence of vegetation. In this case, the histogram of the green band revealed two concentrations of digital values: the first one, with low digital values, meant presence of vegetation; the second one, with a wider range of higher digital values, was related with the soil spectral response (the white colour in the image). In this case, the  $G\_Std$  value was higher. The plot with lower  $G\_Std$  value showed more presence of shrub vegetation (more biomass). This variable in our model contributed to explaining the differences between areas with more and less vegetation as well as the differences between vegetation areas and the soil. Although the dominant species for the plots was *Q. coccifera*, there were some plots with presence of *C. albidus L.* We found that in plots where the presence of this vegetation was important, the  $G\_Std$  value was higher than for the plots with predominance of *Q. coccifera* with a higher value of biomass. This result may be because the *C. albidus L.* species is characterized by an open canopy and therefore the influence of the soil response is higher.

In addition to the models shown, we proved that when the standard deviation and median of the heights were replaced by the mean and the 10th percentile,  $R^2$  and RMSE values of 0.77 and 101.1 kg were obtained, respectively. The selection of mean height is congruous with earlier work indicating that mean height of LiDAR data is a strong predictive variable in tree biomass studies [18,24].

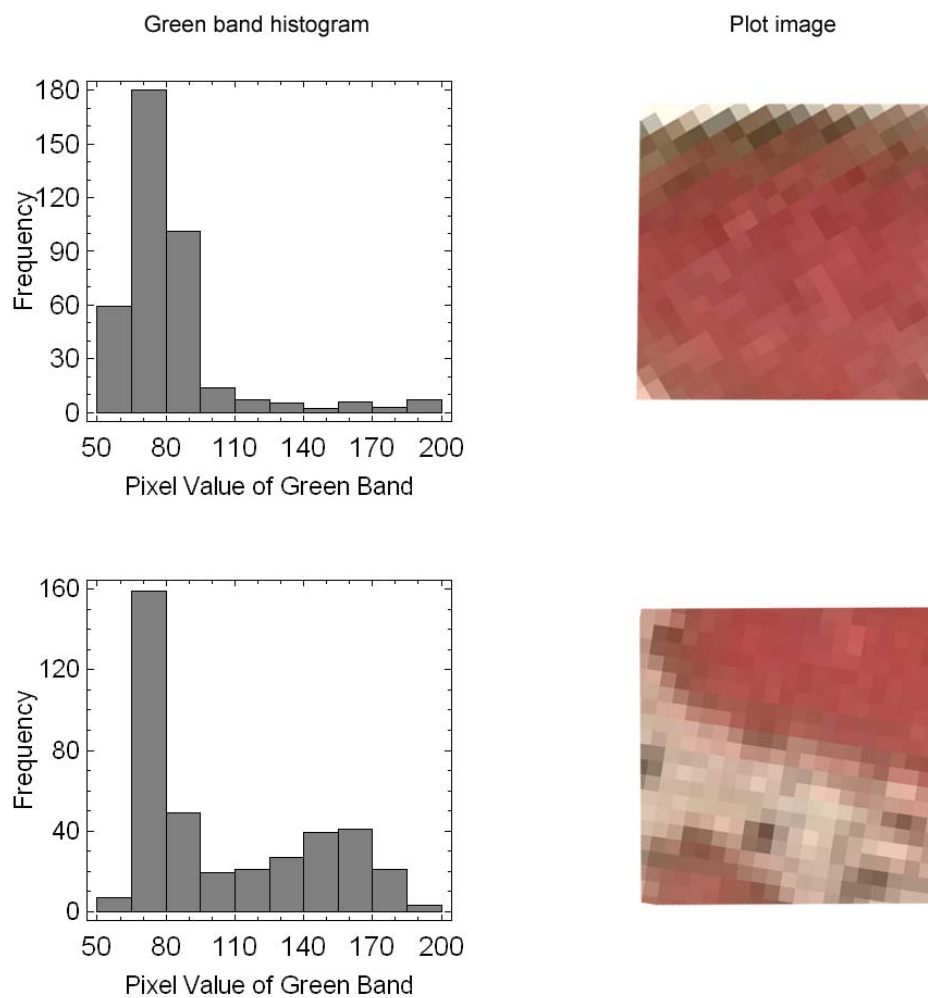


Fig. 6. Green band histograms and images in false color (RGB: infrared, red, green) of two plots with maximum (upper) and minimum presence of vegetation (bottom). Upper plot with dry biomass of 821.27 kg and standard deviation of the green band 26.90; bottom plot with dry biomass of 275.5 kg and 38.33 for the value of standard deviation of the green band.

The map of dry biomass (Fig. 7) of the study area was generated by selecting the model with the highest coefficient of determination (combination of LiDAR data and the airborne image). Clear areas on the image indicate open areas with no vegetation, and tree areas. The values of biomass are concentrated on north facing slopes and show a great degree of concordance with the field data. It is hoped that this map will contribute to improving knowledge of Mediterranean forests and be applied to create fuel-type maps for better accuracy in fire behavior modeling, biodiversity, forest management, and carbon dynamics.

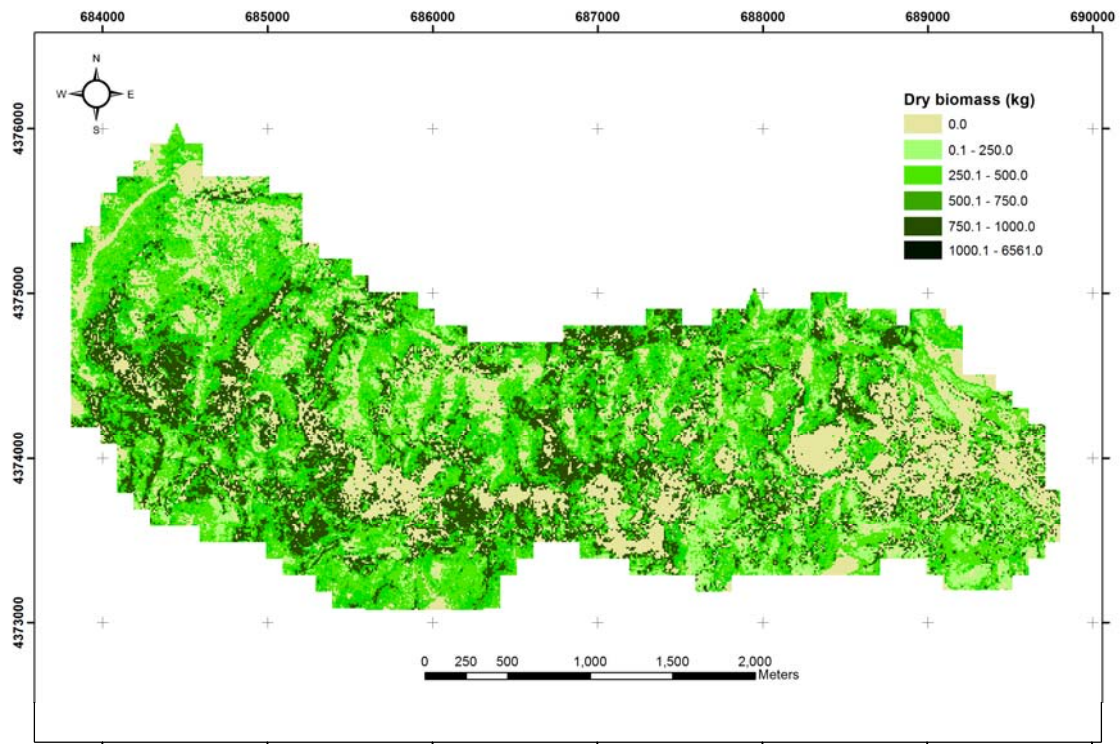


Fig. 7. Map of dry biomass of the study area. Dark green colour shows the areas with more biomass, light green areas with less biomass, and light brown areas without shrub biomass.

