

**AUTOMATED MAPPING WITH LiDAR AND SPECTRAL
CHARACTERIZATION IN MEDITERRANEAN FOREST
AGROECOSYSTEMS**

Ricardo MARTÍNEZ^{1,2*}, César VICENTE², Nuria SÁNCHEZ-LÓPEZ^{2,3},
Javier MONTALVO^{1,2}

¹Department of Ecology and Animal Biology, University of Vigo, Spain

²The Matrix Foundation, Research and Sustainable Development, Spain

³College of Natural Resources, University of Idaho, USA

*Corresponding author: martinez.prentice@gmail.com

ABSTRACT

Mapping with LiDAR data is not a standardized practice, though LiDAR databases are increasing in all countries in Europe. We develop and test a simple method for automated land-cover mapping. The study area was a farm located at a natural park of southern Spain. It comprises 502 ha covered by Mediterranean forest agroecosystems, like dehesa (a very open woodland of scattered evergreen trees used by grazing animals), woodland and scrubland, and transitions among them, composing a heterogeneous landscape. This heterogeneity is caused by variations in holm and cork oak tree density and a sclerophyllous shrub cover, i.e., 3D structure of woody vegetation. Using aerial photographs digitization, Landsat image classification, and image segmentation of tree crowns, land-cover maps were generated. Besides, other maps were produced from LiDAR-derived canopy cover and height of tree vegetation and shrub stratum. These 3D variables allowed to a wall-to-wall characterization of woody vegetation land-cover classes in the study area, that was completed with a NDVI assessment. The results show that automated mapping with LiDAR is reliable and accurate enough in comparison with other mapping techniques. It outperforms them because its higher spatial resolution, and can be combined with other remote sensing methods to provides an improved understanding of forest landscapes.

Keywords: *canopy, Dehesa, forest structure, GIS, LiDAR, vegetation structure.*

INTRODUCTION

Remote sensing and Geographic Information Systems (GIS) are used for land-cover mapping, characterization and monitoring land-cover changes at a local, regional and global scale (Rogan & Chen 2004; Giri 2012). Point cloud data from active sensors like LiDAR provide a 3-dimensional (3D) information of features of land-cover classes and, specifically, 3D structure of woody vegetation (e.g. Parent et al. 2015). Land-cover mapping at high or moderate spatial resolution is a

challenge in complex Mediterranean landscapes with mosaic vegetation and high spatial variability of tree and shrub cover. In addition, recent studies combine LIDAR with satellite images, demonstrating its capacity to characterize the Mediterranean vegetation (Maselli et al. 2017; Gouveia et al. 2017). Land-cover mapping with low-density LiDAR point clouds in Mediterranean agroecosystems is not tested yet.

The goal of this research is to develop and to test an automated simple method for land-cover mapping from airborne LiDAR data that can be applied to a heterogeneous landscape. The specific objectives are: (1) to identify and to map the woody vegetation land-cover classes with LiDAR data in a relatively large area, and (2) to characterize the woody vegetation of land-cover classes and their internal spatial structure from 3D variables extracted from LiDAR and multi-spectral satellite data.

MATERIALS AND METHODS

The study area is the farm Zahurdillas, which is part of the Sierra de Hornachuelos Natural Park located in Sierra Morena mountain range (Southern Spain). It comprises 502 ha covered by different Mediterranean forest agroecosystems. Dehesa is the Spanish name of a common type of vegetation similar to savanna. It is a very open woodland of scattered evergreen trees –mainly holm and cork oak– where grasses, tree and shrub biomass, and acorns are used by livestock, sheep and pigs. Besides the dehesa, other woody vegetation were present, such as woodland and scrubland, as well as transitions between them, composing a heterogeneous land-cover landscape. The study area has been divided into a grid of 30x30 m cells. The coordinate system used for the study area was ETRS89 and UTM projection, zone 30.

