

# Canopy Density Model: A New ALS-Derived Product to Generate Multilayer Crown Cover Maps

António Ferraz, Clément Mallet, Stéphane Jacquemoud, Gil Rito Gonçalves, Margarida Tomé, Paula Soares, Luísa Gomes Pereira, and Frédéric Bretar

**Abstract**—The canopy density model (CDM), a new product interpolated from airborne laser scanner (ALS) data and dedicated to forest structure characterization is presented. It exploits both the multiecho capability of the ALS and a nonparametric density estimation technique called kernel density estimators (KDEs). The CDM is used to delineate the outmost perimeter of vegetation features and to compute forest crown cover (CrCo). Contrary to other works that focus on single-layer forest canopies, CrCo is derived here for each layer, namely, the overstory, the understory, and ground vegetation. The root-mean-square error of prediction determined by using field data acquired over 44 forest stands in a forest in Portugal allows the testing of the reliability of the method: It ranges from 6.21% (overstory) to 13.76% (ground vegetation). In addition, we investigate the ability of the CDM to map the CrCo for individual trees. Finally, two existing methods have been applied to our study site in order to assess improvements, advantages, and drawbacks of our approach.

**Index Terms**—Density estimation robust algorithm, forestry, lasers, probability density function, remote sensing, vegetation mapping.

Manuscript received August 11, 2014; revised March 23, 2015; accepted May 26, 2015. This work was supported in part by the Portuguese Foundation for Science and Technology under Grant PTDC/AGR-CFL/72380/2006 and Grant Pest-OE/EEI/UI308/2014 and in part by the French National Research Agency (ANR) through the FORESEE Project under Grant ANR-2010-BIOE-008. The work of A. Ferraz was supported in part by the Jet Propulsion Laboratory through the NASA Postdoctoral Program, which was administrated by the Oak Ridge Associated Universities through a contract with NASA. The work of G. Gonçalves was supported in part by the Portuguese Foundation for Science and Technology under Grant UID/MULTI/00308/2013. The work of Margarida Tomé and Paula Soares was supported by the Portuguese Foundation for Science and Technology through the Forest Research Center Project under Grant UID/AGR/00239/2013.

A. Ferraz is with the NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA (e-mail: Antonio.A.Ferraz@jpl.nasa.gov).

C. Mallet is with the Institut National de l'Information Géographique et Forestière (IGN), SRIG, MATIS, Université Paris-Est, 94160 Saint Mandé, France (e-mail: clement.mallet@ign.fr).

S. Jacquemoud is with Institut de physique du globe de Paris, Unités Mixtes de Recherche (UMR) Centre National de la Recherche Scientifique (CNRS) 7154, Sorbonne Paris Cité, Université Paris Diderot, 75013 Paris, France (e-mail: jacquemoud@ipgp.fr).

G. R. Gonçalves is with the INESC-Coimbra and Department of Mathematics, University of Coimbra, 3001-501 Coimbra, Portugal (e-mail: gil@mat.uc.pt).

M. Tomé and P. Soares are with the Forest Research Center, School of Agronomy, University of Lisbon, 1349-017 Lisbon, Portugal (e-mail: magatome@isa.ulisboa.pt; paulasoares@isa.ulisboa.pt).

L. G. Pereira is with the Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro, 3754-909 Águeda, Portugal (e-mail: luisapereira@ua.pt).

F. Bretar is with Consulate General of France in Shanghai, Shanghai 200001, China (e-mail: Frederic.bretar@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2015.2448056

## I. INTRODUCTION

THERE is a growing need to produce maps representing the 3-D structure of vegetation at different time and spatial scales for long-term monitoring of forest ecosystems [1]. It implies a separate characterization of each vegetation layer, namely the overstory, the understory, and the ground vegetation. The overstory corresponds to the upper canopy (dominant, codominant and subdominant trees), whereas the understory (suppressed trees, young trees, and tall shrubs) generally grows in the shade of the overstory. Finally, the ground vegetation is located immediately above the soil and consists of small shrubs and herbaceous plants. Vegetation cover, i.e., the proportion of the forest floor covered by the vertical projection of tree crowns, is an important biophysical variable widely used to assess vegetation density, volume, and spatial distribution. Indeed, it is an input for a variety of process-based ecological and biochemical models to assess microclimate [2], wildlife habitat quality and biodiversity [3]–[5], carbon storage [6], [7], energy exchange, and evapotranspiration [8], [9]. Moreover, it plays an important role in classifying stand structure and determining fuel distribution, which both impact fire behavior [10]–[13]. Determining vegetation cover with conventional field techniques is expensive and time-consuming [14]. As a result, several remote sensing techniques have been investigated to map it. For instance, Carlson and Ripley [15] and Schlerf and Atzberger [16] studied various biomes with optical imagery. The implementation of such techniques on multilayered forests is nevertheless limited because they are unable to stratify vertical structures [1], [7]. A significant breakthrough has been made with the emergence of small-footprint airborne laser scanners (ALSs) that allow vertical stratification of forest ecosystems [17], [18]. They are an active remote sensing technique: A laser rangefinder on board a moving platform measures the distance to the Earth surface. The area illuminated by the laser beam, which is called the footprint, and usually ranges from 0.2 to 0.8 m in diameter [19]. The telescope, which is aligned with the laser, receives several echoes per pulse, which are converted into 3-D coordinates in the form of a point cloud. Therefore, this technique offers a unique capability to estimate vegetation cover for a single layer as the laser beams penetrate the vertical structure of the forest [11].

The forest vegetation cover can be described by means of various biophysical variables that describe the spatial pattern formed by the plant material and the vegetation gaps using different levels of detail [20], [21]. In this paper, we only investigate the crown cover (CrCo) defined as the total vertical

projection area of the tree crowns divided by the horizontal area of observation unit that the trees are growing on [22]. Overlapping crowns are not counted twice, and the crown area is defined by the outermost perimeter (envelope) of the crowns including within crown gaps.

Two methods have been used to extract CrCo from ALS data: the proportion-based method (PBM) and the grid-based method (GBM). In the first case, appropriate metrics allow to correlate CrCo to the fraction of echoes reflected within a given layer. Several metrics have been published in the literature [21], [23]–[25]. It is the most popular method because it is simple to implement and robust: It has already proved to be effective on many study sites. However, it computes CrCo at a coarse resolution, and results are biased if the plots do not stand vertically below the flight line. Such biases can be readily explained by an increase in the apparent canopy density when vegetation is measured at viewing angles greater than  $14^\circ$ . Therefore, the PBM does not take full advantage of the ALS ability to characterize forest structure. For instance, greater scanning angles may reduced the occultation of underneath vegetation by topmost vegetation and provide additional information [26]. Therefore, when applied to ALS surveys performed with spaced flight lines and wide scanning angles, for instance if cost issues are critical, such an approach may provide CrCo maps with variable accuracy according to the distance to the nearest flight line [27].