The coefficients of determination for predicting wet biomass (Table 5) were slightly lower than those obtained for dry biomass (Table 4). This could be explained considering that wet biomass of shrubs is more variable as their irrigation can be very different depending on the slope and heatstroke, as it was demonstrated in [5]. This fact does not affect in the variability of dry biomass, which produced more accurate models. This means that every shrub has the approximately same density and the dry biomass only depends on the plant sizes, that area estimated using LiDAR techniques. As with dry biomass, the best models were found when variables derived from both LiDAR and spectral data were used with  $R^2$  and RMSE values of 0.77 and 150.2 kg, respectively.

Table 5. Estimation of wet biomass

Input data	Model	$R^2$	RMSE (kg)
LiDAR	$B_w = -932.9 + 4411.5 H\_Std + 4782 H\_Median - 71.4 p_1 - 27.9 p_{0.75-1} - 44.7 p_{0.5-0.75}$	0.64	185.9
Airborne image	$B_w = 1428.6 + 1191.7 NDVI\_Max - 5.3 IR\_Max$	0.32	239.3
LiDAR and airborne image	$B_w = -475.1 + 4625 H\_Std + 4486.7 H\_Median - 71.9 p_1 - 26.6 p_{0.75-1} - 50.8 p_{0.5-0.75} - 11.9 G\_Std$	0.77	150.2

$p < 0.05$

$B_w$  wet biomass in kg in 100 m<sup>2</sup> plots; Independent variables derived from LiDAR data by plots: Median of the heights ( $H\_Median$ ) and Standard deviation of the heights ( $H\_Std$ ); variables derived from the point distribution in height by plots (density metrics): percentage of points in a plot whose height is above 1 m ( $p_1$ ), percentage of points in a plot whose height is between 0.75 m to 1 m ( $p_{0.75-1}$ ); percentage of points in a plot whose height is between 0.50 m to 0.75 m ( $p_{0.5-0.75}$ ); Variables derived from the airborne image by plots: Standard deviation of the green band ( $G\_Std$ ), maximum value of the infrared band ( $IR\_Max$ ), maximum value of the NDVI image ( $NDVI\_Max$ ).

### 3.4 Estimation of volume

The stepwise regression models for each group of selected variables are shown in Table 6. The model with the lowest  $R^2$  value was found when only variables derived from spectral data were used. These data were only able to explain 0.29 of the variance associated with the total volume. As with biomass, these results indicated the low ability of spectral data to predict structural variables of shrub vegetation when it is used alone. In contrast, better results were obtained when LiDAR data was used for predicting volume ( $R^2 = 0.55$ ).

The combination of variables derived from LiDAR and spectral data provided the best fit (Fig. 4b), with values of  $R^2$  and RMSE of 0.84 and 0.1 m<sup>3</sup>, respectively. This stepwise model included seven significant variables derived from image (*G\_Std*, *NDVI\_Mean*), metrics height (*H\_Std*, *H\_Median*) and density metrics of LiDAR data (*p1*, *p0.75-1*, *p0.5-0.75*). The most important of which in terms of explanation of variance was standard deviation of the green band (*G\_std*), followed by the percentage of points whose height was between 0.50 m to 0.75 m (*p0.50* and *p0.75*) and median height (*H\_Median*). What was interesting was the contribution of the spectral data for predicting volume: when *G\_Std* and *NDVI\_Mean* were added, the  $R^2$  value increased from 0.55 to 0.84. These variables explained differences between soil and vegetation. These results revealed spectral data showed complementary information to the LiDAR data for predicting volume.