Two conventional and two more modern mapping methods have been used. High-resolution aerial images (50 cm pixel size) of June 2014, provided by the National Geographic Institute of Spain (IGN), have been used for an accurate photointerpretation and manual digitization. This method allowed the identification of six land-cover classes (digitized map). The criteria for the delimitation of these classes with woody vegetation were the tree density, tree cover, and shrub abundance. Landsat 8 satellite images (30 m pixel size and 7 spectral bands) of July 2016, downloaded from the Earth Explorer website (USGS 2017) were used to generate a map of land-cover classes by the supervised classification method. These images were also used to generate a NDVI (Normalized Difference Vegetation Index) map, which is related to chlorophyll concentration and vigor of vegetation (Glenn et al. 2008).

Segmentation of tree crowns from aerial images was performed using a multiresolution algorithm, and the nearest neighbor algorithm for classification (Aldrich 1997), both implemented on the eCognition software (Trimble 2017). The multiresolution segmentation algorithm is a bottom up segmentation which departs from one pixel and merges neighbor pixels according to a heterogeneity and size parameters (Hamilton et al. 2007). A relatively simple segmentation was carried

out. Compactness and shape parameters were set at default values and scale parameter was defined at 50. The three bands of the image were used as input data for segmentation, assigning the green band a double weight compared to the red and blue bands. Classification was performed using two main classes (vegetation and non-vegetation) to identify objects that were likely to represent tree crowns. Several objects were manually selected for each class to be used as training data for classification. Tree crown area within each cell estimated by segmentation was used to calculate the relative total tree canopy cover (RTCC_s) and delimitate land-cover classes.

Airborne LiDAR data was acquired from the IGN, between December 2014 and January 2015. The density of the point cloud was 0.5 points/m². Data was reclassified automatically. A Digital Terrain Model (MDT, 1x1 m), based on LiDAR points classified as soil was obtained using the IDW method (interpolation by weighted inverse distance; Watson et al. 1985) to calculate the normalized height of the LiDAR point cloud. In each cell, four LiDAR-derived attributes were estimated: tree canopy height (TCH) defined by the 85th percentile of height, establishing a threshold of 3 m (Gopalakrishnan et al. 2015), after calibration with an empirical model; shrub canopy height (SCH): 95th percentile height of shrub stratum (points below 3 m); tree canopy cover index (TCCI), and shrub canopy cover index (SCCI). These last two indexes have been estimated as the number of high vegetation points and the number of vegetation points below 3 m, respectively, expressing their units in points/ha.

ArcGIS 10.4 software has been used for the spatial data analysis. An additional tree crown segmentation has been performed by the aggregation of LiDAR points of high vegetation in polygons with a smoothed geometry using GIS tools. Relative tree canopy cover derived from this segmentation (RTCC_L) was used to evaluate the TCCI significance. From the combination of the TCCI and SCCI maps using GIS tools, an automated map of woody vegetation land-cover classes was produced. The characterization of each vegetation class has been done by limiting the descriptive statistical analysis to those cells which were completely within each vegetation class, avoiding the other cells which contained different vegetation classes. Coefficient of variation (CV) was used as a measure of total spatial variability of a variable within a land-cover class. The automated map was compared to the digitized map, without assuming its inherent quality. For those cells that belonged to two or more land-cover classes derived from the manual digitization, the class which occupied a larger area within the cell was assigned. The agreement between automated and digitized (reference) maps was evaluated using the confusion matrix method and Kappa index, which is a measure of agreement for multinomial data commonly used for thematic mapping accuracy assessment (Rossiter 2014). We have analysed both global accuracy and land-cover class accuracy. R software was used for these data analyses.