In GBMs, the echoes are projected on a binary grid with a certain resolution [28]. CrCo is computed as the number of cells containing data divided by the number of empty ones. Alternatively, Lee and Lucas [29] projected the ALS echoes on a high-resolution voxel grid. Then, CrCo is related to the relative penetration of lidar pulses into the canopy, which is assessed by means of the height-scaled crown openness index (HSCOI) expressed as a percentage. HSCOI calculates the echo distribution within each column of the grid cell. Near-zero values correspond to low penetration, i.e., dense vegetation. The success of GBM highly depends on the grid resolution: A coarse resolution does not preserve the sharp edges of vegetation, overestimating the crown areas and, consequently, CrCo, and a fine resolution usually leads to an underestimation because the ALS point cloud provides an irregular vegetation sampling, i.e., many cells located within the outermost perimeter of the crowns are empty. The method may be improved by applying an image processing operation such as smoothing. Empty cells surrounded by cells with data are assumed to bound crowns. For instance, Lee and Lucas [29] analyzed the  $n$  nearest neighbors of each empty cell to locate the outmost perimeter of the crowns. Morphological operators such as closing and opening [30] can be applied to binary images [21], [28]. However, contrary to the PBM, the GBM lacks robustness due to variability in the ALS pulse density. Vegetation features are measured at different spatial sampling rates so that the size of the smoothing filter should be optimized [21]. No automatic approach has been proposed so far to adapt the filter size to the pulse density. It is either tuned manually or calibrated using field data.

In this paper, we present a new tool to characterize forest structure called canopy density model (CDM). It is a smooth

vegetation density surface determined from ALS echoes using kernel density estimators (KDEs). The CDM approach is unbiased with respect to the scanning angle, which allows keeping all of the ALS measurements. Contrary to existing methods dedicated to interpolate canopy cover maps from ALS measurements, the degree of smoothness is automatically defined as a function of the ALS spatial sampling rate variability. The reliability of such normalized CDM is assessed in terms of CrCo estimates. We first present the experimental data (see Section II). Then, we introduce the KDE theory (see Section III-B) that allows deriving the CDM (see Section III-C). Next, CrCo is computed both from the CDM and existing methods (see Sections III-D and III-E). The reliability of all these methods is compared and discussed in Section IV.

## II. EXPERIMENT

### A. Study Area

The study area, which totals  $9 \text{ km}^2$  and rises from 70 to 220 m, is located near the city of Águeda in northwest Portugal ( $40^\circ 36' \text{ N}$ ,  $8^\circ 25' \text{ W}$ ). It is a woodland, predominantly composed of eucalyptus (*Eucalyptus globulus* Labill.), with some stands of maritime pine (*Pinus pinaster* Ait.), shrublands, and agricultural fields. It is characterized by a well-developed understory, typical in Mediterranean forests, as well as vigorous shrubby vegetation. The lowermost layer mainly contains suppressed trees (eucalyptus, maritime pine, acacia, and oak), gorse bush (*Ulex* spp.), heath (*Erica* spp.), carquesia (*Pterospartum* spp.), gum cistus (*Cistus* spp.), blackberry (*Rubus* spp.), broom (*Cytisus* spp.), ferns, and herbaceous plants.

### B. Field Data Collection

A forest inventory was specifically carried out over our study site for result validation purposes. The measurements were collected according to a field protocol set up by the Portuguese National Forest Inventory [31]. Forty-seven plots were identified for systematic sampling [32]. However, three of them were discarded due to the difficulty in accessing them; thus, in total, 44 plots have been studied: 42 with eucalyptus and 2 with maritime pines (plots #100 and #200, see Fig. 1). Each plot consists of two nested circles, an outer one ( $400 \text{ m}^2$ ) and an inner one ( $200 \text{ m}^2$ ), bounded using a distance measurer. A stand is defined as a uniform plant community in terms of species, age, and spatial arrangement [33]: Because the forest belongs to many landowners, a plot may contain more than one stand. In such a case, only the stand located in the center of the plot is described. To refine the forest boundaries, all the trees in the outer circle were numbered using a marker pen.

The individual trees (within the inner circle) and the forest layers (within the outer circle) have been carefully studied [34]. All individual trees higher than 2 m were characterized by means of their diameter at breast height (DBH), the tree height (TH), and their crown base height. DBH was measured using a caliper, and TH was measured using a telescopic stick or a Vertex hypsometer. Vertical stratification of the forest layers is appraised in terms of height: Seven classes have been defined (0–0.6, 0.6–1, 1–2, 2–4, 4–8, 8–16, and  $>16 \text{ m}$ ), which can

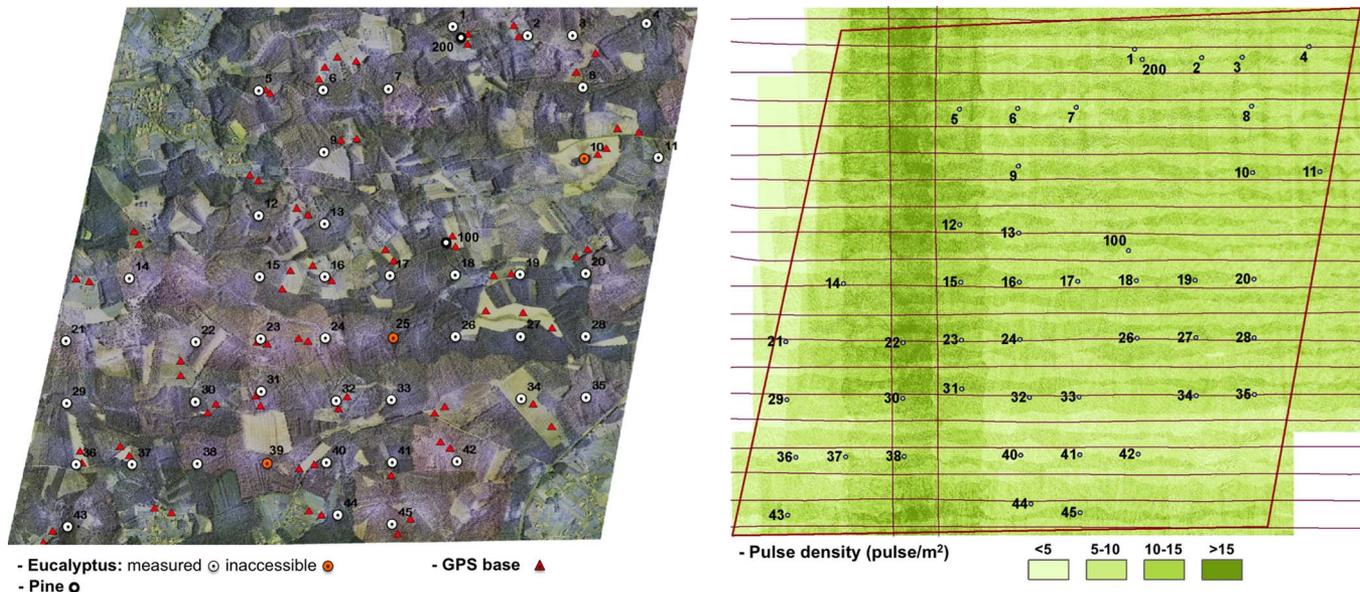


Fig. 1. (Left) Orthoimage and (Right) pulse density map (1 m  $\times$  1 m resolution) of the study area. The average pulse density within each forest stand equals 9.5 pulse/m<sup>2</sup> (min = 4.7 pulse/m<sup>2</sup>, max = 15.5 pulse/m<sup>2</sup>, and  $\sigma$  = 1.9 pulse/m<sup>2</sup>).