As with biomass, density metrics showed a strong explanatory power for predicting shrub volume. The plots with the maximum number of points in the intervals 0.50 m to 0.75 m, 0.75 m to 1 m, and heights greater than 1 m are shown in Fig. 5. As expected,

the plot with the greatest percentage of points in the interval 0.50 m to 075 m had lower volume. In contrast the plot with the maximum percentage of points whose height was greater than 1 m had the maximum volume. Consequently, the distribution of the points was related to the plot volume occupied by shrub vegetation.

Table 6. Estimation of volume

Input data	Model	R <sup>2</sup>	RMSE (m <sup>3</sup> )
LiDAR	$V = -0.47 + 2.53 H\_Std + 3.07 H\_Median - 0.064 p_1 - 0.02 p_{0.75-1} - 0.03 p_{0.5-0.75}$	0.55	0.16
Airborne image	$V = 1.77 - 0.01 IR\_Max$	0.29	0.19
LiDAR and airborne image	$V = 0.12 + 3.00 H\_Std + 4.08 H\_Median - 0.06 p_1 - 0.02 p_{0.75-1} - 0.04 p_{0.5-0.75} - 0.01 G\_Std - 1.58 NDVI\_Mean$	0.84	0.01

p<0.05

Volume in m<sup>3</sup> in 100 m<sup>2</sup> plots; Independent variables derived from LiDAR data by plots: Median of the heights (*H\_Median*) and Standard deviation of the heights (*H\_Std*); variables derived from the point distribution in height by plots (density metrics): percentage of points in a plot whose height is above 1 m (*p<sub>1</sub>*), percentage of points in a plot whose height is between 0.75 m to 1 m (*p<sub>0.75-1</sub>*); percentage of points in a plot whose height is between 0.50 m to 0.75 m (*p<sub>0.5-0.75</sub>*); Variables derived from the airborne image by plots: Standard deviation of the green band (*G\_Std*), maximum value of the infrared band (*IR\_Max*), mean of the NDVI image (*NDVI\_Mean*).

#### 4. Conclusions

This study demonstrates the potential of LiDAR and spectral data to estimate the biomass and volume of shrub vegetation. The best results for biomass and volume estimation were found when LiDAR data and an airborne image with high spatial resolution were combined. We must highlight here the importance of the density metrics to estimate biomass and volume. Our study also shows the importance of

adding spectral data to LiDAR data variables to explain above all, the variation of volume. We also noted that the accuracy in predicting volume was lower than for biomass when only LiDAR data was used. This could be because field volume is an indirect calculation from a base diameter and height measurements, whereas biomass is a direct measurement. However when LiDAR and spectral data are combined, the volume is predicted with more accuracy.

The results of this study show how biomass and volume estimation and mapping can be obtained in shrub forests in Mediterranean areas using medium density LiDAR data, visible and NIR aerial images, and applying an appropriate variable selection procedure. The models proposed could be used in similar dense forest having a restricted number of similar species and low shrub.

## **5. References**

- [1] Rango A, Chopping M, Ritchie J, Havstad K, Kustas W, Schmugge, T. Morphological characteristics of shrub coppice dunes in desert grasslands of southern New Mexico derived from scanning LiDAR. *Remote Sens Environ* 2000;74(1): 26-44.
- [2] Mundt JT, Streutker DR, Glenn NF. Mapping sagebrush distribution using fusion of hyper spectral and LiDAR classifications. *Photogramm Eng Remote Sens* 2006;72(1): 47-54.
- [3] Riaño D, Chuvieco E, Ustin SL, Salas J, Rodríguez-Pérez JR, Ribeiro LM et al. Estimation of shrub height for fuel-type mapping combining airborne LiDAR and simultaneous color infrared ortho imaging. *Int J Wildland Fire* 2007;16(3): 341-48.