RESULTS AND DISCUSSION

Maps derived from digitization and image analysis

Figure 1a-c shows three maps. The first map represents a conventional map based on digitization of aerial photography (Figure 1a). From this map, four main land-cover classes with woody vegetation and different tree density and shrub cover were identified and mapped: Woodland (W); Woodland/Scrubland Complex (W/SC); Dehesa with a sparse shrub cover (D-ss), and Dehesa with a dense shrub cover (D-ds). The relative area of these land-cover classes in the farm was: 42% (W), 4% (W/SC), 48% (D-ss) and 5% (D-ds). Other minor land-cover classes were also mapped but they do not include woody vegetation. The second map was derived from supervised classification of Landsat images (Figure 1b). This map only discriminated approximately between D-ss (48% of total area) and the rest (W, W/SC and D-ds; 52%). The product of segmentation of tree crowns from aerial photography (Figure 1c) produced similar land-cover units: D-ss (56%, $10\% < RTCC_s < 70\%$); W, W/SC and D-ds (43%; woody canopy cover $>70\%$); besides, Grassland unit (1%) was identified ($RTCC_s < 10\%$). Though $RTCC_s$ was estimated and used to delimitate land-cover classes, there was an overestimation of this variable, due to the limitations derived from the segmentation, but also mapping the D-ss and the rest (W, W/SC and D-ds) was done approximately.

Tree canopy cover and height

LiDAR-derived TCCI was useful to discriminate vegetation classes (Figure 1d). There was a strong nonlinear relationship between TCCI and a LiDAR estimation of tree canopy area, $RTCC_L$ ($N=4,216$; $R^2_{aj}=0.86$; $p<0.001$). TCCI was above 2,000 in most parts of W and W/SC, and below 2,000 in most parts of D-ss and D-ds. LiDAR allowed to tree canopy height mapping (Figures 1h and 2a). TCH was 6-8 m in 92% of W area, but was more concentrated, 7-8 m, in W/SC, where represented 73% of area. Average TCH was 7.1 m in W and 7.3 m in W/SC. TCH was lower in dehesa: 6-7 m in most area of dehesa (76% in D-ss and 67% in D-ds). Average TCH was 6.7 m in D-ss and 6.5 m in D-ds. The CV in TCH in the four woody vegetation classes was 6-8%.

Spatial internal structure of land-cover classes

Figure 2b-c shows the spatial variability within each woody vegetation land-cover class considering two different LiDAR-derived attributes, SCCI and SCH. Shrub canopy cover was clearly mapped (Figure 1e). W and D-ss showed strong skewness of SCCI values (Figure 2c). The distribution of SCCI was right-tailed in both vegetation classes, which indicates a spatial uniformity with predominance of low and very low shrub cover classes in 75-80% of the total area, being average shrub cover slightly less in D-ss than in W. Oppositely, W/SC and D-ds showed a more symmetric distribution of SCCI values (Figure 2c). SCCI was medium in about a third of total area of both woody vegetation classes, and almost 40% of total area showed higher SCCI in D-ds. Shrub cover map clearly showed spatial variability and differences among woody vegetation classes (Figure 2e). Their

internal variability could be partly the effect of a misidentification of training objects within the vegetation classes (Martin et al. 2001). For example, some isolated areas with SCCI over 400 in northern and southern parts of W area should have been classified as W/SC. W and D-ss showed a left-tailed skewness of SCH values (Figure 2b) with a relatively large variation range in canopy height classes in most part of the total area within these vegetation classes. SCH was between 1.5 and 2.5 m, but SCH in 35% of the total area was below 1.5 m in D-ss, where variability in SCH was greater than in W. Conversely, W/SC and D-ds showed a more concentrated distribution of SCH values (Figure 2b). SCH was between 1.5 and 2.5 m in 99% of total area of these woody vegetation classes, though higher SCH classes were more relevant in D-ds.