TABLE I  
 BIOPHYSICAL CHARACTERISTICS OF STAND #30

| Height class (m) | Layer             | Species           | % dominance | $\bar{h}$ (m) | CrCo (%) |
|------------------|-------------------|-------------------|-------------|---------------|----------|
| 0 – 2            | Ground vegetation | Ferns             | 95          | 1.2           | 50       |
|                  |                   | <i>Ulex</i>       | 5           |               |          |
| 2 – 8            | Understory        | <i>Acacia</i>     | 70          | 6             | 8        |
|                  |                   | <i>Pinus</i>      | 30          |               |          |
| > 8              | Overstory         | <i>Eucalyptus</i> | 100         | 21.2          | 20       |

be aggregated *in situ* to better describe the vegetation strata. In practice, the forest plots include a maximum of three layers that we named overstory, understory, and ground vegetation (see Table I). For each class, the CrCo, the relative dominance (species composition expressed as a percentage), and the vegetation mean height were assessed. To reduce the inherent subjectivity of a visual determination, the Portuguese manual adopted in this paper for forest inventory suggests that CrCo should be estimated by two field operators. If their appraisal is similar, they record the average value; otherwise, they shall reach an agreement.

The geolocation of information collected by forest inventory is usually less accurate than that of ALS systems. To improve it, a local geodetic network made of 41 pairs of GPS-derived points was built on the same map projection as the ALS data (see Fig. 1). All stand centers, as well as the tree positions within the outer circle, were surveyed using a tachymeter [35]. Finally, the data were integrated into a single 3-D geometry. More details of this experiment can be found in Gonçalves and Pereira [35] and Ferraz *et al.* [32].

### C. Characterization of the Stands

The eucalyptus stands are from 1 to 13 years old, younger than the two maritime pine stands that are 30 and 60 years old.

We distinguished 12 juvenile stands (1–4 years) and 32 mature stands (> 4 years). The latter may display three layers (ground vegetation, understory, and overstory), whereas the former usually display two layers (ground vegetation and overstory) or ground vegetation only (#9 and #34). In total, we identified 118 layers in the 44 forest stands: 44 ground vegetation, 32 understory, 10 juvenile overstory, and 32 mature overstory. Table II provides their characteristics. In Mediterranean forests, CrCo is quite variable: The overstory is generally fairly sparse, whereas the understory is well developed. The ground vegetation, which can reach up to 1.3 m high, is the densest layer. It often receives direct sunlight because many stands within the study area are crossed by roads or cleared [32]. On the contrary, many stands have little understory or ground vegetation because the owners removed the undergrowth.

### D. ALS Data

A plane equipped with a global positioning system/inertial measurement unit (GPS/IMU) device flew over the study area on July 14, 2008 (see Fig. 1). Lidar data were acquired along predetermined flight lines using a full-waveform ALS (LiteMapper 5600) with a scanning angle of  $\pm 22.5^\circ$ . Note that this angle is larger than those reported in forest studies [26]. For that reason, we have split the ALS echoes in two groups

TABLE II  
SINGLE-LAYER CHARACTERIZATION

|                    |                    | Minimum | Maximum | Mean  | Standard deviation |
|--------------------|--------------------|---------|---------|-------|--------------------|
| Mean height<br>(m) | Ground vegetation  | 0.15    | 1.3     | 0.53  | 0.3                |
|                    | Understory         | 0       | 6       | 2.41  | 1.64               |
|                    | Juvenile overstory | 2.2     | 6.32    | 4.16  | 1.43               |
|                    | Mature overstory   | 9.65    | 22.34   | 15.24 | 4.4                |
| CrCo<br>(%)        | Ground vegetation  | 2       | 100     | 52.1  | 33                 |
|                    | Understory         | 0       | 95      | 15.6  | 20.2               |
|                    | Juvenile overstory | 10      | 50      | 26.5  | 11.8               |
|                    | Mature overstory   | 10      | 90      | 33.91 | 18.46              |

TABLE III  
ALS ACQUISITION PARAMETERS

| ALS sensor                         | RIEGL LMS-Q560            |
|------------------------------------|---------------------------|
| Wavelength                         | 1550 nm                   |
| Scanning angle                     | $\pm 22.5^\circ$          |
| Pulse rate                         | 150 kHz                   |
| Effective measurement rate         | 75 kHz                    |
| Beam divergence                    | 0.5 mrad                  |
| Ground speed                       | 46.26 m/s                 |
| Flying level                       | 600 m                     |
| Scanning swath                     | 479 m                     |
| Strip overlap                      | 70%                       |
| Nominal distance between two lines | 150 m                     |
| Footprint diameter                 | 0.3 m                     |
| Single run density                 | 3.3 pulses/m <sup>2</sup> |
| Expected pulse density (epd)       | 9.9 pulses/m <sup>2</sup> |

corresponding to angles lower and higher than  $14^\circ$ . The combination of the scanning angle with the aboveground flight height determined the scanning swath (see Table III). Adjacent laser strips overlapped to guarantee wall-to-wall coverage. The GPS/IMU measurements were corrected by analyzing overlapping strips from the cross-strips (see Fig. 1). Detailed information about strip adjustment, geo-spatial data correction, and removal of systematic errors can be found in [35].

The operator of the lidar system supplied a point cloud obtained by processing the full-waveform data [36]. Each pulse gave rise to one to five echoes that were classified into first, last, intermediate, and single echoes. Ground and vegetation echoes were separated using TerraScan [37] to produce a Delaunay triangulation representing the digital terrain model (DTM). Then, the point cloud was normalized to calculate the actual height of the objects in the scene. Note that the points identified as ground were kept in the data set.

The expected pulse density (epd), i.e., the average number of emitted laser pulses per square meter, is a standard metric used in ALS surveys. It depends on the acquisition parameters and allows qualifying and comparing different data sets. The epd should determine the spatial sampling rate at which ALS measures the forest structure, but in reality, it differs from that specified in the project [38]. Fig. 1 shows the epd map over the

study area: The variability may be explained by the difficulty to fly along evenly spaced lines, by the scanning pattern, or by the number of overlapping strips. For instance, if the overlapping is high (70%), linear patterns are present in the image. Moreover, we kept cross-strip measurements, which highly increase the variability. On average,  $\text{epd} = 9.9 \text{ pulse/m}^2$ . In Section III-C1, we introduce the observed pulse density (opd), a variable intended to measure the actual pulse density locally.

### III. METHODS

This section is divided as follows. First, we describe the method developed by Ferraz *et al.* [32], which consists of extracting three forest canopy layers (overstory, understory, and ground vegetation) and delineating individual plant crowns. Note that the CDM-based method applies to each forest layer beforehand computed by this approach. Next, we introduce the kernel density estimation (KDE) theory, the main product of which is the CDM described in Section III-C. The way the CrCo is delineated from the CDM is detailed in Section III-D.