- [4] Mikšys V, Varnagiryte-Kabasinskiene I, Stupak I, Armolaitis K, Kukkola M, Wójcik J. Above-ground biomass functions for Scots pine in Lithuania. *Biomass Bioenerg* 2007;31(10): 685-92.
- [5] Velázquez-Martí B, Fernández-González E, Estornell J, Ruiz LA. Dendrometric and dasometric analysis of the bushy biomass in Mediterranean forests. *For Ecol Manage* 2010;259(5): 875-82.
- [6] Zheng D, Rademacher J, Chen J, Crow T, Bresee M, Le Moine J, et al. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sens Environ* 2004;93(3): 402-11.
- [7] Top N, Mizoue N, Ito S, Kai S, Nakao T, Ty S. Re-assessment of woodfuel supply and demand relationships in Kampong Thom Province, Cambodia. *Biomass Bioenerg* 2006;30(2): 134-43.
- [8] Cremer T, Velázquez-Martí B. Evaluation of two harvesting systems for the supply chips in Norway spruce forest affected by bark beetles. *Croat J For Eng* 2007;28(2): 145-55.
- [9] Mani S, Parthasarathy N. Above-ground biomass estimation in ten tropical dry evergreen forest sites of peninsular India. *Biomass Bioenerg* 2007;31(5): 284-90.
- [10] Popescu SC. Estimating biomass of individual pine trees using airborne LiDAR. *Biomass Bioenerg* 2007;31(9): 646-55.
- [11] Means JE, Acker SA, Brandon JF, Renslow M, Emerson L, Hendrix CJ. Predicting forest stand characteristics with airborne scanning LiDAR. *Photogramm Eng Remote Sens* 2000;66(11): 1367-71.
- [12] Hyyppä J, Kelle O, Lehtikoinen M, Inkinen MA segmentation-based method to retrieve stem volume estimates from 3-d tree height models produced by laser scanners. *IEEE Trans Geosci Remote Sensing* 2001;39(5): 969-75.

- [13] Lefsky MA, Cohen WB, Harding DJ, Parker GG, Acker SA, Gower ST. LiDAR remote sensing of above-ground biomass in three biomes. *Glob Ecol Biogeogr* 2002;11(5): 393–99.
- [14] Hyyppä J, Hyyppä H, Leckie D, Gougeon F, Yu X, Maltamo M. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *Int J Remote Sens* 2008;29(5): 1339-66.
- [15] Nelson R, Krabill W, Tonelli J. Estimating forest biomass and volume using airborne laser data. *Remote Sens Environ* 1988;24(2): 247- 67.
- [16] Næsset E. Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic full-scale operational project. *Scand J Forest Res* 2004;19(6): 554-57.
- [17] Andersen H-E, McGaughey RJ, Reutebuch SE. Estimating forest canopy fuel parameters using LiDAR data. *Remote Sens Environ* 2005;94(4): 441-49.
- [18] Li Y, Andersen H-E, McGaughey RA. A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *West J Appl For* 2008;23(4): 223-31.
- [19] Persson A, Holmgren J, Söderman U. Detecting and measuring individual trees using an airborne laser scanner. *Photogramm Eng Remote Sens* 2002;68(9): 925-32.
- [20] Maltamo M, Eerikäinen K, Pitkänen J, Hyyppä J, Vehmas M. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sens Environ* 2004;90(3): 319-30.
- [21] Streutker DR, Glenn NF. LiDAR measurement of sagebrush steppe vegetation heights. *Remote Sens Environ* 2006;102(1-2): 135-45.
- [22] Popescu SC, Wynne RH, Nelson RF. Estimating plot-level tree heights with LiDAR: local filtering with a canopy-height based variable window size. *Comput Electron Agric* 2002;37(1-3): 71-95.