Spectral characterization of land-cover classes

NDVI variability was clearly related with the relevance of leaf biomass in woody vegetation (Figure 1g). The higher the TCH, the greater the values of NDVI ($N=4,245$; $R^2_{aj}=0.28$; $p<0.001$). Landsat images dates from June, meaning that during that month, almost all the natural grass vegetation in southern Spain is dry, losing all its chlorophyll (Figure 1c). Average NDVI in the small grassland patches in the farm was the lowest, 0.20. Average NDVI in both types of dehesa were 0.28 and 0.29, reflecting probably both, a large area of dry grass stratum and a low density of tree and shrub leaf biomass, being its overall spatial variability relatively low ($CV<10\%$). NDVI values were below 0.3 in most of the area of dehesa (81% of D-ss and 67% of D-ds); this last class showed NDVI between 0.3 and 0.4 in 34% of total area, probably due to the contribution of shrub leaf biomass. Average NDVI in W and W/SC was 20-25% higher than in dehesa classes, which could be attributed to a greater density of leaf biomass of trees and trees with shrubs, respectively. NDVI range of 0.3-0.4 was observed in 98% of total area of W/SC, while the same range was limited to 75% of W area, which interestingly showed 10% of area in the top values recorded in the farm, range 0.4-0.5. The overall internal variability of W and W/SC classes was also low, with CV of 12% and 6%, respectively. Spectral discrimination through NDVI was useful but not enough to differentiate between dehesa classes nor between W and W/SC. NDVI allows to differentiate among ecosystems (Pettorelli et al. 2005), but not among woodlands with similar species composition as occur in the study area.

Automated map generation and accuracy

Combining two LiDAR-derived indexes related to tree cover (TCCI) and shrub cover (SCCI), allowed to generate and automated land-cover map of woody vegetation (Figure 1f). This map represented a similar image of the land-cover classes area and distribution identified by digitization. In fact, its accuracy can be considered higher than other maps because it was based in high spatial resolution measures of 3D the vegetation structure. The comparison of digitized and automated maps of land-cover classes revealed a reasonable accuracy of the automated mapping method based just in LiDAR point-clouds.