#### A. Single-Layer Extraction

Ferraz *et al.* [32] published the adaptive mean shift (AMS 3D) algorithm that segments ALS point clouds into individual crowns within single layers. The main results of their work are summarized here. When applied to the point cloud, the algorithm produces 3-D segments that correspond to individual vegetation features such as shrubs or tree crowns. Those segments are assigned to a given forest layer (see Fig. 2). Forest stratification has been validated over the 44 forest stands described in Section II-C. As for the mature overstory, the AMS 3D algorithm could extract 486 trees out of 701 (69.3% success rate). This rate actually depends on crown dominance: It equals 98.6%, 85.2%, 61.4%, and 12.8% for dominant, codominant, dominated, and suppressed eucalyptus, respectively, and 92.3% for pine trees. The mean height of juvenile overstory, understory, and ground vegetation was validated at the forest stand level. The method generated 6.8%, 15.6%, and 20% of outliers, respectively. The mean height of the remaining forest stands was extracted with a RMSE of 0.31, 0.96, and 0.15 m for juvenile overstory, understory, and ground vegetation. Finally, all the points classified as ground were assigned to the ground vegetation layer. In this paper, echoes lower than 0.1 m are considered as ground points. This threshold, which corresponds

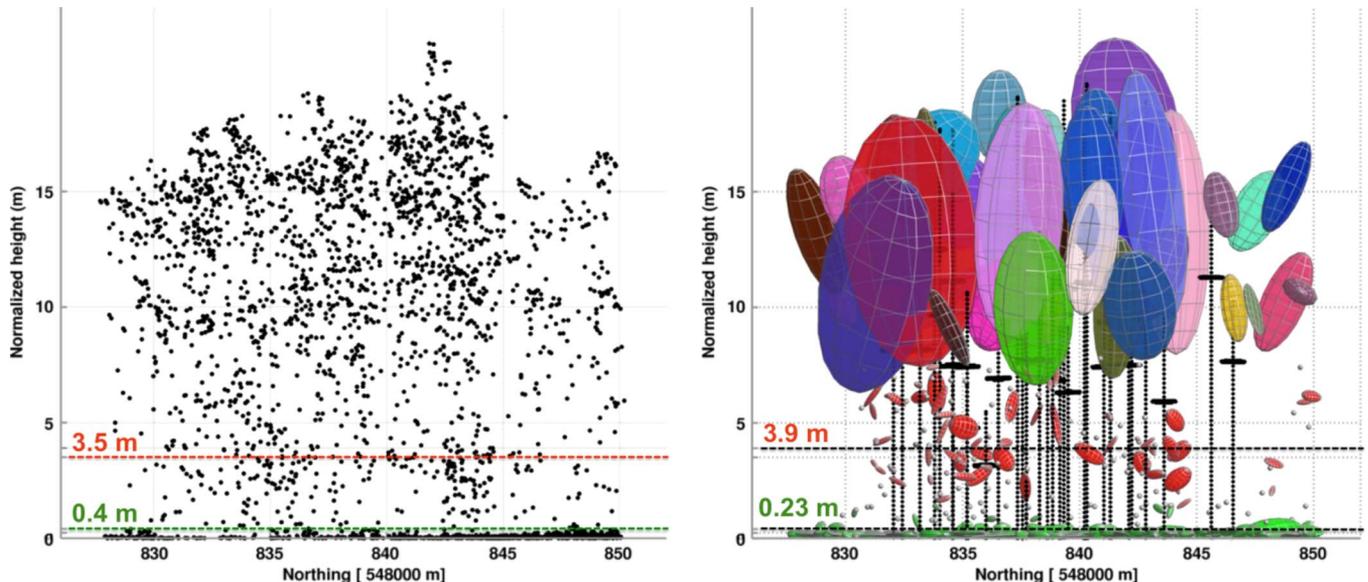


Fig. 2. (Left) ALS point cloud of stand #20 decomposed into (Right) 3-D segments. The multicolor, red, and green ellipsoids correspond to individual crowns within the overstory, understory, and ground vegetation, respectively. The surveyed trees are marked in black (dotted lines). The values correspond to (left) field-measured and (right) ALS-derived mean height for understory and surface vegetation.

to the mean height of the ground vegetation layer, is lower than the DTM vertical accuracy (0.12 m) determined in this area [35].

### B. KDEs

In statistics, the KDE is a nonparametric way to estimate the probability density function (pdf) of a random variable defined in a  $d$ -dimensional space  $\mathbb{R}^d$  [39]. This method relies on a single parameter, i.e., the kernel size, which is also called bandwidth and the choice of which is critical because it strongly impacts on the results. To limit such effects, we applied variable bandwidths that are suited to accounting for the local characteristics of the distribution. They proved to perform better than fixed bandwidths in many cases [40]. In the following, we describe a variable bandwidth approach based on the estimation of the sample point density [41]. Let  $X_i \in \mathbb{R}^d$  ( $i = 1, \dots, n$ ) be a point distribution with an unknown density. The shape of the pdf on a given point  $X \in X_i$  according to the KDE is computed as

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

where  $n$  is the number of samples,  $h(X_i)$  is the bandwidth,  $k(\cdot)$  is the kernel profile, and  $w(\cdot)$  is a weight function. A range of kernel profiles, such as the Laplacian, can be applied as follows:

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

The kernel profile is a radially symmetric function with the integral equal to 1. The contribution of the points to the pdf depends on their distance from the kernel center. Regardless of the kernel profile selected, the value of (1) increases with the number of points that are close to the kernel center. The

bandwidth controls the extent of the kernel profile: Data points located at a distance from the kernel center higher than  $h(X_i)$  have no or little influence on the pdf shape. Finally,  $w(\cdot)$  is an adjustable weight function that controls the weight of each data point into the pdf. It neither depends on the distance from the kernel center nor on the bandwidth. Moreover, it is not part of the original definition of a KDE; therefore, it is optional.

Fig. 3(a) shows the pdf of a random distribution defined in  $\mathbb{R}$  and computed according to (1) using the Laplacian kernel profile defined by (2), with  $h(X_i) = 1$  m and  $w(X_i) = 1$  for all  $X_i$ . In practice, the kernel is shifted along the space in order to calculate the density estimation for any location. The resulting pdf is a continuous function, the local maxima of which correspond to high point density. Naturally, the choice of the KDE parameters has an effect on the pdf. We will focus on the bandwidth that has a higher impact than the kernel profile [42]–[44]. Fig. 3(a) and (b) shows that larger bandwidths spread the influence of the distribution points. Therefore, the bandwidth may be viewed as a scale parameter that defines the smoothness of the density estimation. As for the weight function, it controls the influence of the distribution points in the pdf; therefore, its choice depends on the application. For example, if the distribution contains outliers, one can minimize their influence on the pdf by assigning them a low weight.

### C. CDMs

The KDE transforms discrete measurements into continuous smooth surfaces that represent spatial density variability. It has been used to identify animal home range and movement path [45], [46], analyze historical forest fire occurrence pattern [47] or seismic risk [48], visualize urban crime cluster [49], and identify road accident hotspots [50]. Here, it is tested for the first time on ALS data. Since our method deals with single layers, for the sake of clarity, the echoes  $X_i$  are divided into subsets  $X_{p,l,j} \subset X_i$  extracted by the AMS 3D method:  $p = 1, \dots, 44$

