

# SINGLE STRATA CANOPY COVER ESTIMATION USING AIRBORNE LASER SCANNING DATA

*António Ferraz<sup>1,2</sup>, Clément Mallet<sup>1</sup>, Gil Gonçalves<sup>2,3</sup>, Margarida Tomé<sup>4</sup>, Paula Soares<sup>4</sup>, Luisa Pereira<sup>5</sup>, Stéphane Jacquemoud<sup>6</sup>*

- 1- IGN/SR, Laboratoire MATIS, Université Paris Est, Saint-Mandé, France
- 2- Instituto de Engenharia de Sistemas e Computadores de Coimbra, Coimbra, Portugal
- 3- Departamento de Matemática, Universidade de Coimbra, Coimbra, Portugal
- 4- Universidade Técnica de Lisboa, Instituto Superior de Agronomia, Centro de Estudos Florestais, Lisboa, Portugal
- 5- Universidade de Aveiro, Escola Superior de Tecnologia e Gestão de Águeda, Águeda, Portugal
- 6- Institut de physique du globe de Paris, Géophysique spatiale et planétaire, Paris, France

## ABSTRACT

Canopy height and canopy cover are two important biophysical variables in forest characterization. As far as airborne laser scanning (ALS) studies is concerned, canopy height has been mapped at a high spatial resolution through canopy height models (CHM). Conversely, canopy cover had been computed at a coarser resolution (plot-level) by means of penetration rate metrics (PRM). In this work we present a method dedicated to compute single strata canopy cover at a spatial resolution similar to the CHM. Vegetation density maps are interpolated from the ALS point cloud pattern using a novel method based on the kernel density estimators (KDE). We developed an automatic variable-bandwidth approach that accommodates for both the point cloud density variability inherent to any ALS survey and the shading effect induced by taller canopy on laser beams. Results are compared with the PRM approach.

*Index Terms*— airborne laser scanning, forest, canopy cover.

## 1. INTRODUCTION

Single strata canopy cover is an important biophysical variable commonly used to assess the volume and spatial distribution of vegetation in forest ecosystems. Therefore, it is crucial for many process-based ecological and biochemical models such as biomass stocks estimation, fuels distributions or biodiversity assessment [1]. Contrary to both optical and radar imagery, small-footprint ALS systems are able to stratify the forest structure allowing to characterize single strata such as overstory, understory and ground vegetation. It is an active remote sensing technique that emits successive laser beams providing direct distance measurements between a moving platform and the earth surface [2]. Depending on the nature of the target, a single laser pulse emission may induce several backscattered echoes. Therefore, the laser beams are able to penetrate down into the forest canopy layers providing 3-D point cloud that is a discrete model of the forest structure. However, there is no study assessing the ALS reliability to derive single strata canopy cover. Former works either estimated canopy cover without stratifying vegetation structure or focused on a single layer such as the overstory [1, 3] or the understory [4]. There are two main approaches to derive canopy cover from ALS data: PRM and the grid-based methods. The first approach compares the number of ALS echoes corresponding to the stratum

of interest, e.g. the overstory, to the number of echoes that lie below it. The more echoes gathered by a given strata the denser the canopy cover. However, this method is biased on plots that are not located directly below a flight line [3]. In fact, laser beams emitted with no near-vertical scan angles do not measure vegetation vertically, which affects the proportion of echoes between the stratum of interest and the area vertically below it. To minimize this effect, researchers have been using only the echoes derived from laser beams emitted with near-vertical scan angles, e.g. scan angles less than 15° [1]. However, measurements obtained from different perspectives can provide additional information on vegetation structure.