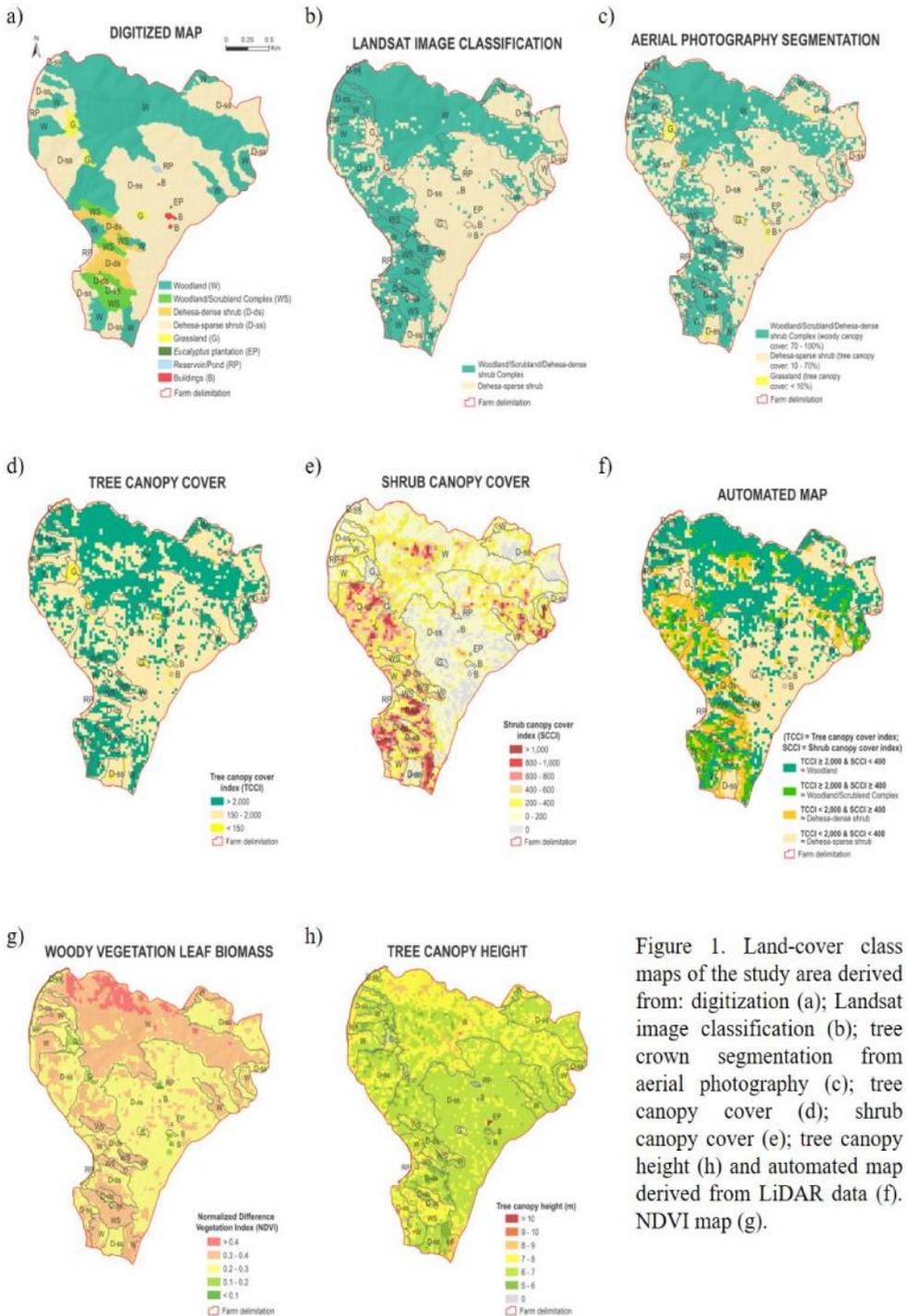


Figure 1. Land-cover class maps of the study area derived from: digitization (a); Landsat image classification (b); tree crown segmentation from aerial photography (c); tree canopy cover (d); shrub canopy cover (e); tree canopy height (h) and automated map derived from LiDAR data (f). NDVI map (g).

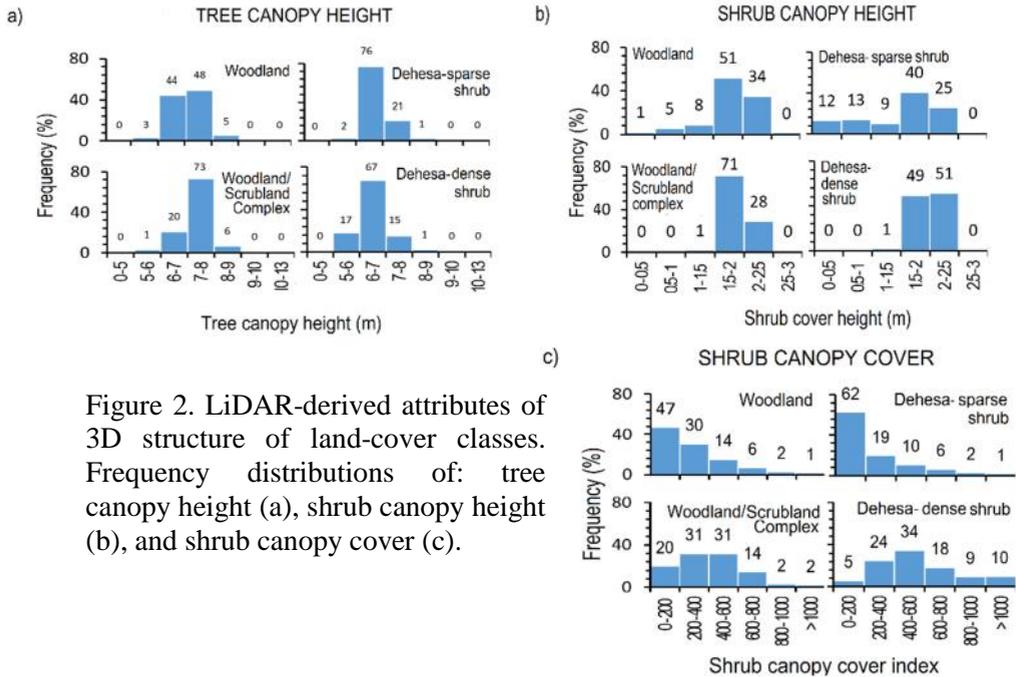


Figure 2. LiDAR-derived attributes of 3D structure of land-cover classes. Frequency distributions of: tree canopy height (a), shrub canopy height (b), and shrub canopy cover (c).