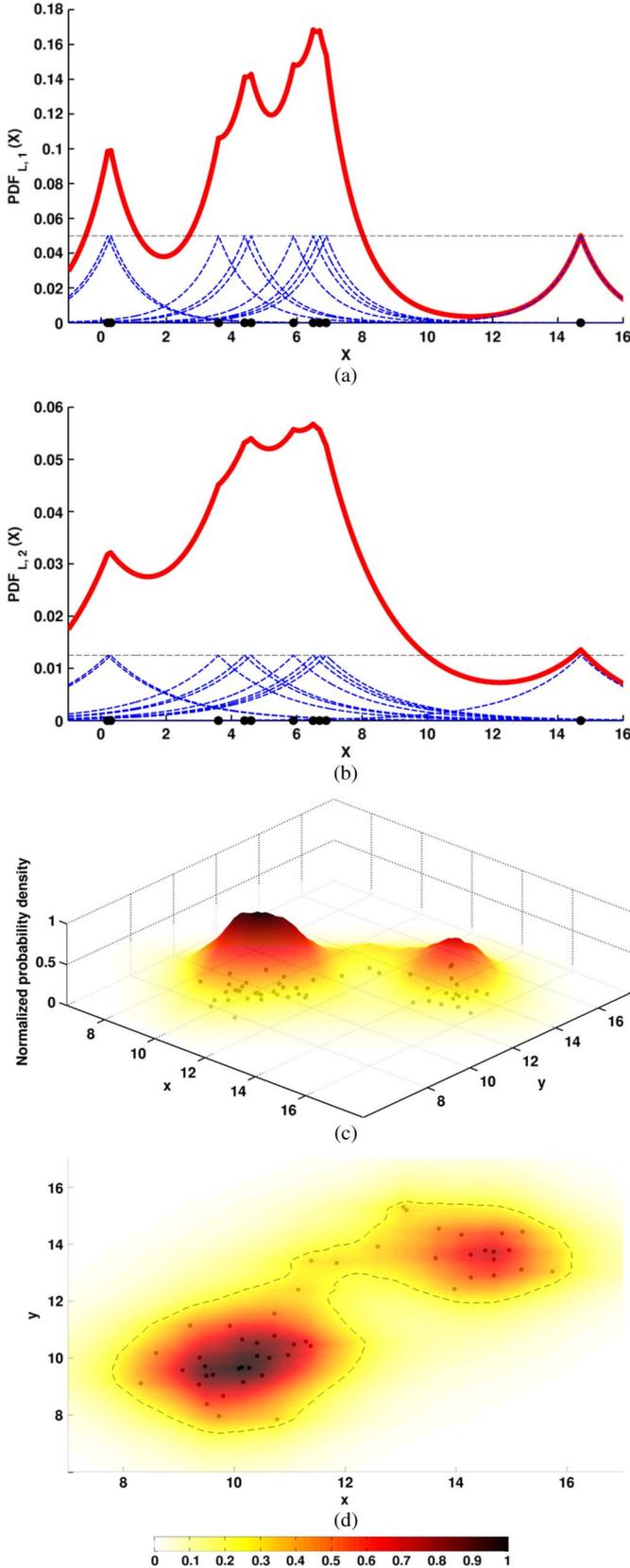


Fig. 3. Examples of a pdf in (a) and (b)  $\mathbb{R}$  and (c) and (d)  $\mathbb{R}^2$ . The bandwidth is set to 2 m in (b) and to 1 m in the other images. In (a) and (b), the point distribution, the Laplacian kernel profiles, and the pdf are represented by black dots, blue dashed lines, and a red line, respectively. (c) and (d) show a normalized pdf from a different point of view. The black dashed line in (a), (b), and (d) corresponds to a density contour used for CrCo estimates.

corresponds to the 44 forest stands,  $l = \{\text{os, us, gv, gr}\}$  refers to the forest stand layer (overstorey, understorey, ground vegetation, and ground), and  $j = 1, \dots, m_l$  represents the ALS echoes. Our method applies to  $\mathbb{R}^2$ ; thus,  $X_{p,l,j} = (x_{p,l,j}, y_{p,l,j}) \in \mathbb{R}^2$  are the planimetric coordinates of the ALS echoes. The CDM of a single layer is defined as

$$\begin{aligned} \text{CDM}(X) = \text{pdf}_{L,h(X_{p,l,j})}(X) &= \frac{1}{m \times h(X_{p,l,j})^2} \\ &\times \frac{1}{2 \times h(X_{p,l,j})} \times \sum_{i=1}^n w(X_{p,l,j}) \times \exp\left(-\frac{\|X - X_{p,l,j}\|}{h(X_{p,l,j})}\right). \end{aligned} \quad (3)$$

This equation follows from (1) by setting  $d = 2$ , where  $L$  stands for the Laplacian kernel function. As a result, the CDM is a pdf dedicated to computing the probabilistic models of vegetation density. It maps the spatial density variability by means of a smooth surface, the contours of which follow the ALS echo pattern. Fig. 3(c) and (d) shows the CDM that have been normalized by scaling between 0 and 1. They correspond to a simulated ALS echo distribution computed according to (3) with  $h(X_{p,l,j}) = 1$  m and  $w(X_{p,l,j}) = 1$  for all  $X_{p,l,j}$ . Therefore, a higher probability of vegetation occurrence is assigned to areas with higher echo densities. In the following, we will specify a variable bandwidth and a weight function both dedicated to compute single-layer CDM using ALS point clouds.

1) *Automatic Kernel Bandwidth Selection*: The CDM can be similarly implemented, but its effectiveness in producing reliable results is intricately linked to the selection of the bandwidth, which remains a scientific issue [51]. In our case, the pulse density variability and the occlusion effect introduced by the topmost vegetation should size the bandwidth [see Fig. 1(b)]. These two factors may cause either undersampling or oversampling of similar vegetation features depending on their spatial position or the forest vertical structure. It is likely that low vegetation growing under a dense overstorey be undersampled [52], [53]. We consequently designed a variable bandwidth depending on the local pulse density and the forest layer of interest, i.e.,

$$h(X_{p,l,j}) = h^* \times h^n(X_{p,l,j}) \quad (4)$$

where  $h^*$  is a default bandwidth, and  $h^n(X_{p,l,j})$  is a normalization factor driving the variable bandwidth settings. In this paper, the default bandwidth has been set to the nominal ALS pulse footprint  $h^* = 0.3$  m. Note that, although it was tuned empirically, this value is well suited to our experiment. The normalization factor is written as

$$h^n(X_{p,l,j}) = \frac{\text{epd}}{\text{opd}(X_{p,l,j})} \quad (5)$$

where  $\text{epd}$  is  $\text{epd} = 9.9$  pulse/m<sup>2</sup>, and the  $\text{opd}$  is calculated by

$$\text{opd}(X_{p,b,j}) = \frac{\sum_{l=b}^{\text{gr}} |X_{p,l,j}^{\text{single}}| + \sum_{l=b}^{\text{gr}} |X_{p,l,j}^{\text{first}}|}{A_p} \quad (6)$$

for a given layer  $b \in l$ .  $|\cdot|$  refers to the cardinality of a set, single and first stand for single and first echoes (see Section II-D),

and  $A_p$  is the area of the forest stand  $p$ . Note that the numerator in (6) accounts for the echoes assigned to the layer  $b$ , as well as those underneath including the ground. Contrary to  $epd$ , which is a theoretical value,  $opd$  measures the actual pulse density. One can assess the occlusion effect due to the overstorey with (6). Its impact on the sampling of the understorey can be estimated by setting both  $b = os$  and  $b = us$ . According to (6), the lack of single or first echoes that would suggest a lack of understorey do not have an impact on the estimation of the occlusion because the ALS pulses are supposed to hit either the ground vegetation or the ground.

We are well aware that equaling the spatial sampling rate of each layer to  $opd$  is an approximation. The reality might be different because the  $opd$  only considers single and first echoes, whereas the last and intermediate echoes increase the spatial sampling rate. In fact, the latter represent about 21% and 15% of the total number of echoes of ground vegetation and understorey, respectively.