Conversely, grid-based methods are not biased with respect to echoes issues from laser beams emitted with large scan angles. Echoes are either projected or interpolated into a binary image with a given resolution. The canopy cover is calculated through the ratio of pixels with data to those with no-data. The success of this approach highly depends on the selection of the image resolution. A coarse one leads to the “horizontal expansion” of vegetation features and consequently to an over-estimation of the canopy cover. On the contrary, a finer image resolution leads to its under-estimation: many pixels corresponding to the areas within the outermost perimeter of crowns contain no-data due to the lack of a corresponding ALS measurement. In such a case, results can be improved by applying image processing techniques such as morphological operators, e.g. openings and closings. The goal is to minimize the canopy cover under-estimation by eliminating “lakes”, “gulfs” and “channels” present within the crowns perimeters. However, the filter size used in the procedure, i.e. the structuring element size, is critical. So far, it has been tuned manually. Moreover, due to the point cloud density variability inherent to any ALS survey, a static morphological operator size might not be optimal to deal with a single study area [1].

In this work we propose a novel method for canopy cover estimates. Smooth maps of vegetation density are interpolated from the ALS echo pattern using a technique based on a statistical tool called kernel density estimators (KDE). The degree of smoothing is driven by the size of the kernel, which is commonly called bandwidth. The latter is here automatically adapted to accommodate for both the point cloud density variability among the study area and the shade effect provoked by the taller canopy in the underneath strata surveying.

## 2. KERNEL DENSITY ESTIMATORS

The kernel density estimators are a non-parametric technique for estimating the probability density function (PDF) of a point distribution defined in a  $d$ -dimensional space  $\mathbb{R}^d$ . It is a non-parametric approach that depends on a single parameter: the kernel size, i.e. the bandwidth. However, the choice of the latter is critical because strongly impacts on the results. Variable-bandwidths KDE techniques have been applied in order to limit such impact: the bandwidth varies depending upon the local characteristics of the point distribution. In the following we describe a variable-bandwidth approach based on the so-called sample point density estimator [5]. Let  $X_i \in \mathbb{R}^d$  ( $i = 1, \dots, n$ ) be a point distribution with an unknown density. The shape of the PDF in a given point  $X \in X_i$  according to the KDE can be computed using

$$PDF_{k,h(X_i)}(X) = \frac{1}{n \times h(X_i)^d} \sum_{i=1}^n k(X, X_i, h(X_i)) \times w(X_i) \quad 1)$$

Where  $k$  is called the kernel profile,  $h$  the bandwidth and  $w(X_i)$  corresponds to a weight function. The kernel profile is a function that determines the contribution of the distribution points into the PDF according to their distance from the kernel center. A range of kernel profiles can be applied such as the Laplacian one:

$$k(X, X_i, h(X_i)) = \frac{1}{2 \times h(X_i)} \exp\left(-\frac{\|X - X_i\|}{h(X_i)}\right) \quad 2)$$

The bandwidth is a search radius that controls the extent of the kernel profile: distribution points farther than  $h(X_i)$  from the kernel center have little or no influence on the PDF. Finally, the weight function  $w(X_i)$  it is not part of the original definition of a KDE and its use is optional. However, it is commonly used within task-dependent methods to control the role of each individual point into the PDF.

## 3. METHODS

### 3.1. Material

The study area is located in northwest Portugal covering 9 km<sup>2</sup> with slopes ranging from 2.5% to 34.2% and altitude varying from 70 m to 220 m. The landscape is predominantly composed of woodlands dominated by blue gum eucalyptus (*Eucalyptus globulus* Labill) with some stands of maritime pine (*Pinus pinaster* Ait.). The forest stands are usually strongly populated by understory and ground vegetation. A systematic grid over the area led to the selection of 44 plots, 42 covered by eucalyptus: 30 mature (>4 years old) and 12 juvenile (1-4 years old) and 2 plots populated by pine of 30 and 60 years old. Field measurements were performed on each plot (400 m<sup>2</sup>) according to a protocol adapted from the Portuguese National Forest Inventory. Two experimented field operators estimated visually the canopy cover (%) for each forest strata: overstory, understory and ground vegetation.