First, the confusion matrix derived had a Kappa index of 0.297. Although it was not very high, it was a statistical evidence of a general agreement of automated and digitized classification of land-cover units. Accuracy in automated identification of W and D-ss was high (Kappa values of 0.415 and 0.412, respectively). These results are relevant because the 65% of the W area and the 71% of the D-ss area were correctly identified. These vegetation classes represented the 90% of total woody vegetation in the study area. The automated identification of the other two classes, W/SC and D-ds, was less accurate. Nevertheless, these apparent incorrect classifications were probably related to the identification errors, caused by the manual digitization of both land-cover classes, that included a mixture of tree and shrub strata, not easily perceived by the human eye. The accuracy of the automated map was the effect of the quality of 3D data and the standardized method applied, with quantitative thresholds. The automated map can be considered better than the real reference digitized map. Besides, the high spatial resolution of the automated map suggested that its quality outperforms the digitized map. These results are consistent with previous applications of LiDAR data for automated land cover mapping in non-Mediterranean areas (e.g. Parent 2015).

CONCLUSIONS

Preliminary results indicate that airborne LiDAR data allows to automate the production of land-cover maps, and to characterize their 3D vegetation structure in relatively large areas of complex Mediterranean vegetation with a high spatial heterogeneity.

ACKNOWLEDGEMENT

This work is part of the 'Ecological and Territorial Research Program' of The Matrix Foundation, co-financed by the Spanish Ministry of Agriculture and Fisheries, Food and Environment. We would like to thank the owners of the Zahurdillas farm. Lourdes García and Sara Ortega who effectively assisted in map digitization and accuracy assessment.

REFERENCES

- Aldrich, J. (1997). R.A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12, 162-176.
- Trimble (2017). eCognition Land Cover Mapping Software website. <http://www.ecognition.com>
- Giri, C.P. (2012). *Remote sensing of land use and land cover: principles and applications*. CRC Press, Boca Raton.
- Glenn, P.E., Huete, R.A., Nagler, L.P. & Nelson G.S. (2008). Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: what vegetation indices can and cannot tell us about the landscape. *Sensors*, 8, 2136-2160.
- Gouveia, C.M., Trigo, R.M., Beguería, S. & Vicente-Serrano, S.M. (2017). Drought impacts on vegetation activity in the Mediterranean región: An Assessment using remote sensing data multi-scale drought indicators. *Global and Planetary Change*, 151, 15-27.
- Gopalakrishnan, R., Thomas, V.A., Coulston, J.W., Wynne, R.H. (2015). Prediction of canopy heights over a large region using heterogeneous lidar datasets: efficacy and challenges. *Remote Sensing*, 7, 11036-11060.
- Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proceedings of the Eighth International Conference on Computer Vision (ICCV'01)*, Volume 2.
- Maselli, F., Vaccari, F.P., Chiesei, M., Romanelli, S. & D'Acqui, L.P. (2017). Modelling analyzing the water and carbon dynamics of Mediterranean macchia by the use of ground and remote sensing data. *Ecological Modelling*, 351, 1-13.
- Parent, J.R., Volin, J.C., Civco, D.L. (2015). A fully automated approach to land cover mapping with airborne LiDAR and high resolution multispectral imagery in a forested suburban landscape. *SPRS J. Photogramm. Remote Sens.*, 104, 18-29.
- Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.M., Tucker, C.J. & Stenseth, N.C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol. Evol.* 20, 503-510.
- Rogan, J., Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning*, 61, 301-325

- Rossiter D.G. (2014). Technical Note: Statistical methods for accuracy assessment of classified thematic maps. http://www.css.cornell.edu/faculty/dgr2/teach/R/R_ac.pdf
- USGSG 2017. Earth Explorer website for Landsat data access. United States Geological Service (USGS). <https://earthexplorer.usgs.gov>
- Watson, D.F. & Philip, G.M. (1985). A Refinement of Inverse Distance Weighted Interpolation. *Geoprocessing*, 2, 315–327.