2) *Weight Function*: The weight function aims at quantifying the influence of the echo position on the CDM. Low weights are assigned to echoes coming from crown boundaries, whereas high weights are assigned to echoes coming from the inside. The weight function aims at preventing a horizontal expansion of vegetation and at preserving the edges. To decide which echoes belong to the crowns or to their boundaries, we defined a function called  $vote(X_{p,l,j})$  that draws a circle centered in  $X_{p,l,j}$  with radius  $h(X_{p,l,j})$  [see (4)]. This circle is divided in four quadrants  $Q_t$  ( $t = 1, \dots, 4$ ). If  $X_{p,l,j}$  has no neighbor,  $vote(X_{p,l,j}) = 1$ ; if all the neighbors of  $X_{p,l,j}$  lay in a single quadrant,  $vote(X_{p,l,j}) = 2$ ; and if  $X_{p,l,j}$  has neighbors in two, three, or four quadrants, then  $vote(X_{p,l,j})$  equals 3, 4, or 5. The weight function is defined by

$$w(X_{p,l,j}) = \frac{vote(X_{p,l,j})}{5}, \quad vote(X_{p,l,j}) = \sum_{T=I\chi_{Q_t}}^{IV} (X_{p,l,j}) \quad (7)$$

where  $\chi_{Q_t}$  is the indicator function corresponding to the quadrant  $Q_t$

$$\chi_{Q_t}(X_{p,l,j}) := \begin{cases} 1 & \text{if } \exists X_{r,l,r} : \|X_{p,l,r} - X_{p,l,j}\| \leq h(X_{p,l,j}) \wedge X_{p,l,r} \in Q_t \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Therefore, the weight of echoes regularly distributed in their neighborhood is higher than that of isolated or boundary echoes.

#### D. CrCo Maps

CDMs are used to delineate the outmost perimeter of vegetation features in order to compute CrCo maps. The latter are drawn as a function of a VDT: the areas where  $CrCo > VDT$  are assumed covered by vegetation. As the KDE is a nonparametric technique that does not make any assumption about the shape of vegetation features, the contours of CrCo maps follow the pattern of the ALS measurements [see Fig. 4(a)

and (b)]. In our case, however, a constant VDT is not adapted for two reasons. First, the CDM are computed using variable bandwidths that produce variable probability density values [see vertical axes in Fig. 3(a) and (b)] because each kernel function has integral equal to 1 (see Section III-A). Second, the magnitude of the probability density values also depends on the number of echoes assigned to each forest layer [see  $m$  in (3)]. Therefore, to avoid discontinuities between adjacent areas and properly compare different CrCo maps, we calculate a normalized VDT so that isolated ALS echoes, which are meaningless in terms of forest characterization, do not contribute to the CrCo map. Such an assumption may also remove outliers. The normalized VDT is then written as

$$VDT_{L,h(X_{p,l,j})} = \frac{1}{m \times h(X_{p,l,j})^2} \times \frac{1}{2 \times h(X_{p,l,j})} \times \frac{1}{5} \quad (9)$$

which is derived from (3) for an isolated point, i.e., by setting  $w(X_i) = 1/5$  and  $\|X - X_i\| = 0$  m. The dashed lines in Fig. 3(a) and (b) show an example of a VDT. Finally, a CrCo map is a binary mask delineated on the CDM according to

$$\forall X \in \mathbb{R}^2, CDM(X) \geq VDT_{k,h(X_{p,l,j})}. \quad (10)$$

#### E. Result Assessment

We implemented the computation of single-layer CrCo maps (ground vegetation, understorey, and overstorey) in MATLAB and compared the results of our method with *in situ* measurements. The relationship is evaluated using a linear regression that beforehand identifies outliers [54]. In addition, we implemented two approaches described in Section III-E1: the PBM and the smallest enclosing ellipse (SEE) approach. On one hand, the reliability of our approach in estimating single-layer CrCo can be compared with existing methods. On the other hand, this cross comparison allows assessing the reliability of the field CrCo measurements, the accuracy of which is unknown (see Section II). For instance, it is recognized that the PBM (see Section I) is suitable for estimating the CrCo of the overstorey layer. As a result, as far as this layer is concerned, a strong relationship between field- and PBM-estimated CrCo ensures the reliability of the field-based method.

1) *State-of-the-Art Methods for CrCo Estimation*: The reliability of the CDM-based method to estimate CrCo is assessed both at stand and individual crown levels [see Fig. 4(a) and (b)]. In the first case, CrCo maps result from ALS echoes assigned beforehand to single layers (AMS 3D method, see Section III-A). Hereafter, this approach is simply called CDM. In the second case, CrCo maps result from ALS echoes corresponding to individual crowns also provided by the AMS 3D method (ellipsoids in Fig. 2, Section III-A). Hereafter, this approach is called IC-CDM for individual-crown CDM. Individual crowns have been merged to compare with field measurements, which have been carried out at the stand level. This avoids double-counting overlapping crowns. Note that the weights [see (7)] and VDT [see (9)] functions vary as a function of individual tree crown characteristics, whereas the variable bandwidth is computed at the forest stand level [see (5) and (6)].