During the field inventory, an ALS also surveyed the study area. The data were acquired on July 14, 2008 in a full-waveform mode using a LiteMapper 5600 system with a scanning angle of  $\pm 22.5^\circ$ . The footprint equals 30 cm. From the digitized waveforms a 3D point cloud was extracted using a Gaussian Pulse Estimation technique. Each laser pulse gave rise to 1-5 ALS echoes. On that basis, the latter are commonly classified into single, first, last,

intermediate and last echoes. According to the ALS acquisition parameters a point cloud density of 9.9 pt/m<sup>2</sup> is expected. This variable will be hereafter called *expected point density (epd)*. However, real-world ALS point clouds always show variable point density values that are locally either smaller or larger than those specified in an ALS project [6]. For further details on the field sampling design, the forest stand characterization or the ALS data, please refer to [7].

Our method for single strata canopy cover estimation demands a ALS point cloud stratified beforehand. In this study we take advantage of a former method called adaptive mean shift (AMS 3D) that segment 3D point clouds into 3D segments corresponding to individual crowns [7]. Additionally, such segments are assigned to forest strata: overstory, understory and ground vegetation.

### 3.2. KDE forest canopy maps for single strata

Maps of vegetation density are interpolated from the ALS point cloud by means of the KDE technique (Fig. 1). We establish both a variable-bandwidth and a weight function oriented to perform such a task.

The variable-bandwidth allows applying a multi-scale interpolation to accommodate for the high point cloud density variability inherent to any ALS survey (Section 3.1). For instance, underneath strata are expected to be under-sampled comparing to the topmost layers due to the shade effect caused by a taller canopy. In the following, we develop a specific bandwidth to apply to each stratum of each plot in order to compensate such an effect. Therefore, we can divide the point distribution  $X_i$  (Section 2) into several subsets  $X_{p,l,j}$  corresponding to single strata that had been computed by the AMS 3D (Section 3.2):  $p = 1, \dots, 44$  corresponds to the forest plots within our study area,  $l = \{os, us, gv, gr\}$  stands for a given layer (overstory, understory, ground vegetation, and ground, respectively) and  $j = 1, \dots, m$  corresponds to the specific layer echoes. The strata specific bandwidth is defined by:

$$h(X_{p,l,j}) = h^* \times nh^*(X_{p,l,j}) \quad 3)$$

Where  $h^*$  is a default bandwidth defined to apply to all study area and  $nh^*(X_{p,l,j})$  a normalization factor that will drive the variable-bandwidth settings. In this work we set  $h^* = 0.3$  m, i.e., it equals the ALS footprint (Section 3.1). The normalization factor is defined by:

$$nh^*(X_{p,l,j}) = \frac{opd(X_{p,l,j})}{epd} \quad 4)$$

where  $epd$  corresponds to the *expected point density (epd)* = 9.9 pt/m<sup>2</sup>, (Section 3.1), whereas  $opd$  stands for *observed point density*. The  $opd$  for a given layer  $b \in l$  is computed by means of:

$$opd(X_{p,b,j}) = \frac{\sum_{l=b}^{gr} |X_{p,l,j}^{single}| + \sum_{l=b}^{gr} |X_{p,l,j}^{first}|}{A_p} \quad 5)$$

where  $|\cdot|$  refers to the cardinality of a set, *single* and *first* correspond to single and first echoes (Section 3.1). Finally,  $A_p$  (m<sup>2</sup>) is the area of forest plot  $p$ . Thus, the numerator in (Eq. 5) accounts for the echoes assigned to the strata of interest,  $b$ , as well as for the echoes lying below it (including echoes assigned to the ground).

On the other hand, the weight function (Eq. 1) controls the role of each ALS echoes on the vegetation density maps. It intends to assign a lower weight to the echoes corresponding to the boundaries of vegetation features in order to preserve the sharp edges of vegetation on the vegetation density maps. The weight function is defined as follows:

$$w(X_{p,l,j}) = \frac{vote(X_{p,l,j})}{5}, vote(X_{p,l,j}) = \sum_{t=I}^{IV} \chi_{Q_t}(X_{p,l,j}) \quad (6)$$