- [23] Holmgren J, Nilsson M, Olsson H. Estimation of tree height and stem volume on plots using airborne laser scanning. *For Sci* 2003;49(3): 419- 28.
- [24] Kim Y, Yang Z, Cohen WB, Pflugmacher D, Lauver CL, Vankat JL. Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint LiDAR data. *Remote Sens Environ* 2009;113(11): 2499-510.
- [25] Gaveau DLA, Hill RA. Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. *Can J Remote Sens* 2003;29(5): 650-57.
- [26] Hill RA, Thomson AG. Mapping woodland species composition and structure using airborne spectral and LiDAR data. *Int J Remote Sens* 2005;26(17): 3763-79.
- [27] Hopkinson C, Chasmer LE, Sass G, Creed IF, Sitar M, Kalbfleisch W, et al. Vegetation class dependent errors in LiDAR ground elevation and canopy height estimates in a boreal wetland environment. *Can J Remote Sens* 2005;31(2): 191-206.
- [28] Patenaude G, Hill RA, Milne R, Gaveau DLA, Briggs BBJ, Dawson TP. Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sens Environ* 2004;93(3): 368-80.
- [29] Su JG, Bork EW. Characterization of diverse plant communities in Aspen Parkland rangeland using LiDAR data. *Appl Veg Sci* 2007;10(3): 407-16.
- [30] Estornell J, Ruiz LA, Velázquez-Martí B. Study of shrub cover and height using Lidar data in a Mediterranean area. *For Sci* 2011;57(3): 171-79.
- [31] Martinuzzi S, Vierling LA, Gould WA, Falkowski MJ, Evans JS, Hudak AT, Vierling KT. Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sens Environ* 2009;113(12): 2533-46.

- [32] Bork EW, Su JG. Integrating LiDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis. *Remote Sens Environ* 2007;111(1): 11-24.
- [33] Mutlu M, Popescu SC, Stripling C, Spencer T. Mapping surface fuel models using LiDAR and multispectral data fusion for fire behavior. *Remote Sens Environ* 2008;112(1): 274-85.
- [34] Verrelst J, Geerling GW, Sykora KV, Clevers JGPW. Mapping of aggregated floodplain plant communities using image fusion of CASI and LiDAR data. *Int J Appl Earth Obs Geoinf* 2009;11(1): 83-94.
- [35] Popescu SC, Wynne RH, Scrivani JA. Fusion of small-footprint LiDAR and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. *For Sci* 2004;50(4): 551-65.
- [36] Hyde P, Dubayah R, Walker W, Blair JB, Hofton M, Hunsaker C. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sens Environ* 2006;102(1-2): 63-73.
- [37] Van Aardt JAN, Wynne RH, Oderwald RG. Forest volume and biomass estimation using small-footprint LiDAR-distributional parameters on a per-segment basis. *For Sci* 2006;52(6): 636-49.
- [38] García M, Riaño D, Chuvieco E, Danson FM. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sens Environ* 2010;114(4): 816-30.
- [39] Velázquez-Martí B, Annevelink E. GIS application to define biomass collection points as sources for linear programming of delivery networks. *Trans. ASABE* 2009;52(4):1069-78

- [40] Velázquez-Martí B, Fernandez-Gonzalez E. Mathematical algorithms to locate factories to transform biomass in bioenergy focused on logistic network construction. *Renew Energy* 2010;35(9):2136-42.
- [41] Clark ML, Clark DB, Roberts DA. Small-footprint LiDAR estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sens Environ* 2004;91(1): 68-89.
- [42] Estornell J, Ruiz LA, Velázquez-Martí B, Hermosilla T. Analysis of the factors affecting LiDAR DTM accuracy in a steep shrub area. *Int J Digit Earth* 2011;4(6); 521–38 (2011).
- [43] Hudak AT, Crookston NL, Evans JS, Falkowski MJ, Smith AMS, Gessler PE, et al. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return LiDAR and multispectral satellite data. *Can J Remote Sens* 2006;32(2): 126-38.
- [44] McCombs JW, Roberts SD, Evans DL. Influence of fusing LiDAR and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation. *For Sci* 2003; 49(3): 457-66.
- [45] Erdody TL, Moskal LM. Fusion of LiDAR and imagery for estimating forest canopy fuels. *Remote Sens. Environ* 2010; 114(4): 725-37.