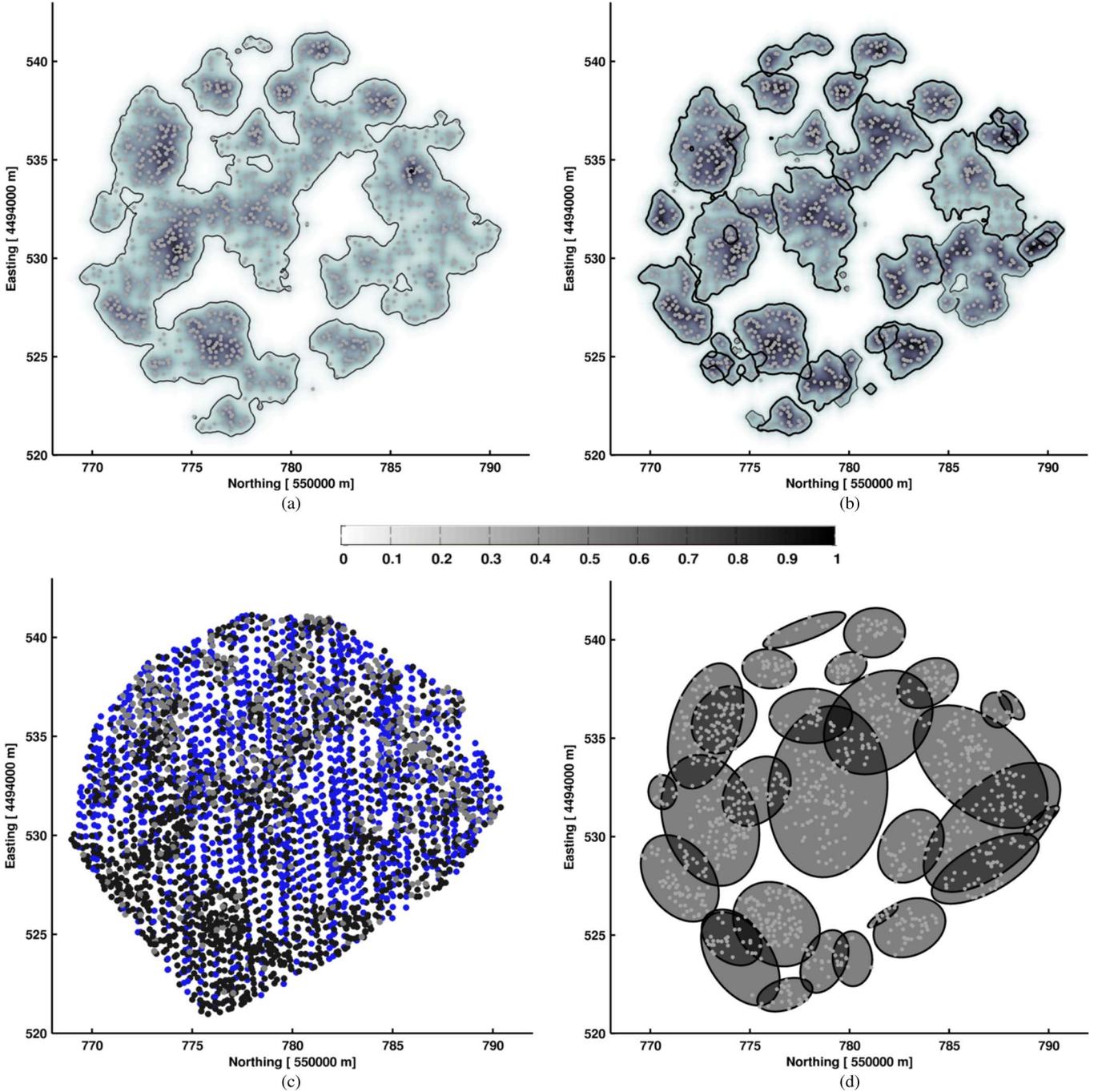


Fig. 4. CrCo estimation for the overstory layer (stand #28), computed using four approaches: (a) CDM, (b) IC-CDM, (c) PBM, and (d) SEE. Gray dots correspond to overstory echoes. In (c), the echoes in blue correspond to the overstory and those in black to scanning angles higher than  $14^\circ$ ; thus, they are not taken into account in the computation of CrCo. The gray level colorbar corresponds to the normalized CDM values.

In the following, we describe two state-of-the-art approaches to compute CrCo at the single layer and individual crown levels [see Fig. 4(c) and (d)]. The first one is a PBM, and the second one models individual crowns by means of the SEE algorithm. In the PBM, we adopt two different metrics. The first echo cover index (FCI) is used for the overstory and understory, i.e., [21]

$$0 \leq \text{PRM}^{\text{FCI}}(X_{p,b,j}) = \frac{|X_{p,b,j}^{\text{single}}| + |X_{p,b,j}^{\text{first}}|}{\sum_{l=b}^{\text{gr}} |X_{p,l,j}^{\text{single}}| + \sum_{l=b}^{\text{gr}} |X_{p,l,j}^{\text{first}}|} \leq 1 \quad (11)$$

with either  $b = \text{os}$  or  $b = \text{us}$ . The understory lidar cover density (ULCD) is used for ground vegetation [25], i.e.,

$$0 \leq \text{PRM}^{\text{ULCD}}(X_{p,b,j}) = \frac{|X_{p,b,j}|}{\sum_{l=b}^{\text{gr}} |X_{p,l,j}|} \leq 1 \quad (12)$$

with  $b = \text{gv}$ . In order to get unbiased results, only narrow ALS echoes colored in blue in Fig. 4(c) were used in (11) and (12). They correspond to scanning angles smaller than  $14^\circ$ , a value which agrees with Korhonen *et al.* [21] and Wing *et al.* [25].

TABLE IV  
OBSERVED PULSE CLOUD DENSITY (OPD, pulse/m<sup>2</sup>) AND VARIABLE-BANDWIDTH ( $h(X_{p,l,j}), m$ )

|                    | Minimum    |                | Maximum    |                | Average    |                | Standard deviation |                |
|--------------------|------------|----------------|------------|----------------|------------|----------------|--------------------|----------------|
|                    | <i>opd</i> | $h(X_{p,l,j})$ | <i>opd</i> | $h(X_{p,l,j})$ | <i>opd</i> | $h(X_{p,l,j})$ | <i>opd</i>         | $h(X_{p,l,j})$ |
| Mature overstory   | 5.8        | 0.18           | 15.4       | 0.47           | 10.3       | 0.31           | 2.6                | 0.08           |
| Juvenile overstory | 8.9        | 0.27           | 16.6       | 0.5            | 11         | 0.33           | 2.3                | 0.07           |
| Understory         | 5.3        | 0.16           | 14.4       | 0.43           | 9.2        | 0.28           | 2.4                | 0.07           |
| Ground vegetation  | 5.1        | 0.15           | 15.5       | 0.47           | 9.2        | 0.28           | 2.3                | 0.07           |

Equations (11) and (12) were the most reliable PBM metrics to compute CrCo of a single layer. They have been chosen out of many others reported in the literature [21], [23]–[25]. It is worth mentioning that the ULCD published by Wing *et al.* [25] applied to ALS points ranging from 0.25 to 1.45 m high. Thus, the layer named understory in that work corresponds to our ground vegetation (see Table II). Moreover, these authors applied a criterion based on the intensity of the ALS echoes to remove hits issued from unwanted features, such as nonvegetated items (rocks) or herbaceous plants. Here, we could not apply this technique because field measurements of CrCo also included herbaceous plants. Moreover, echoes displaying high intensity values may correspond to small shrub stems that are useful to characterize the layer in terms of CrCo [55].

In the SEE approach [see Fig. 4(d)], tree crowns were modeled by the perimeter of the smallest ellipse enclosing the echoes corresponding to individual crowns provided by the AMS 3D method (ellipsoids in Fig. 2, see Section III-A). Note that the IC-CDM and SEE approaches have not been used to model individual crowns in the ground vegetation layer. Indeed, the results would be biased because the AMS 3D method clusters the echoes corresponding to individual crowns together with ground echoes [32].

2) *Forest Stand Delineation*: While the CDM computed for this study are 400-m<sup>2</sup> circular regions, the field-based estimates correspond to areas that may display different shapes and sizes. In fact, if a forest plot includes more than one forest stand, only that located in the center of the plot is characterized (see Section II-B). To properly compare field- and CDM-derived CrCo estimates, the CDM are segmented using a mask, which is defined as a convex hull computed from the planimetric coordinates of the trees that had been surveyed. Moreover, a buffer zone with a radius of 1 m was drawn around the convex hull edges and included to the mask. Note that the PBM metrics are only computed using the ALS echoes within that mask.

## IV. RESULTS AND DISCUSSION

### A. CDMs

The CDMs were calculated for 118 layers: 44 ground vegetation, 32 understory, 10 juvenile overstory, and 32 mature overstory. Considering the variation of *opd*, we applied bandwidths ranging from 0.15 to 0.47 m (see Table IV). Fig. 5(a) and (b) shows the CDM of ground vegetation in stands #2 and #22, for which *opd* equals 9.1 and 15.5 pulses/m<sup>2</sup>, respectively. The corresponding bandwidths are  $h(X_{2,gv,j}) = 0.32$  m and  $h(X_{22,gv,j}) = 0.19$  m. It follows that large bandwidths applied to undersampled layers produce smooth CDM [see Fig. 5(a)],

whereas small bandwidths applied to oversampled layers produce rougher CDM [see Fig. 5(b)]. Its shape is also driven by the weight function: Isolated or edge echoes have less influence on the CDM than the others. As a result, the “horizontal expansion” of vegetation promoted by larger bandwidths is not uniformly distributed along the forest stands, which allows preserving the sharp edges of the vegetation cover. Fig. 5(a) clearly shows that the features of the forest track that crosses stand #2 are conserved.