Where  $\chi_{Q_t}(X_{p,l,j})$  is an indicator function defined by:

$$\begin{cases} 1 & \text{if } \exists X_{r,l,r}: \|X_{p,l,r} - X_{p,l,j}\| \leq h(X_{p,l,j}) \wedge X_{p,l,r} \in Q_t \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Thus, the function  $vote(\cdot)$  evaluates boundary and inner points as a function of a circumference centered in  $X_{p,l,j}$ , radius= $h(X_{p,l,j})$ , divided into four quadrants  $Q_t$  ( $t = I, \dots, IV$ ).  $vote(\cdot)$  equals 1 or 2 if  $X_{p,l,j}$  has no neighbors or all neighbors lay on the same quadrant, respectively. Similarly, it equals 3, 4 or 5 if  $X_{p,l,j}$  has neighbors in 2, 3 or 4 quadrants.

### 3.3. Single strata canopy cover maps

Canopy cover maps are delineated on the vegetation density maps by means of a density threshold (DT). The latter corresponds to a contour line that limits the areas considered as covered by vegetation (Fig. 1). In order to delineate coherently the canopy cover, DT is defined as a function of the variable-bandwidth:

$$DT_{k,h(X_{p,l,j})} = \frac{1}{m \times h(X_{p,l,j})^2} \times \frac{1}{2 \times h(X_{p,l,j})} \times \frac{1}{5} \quad (8)$$

Note that Eq. 8 is derived from Eq. 1 by setting  $w(X_{p,l,j}) = 1/5$  and  $\|X - X_{p,l,j}\| = 0$ , i.e. corresponds to an isolated echo. An area is considered covered by vegetation if there is at least an agglomeration of two echoes, i.e. if:

$$PDF_{k,h(X_{p,l,j})} > DT_{k,h(X_{p,l,j})} \quad (9)$$

Canopy cover for a single layer can be easily derived from the KDE maps (Fig. 1) as a function of the DT: it equals the ratio between the forest plots area to the area enclosed by the density contours.

### 3.4. Results comparison

Contrary to other works, which collect ground truth datasets by means of more reliable techniques, we compare ALS-derived canopy cover with quick-to-apply field visual observations. Since our ground truth is more prone to miscalculations, we apply a robust linear regression to detect outliers [8] that allows figuring out the real trend of the relationship between field- and ALS-derived canopy cover. Additionally, we are interested in examine our method improvements compared with the commonly used PRM techniques. Canopy cover of both overstory and understory, i.e. if  $b = os$  or  $b = us$ , is computed adapting the First echo Cover Index (FCI) described in [1]:

$$prm(X_{p,b,j}) = \frac{|X_{p,b,j}^{single}| + |X_{p,b,j}^{first}|}{\sum_{l=b}^{gr} |X_{p,l,j}^{single}| + \sum_{l=b}^{gr} |X_{p,l,j}^{first}|} \quad (11)$$

Regarding the ground vegetation ( $b = gv$ ) we apply the Understory Lidar Cover Density (ULCD) defined by [4]:

$$prm(X_{p,b,j}) = \frac{|X_{p,b,j}|}{\sum_{l=b}^{gr} |X_{p,l,j}|} \quad (12)$$

It is worth mentioning that [4] focused on the ALS points ranging from 0.25 m to 1.45 m. Thus, the layer named understory in that work corresponds to our ground vegetation, which was between 0.15 m and 1.3 m high [7]. Finally, to avoid unbiased results only the echoes issues from laser beams emitted with a scanning angle less than  $14^\circ$  are used to compute the PRM [1, 4].

## 4. RESULTS AND DISCUSSION

Table 1 shows the highly different spatial sampling rates at which the 44 forest plots within our study area were measured. For instance, the *opd* (Eq. 5) for mature overstory ranges from 5.8 to 15.4  $pt/m^2$ . The inter-strata variability is lower but stills exist: the mature overstory displays in average an *opd* of 10.3  $pt/m^2$  whereas it equals 9.2  $pt/m^2$  for the ground vegetation. However, within plots covered by denser overstory such difference can be significant.

Table 1. Observed point cloud density ( $pt/m^2$ , Eq. 5) by strata.

	Min	Max	Mean	$\sigma$
Mature overstory	5.8	15.4	10.3	2.6
Juvenile overstory	8.9	16.6	11	2.3
Understory	5.3	14.4	9.2	2.4
Ground vegetation	5.1	15.5	9.2	2.3

For instance, the *opd* for Plot #12 corresponding to overstory and surface vegetation equals 13.3  $pt/m^2$  and 7.56  $pt/m^2$ , respectively. As a result, the corresponding density maps were computed applying a bandwidth of 0.22 m and 0.39 m. Consequently, the ground vegetation KDE map is over-smoothed compared with that of the overstory (Fig. 1). In other words, wider bandwidths spread the influence of ALS echoes over a larger neighborhood. In fact, the variable-bandwidth compensates the lower probability that a given vegetation feature within the surface vegetation strata had being hit compared to the overstory features. However, such influence is as well driven by the weight function: echoes surrounded by echoes have a higher influence on the KDE maps compared to isolated or border echoes. As a result, the ‘‘horizontal expansion’’ of vegetation promoted by wider bandwidths is not uniformly distributed along the forest stands, which allows for preserving the sharp edges of vegetation cover

Figure 2 and Table 2 show the relationship between field-estimated and both KDE- and PRM-derived canopy cover. As far as mature overstory is concerned, both the KDE and the PRM approaches produce results well correlated with the field measurements ( $R^2=0.66$  and  $R^2=0.75$ ) and the results in terms of RMSE are very satisfactory (9.11% and 9.18%, respectively). The PRM method produce more outliers (6.25%) compared with the KDE approach (12.5%). With respect to the juvenile overstory stratum, the PRM-derived results correlate worst than the KDE-derived canopy cover ( $R^2=0.68$ ,  $R^2=0.85$ , respectively) but the RMSE is quite similar

(8.37% and 8.54%, respectively). Therefore, the canopy cover is either over- or under-estimated in a similar magnitude (Table 2).

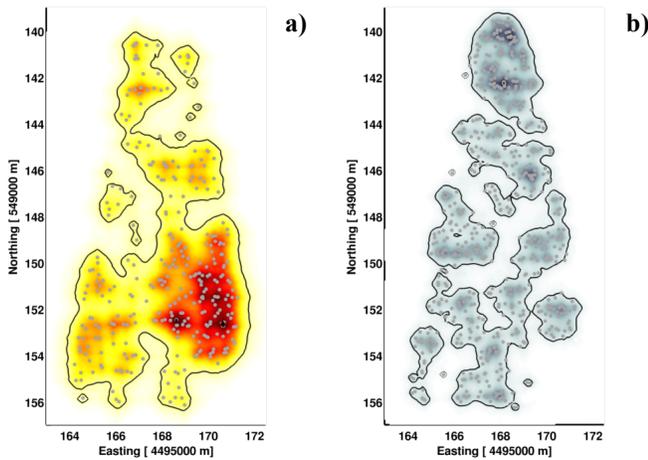


Figure 1. KDE maps for a) surface vegetation and b) overstory of Plot #12. Darker tones correspond to higher densities and ALS echoes are represented by grey dots. DT (Eq. 8) is represented by black contour lines.

With respect to the understory, results of both approaches are as well quite similar. Nevertheless, the KDE-based method improves the RMSE in nearly 5% compared with the PRM approach. Conversely, ground vegetation results are highly improved applying the KDE-based approach: the RMSE was improved up to 13%. Moreover, the underestimation of ground vegetation canopy cover had decreased from 18.37% to 2.49% when applying the KDE-based instead of the PRM method. There are two likely reasons for the poor result of the PRM approach concerning the ground vegetation layer. On the one hand, the PRM discard laser measurements with scanning angles larger than  $14^\circ$  that correspond to 32.2% of the ALS echoes assigned to the ground vegetation that are likely to provide additional information on the characterization of such a strata. On the other hand, the KDE-based approach compensate for the shading effect by means of the variable-bandwidth.

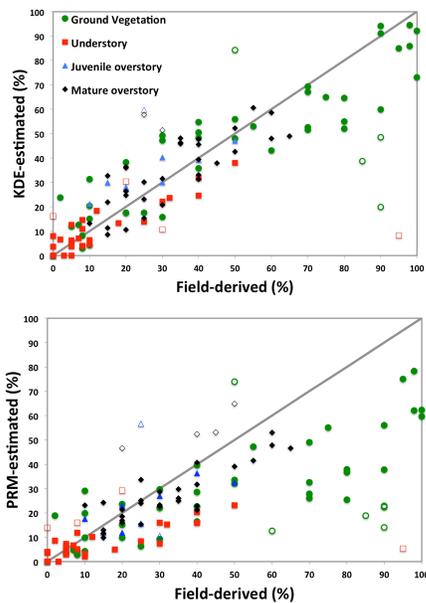


Figure 2. Field canopy cover estimation compared to a) KDE and b) PRM methods. Unfilled symbols correspond to outliers.

Table 2. Linear regression parameters corresponding to Fig. 2. A negative  $\Delta$  means under-estimation.

Stratum	Method	Outliers (%)	$R^2$	RMSE (%)	$\Delta$ (%)
Mature overstory	KDE	6.3	0.66	9.1	1.1
	PRM	12.5	0.75	9.2	-5.1
Juvenile overstory	KDE	20	0.85	8.5	6.1
	PRM	20	0.68	8.4	-4.3
Understory	KDE	12.5	0.85	6.2	-1.9
	PRM	12.5	0.75	11	-5.7
Ground vegetation	KDE	9.1	0.84	13.8	-2.5
	PRM	16.2	0.73	26.6	-18.4

## 5. CONCLUSION

In terms of canopy cover estimates for both mature and juvenile overstory the KDE approach did not meaningfully improve results compared with the PRM approach. However, the KDE map canopy cover at a much higher spatial resolution. Indeed, our approach is able to figure out the canopy cover in spite of the high point cloud density variability within our study area. This is a great advancement with respect to the former high spatial resolution approaches, namely the grid-based approaches. Thus, our method is an automatic technique able to produce single strata canopy cover maps at a high spatial resolution, which could be used jointly to the well know CHM in order to improve ALS canopy cover estimates. We prove that the automatic variable-bandwidth succeeds in compensate for the shading effect. Nevertheless, further investigation is needed to study the methods reliability over other studies areas.

## 6. ACKNOWLEDGMENTS

This work was partially supported by Fundação para a Ciência e Tecnologia (FCT) under project grants PTDC/AGR-CFL/72380/2006 and Pest-C/EEI/UI0308/2011.

## 7. REFERENCES

- [1] L. Korhonen, I. Korpela, J. Heiskanen and M. Maltamo, "Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index", *Remote Sensing of Environment*, 115, pp. 1065-1080, 2011
- [2] C. Mallet and F. Bretar, "Full-waveform topographic lidar: State-of-the-art", *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, pp. 1-16, 2009.
- [3] F. Morsdorf, B. Kötz, E. Meier, K. Itten and B. Allgöwer, "Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction", *Remote Sensing of Environment*, 104, pp. 50-61, 2006.
- [4] B.M. Wing, M.W. Ritchie, K. Boston, W.B. Cohen, A. Gitelman and M.J. Olsen, "Prediction of understory vegetation cover with airborne lidar in an interior ponderosa pine forest" *Remote Sensing of Environment*, 124, pp. 730-741, 2012.
- [5] G. R. Terrel and D.W. Scott, "Variable kernel density estimation", *Ann. Statist.*, 20, pp. 1236-1265, 1992.
- [6] E. P. Baltsavias, "Airborne laser scanning: basic relations and formulas", *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, pp. 199-214, 1999.
- [7] A. Ferraz, F. Bretar, S. Jacquemoud, G. Gonçalves, L. Pereira, M. Tomé, and P. Soares, "3-D mapping of a multi-layered Mediterranean forest using ALS data", *Remote Sensing of Environment*, 121, pp. 210-223, 2012.
- [8] Huber, P.J., *Robust Statistics*, John Wiley & Sons, New York, 1981.