Occlusion impacts on the spatial sampling rate at which single layers are measured: One expects that the ground vegetation be undersampled compared with the overstory [52], [56], [57]. Fig. 5(c) and (d) shows the overstory and ground vegetation of stand #12. They have been measured at different spatial sampling rates: *opd* = 13.3 pulses/m<sup>2</sup>, and *opd* = 7.6 pulses/m<sup>2</sup>, respectively. Therefore, the bandwidths are set to  $h(X_{12,ov,j}) = 0.22$  m and  $h(X_{12,gv,j}) = 0.39$  m. Applying a variable bandwidth favors ground vegetation echoes to the detriment of overstory echoes and counterbalances occlusion effects. Limits of that compensation for occlusion are discussed in Section IV-B1.

Results obtained for individual crowns (IC-CDM) are very similar to those obtained for stand level layers (CDM). The latter method provides a higher vegetation density when there are adjacent crowns that are often connected by isthmuses [see Fig. 4(a)], whereas they remain well separated with the first method [see Fig. 4(b)]. The higher estimates produced by the CDM is actually driven by the weight function. Many ALS echoes considered boundary in the IC-CDM can have additional neighbors from close crowns in the CDM.

### B. Single-Layer CrCo

1) *Stand Level*: The estimation of CrCo with CDM is very satisfactory, both for the mature and juvenile overstory (RMSE = 9.11% and RMSE = 8.54%, respectively) [see Fig. 6(a)]. The correlation is better for juvenile trees ( $R^2 = 0.85$  versus  $R^2 = 0.66$ ), but the outliers are far more numerous (20% versus 6.3%, Table V). The PBM is quite similar in terms of accuracy (RMSE = 9.18% versus 8.37%) [see Fig. 6(c)]. After the outliers are removed, the CrCo accuracy is similar to that reported in Korhonen *et al.* [21] who applied the same metric to Finnish boreal forests (RMSE = 7%). This result ensures the quality of the field-based CrCo measurements since the PBM is considered a robust approach over different ecosystems. Moreover, it indicates that the CDM-based approach can deal with ALS pulse density heterogeneity, similar to the PBM. It is unbiased regarding the pulse angle variability. Finally, it is worth noticing that both approaches are able to determine

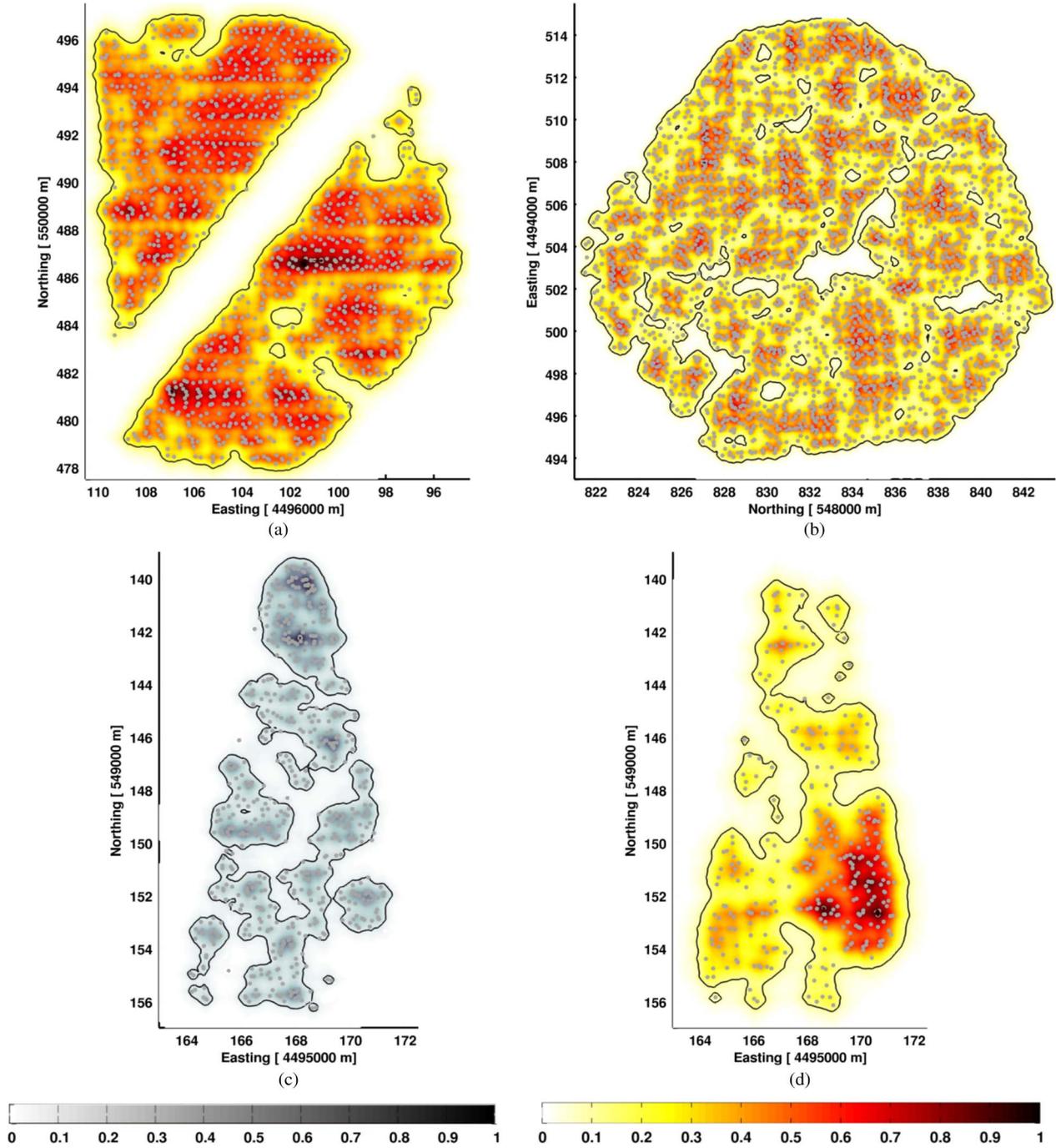


Fig. 5. CDMs for (a) stand #2 and (b) stand #22 ground vegetation. (c) and (d) correspond to the overstory and ground vegetation of stand #12. The ALS echoes are illustrated by gray dots. The colorbars correspond to the normalized CDM values. The VDTs used to delineate the CrCo maps are illustrated by the black lines.

the overstory CrCo in spite of the high rate of unidentified trees: 0.7%, 7.4%, 34.3%, and 85.5% of the total number of dominant, codominant, dominated, and suppressed eucalyptus trees are not represented in the point cloud. As for the pines, the proportion of unidentified trees only equals 2.9% [32]. This shows that the unidentified trees, which are mainly small trees, have little influence on the forest stand CrCo.

As for the understory, the CDM performs better than the PBM. Both methods generate 12.5% of outliers, but the correlation is higher with the first one ( $R^2 = 0.85$  instead of  $R^2 = 0.75$ ).

In addition, it reduces the RMSE by about 5% and the CrCo estimation by 3.84% (see Table V). However, the most significant improvement of the CDM approach concerns the ground vegetation layer. The RMSE decreases from 26.6% to 13.8%, and the CrCo is underestimated only by 2.5%, which represents a decrease of 15.9% with regard to the PBM. Both techniques provide high correlation coefficients ( $R^2 = 0.84$  and  $R^2 = 0.73$ , respectively, for CDM and PBM). Two reasons may explain the bad results obtained with the PBM approach. First, the echoes corresponding to scanning angles larger than  $14^\circ$

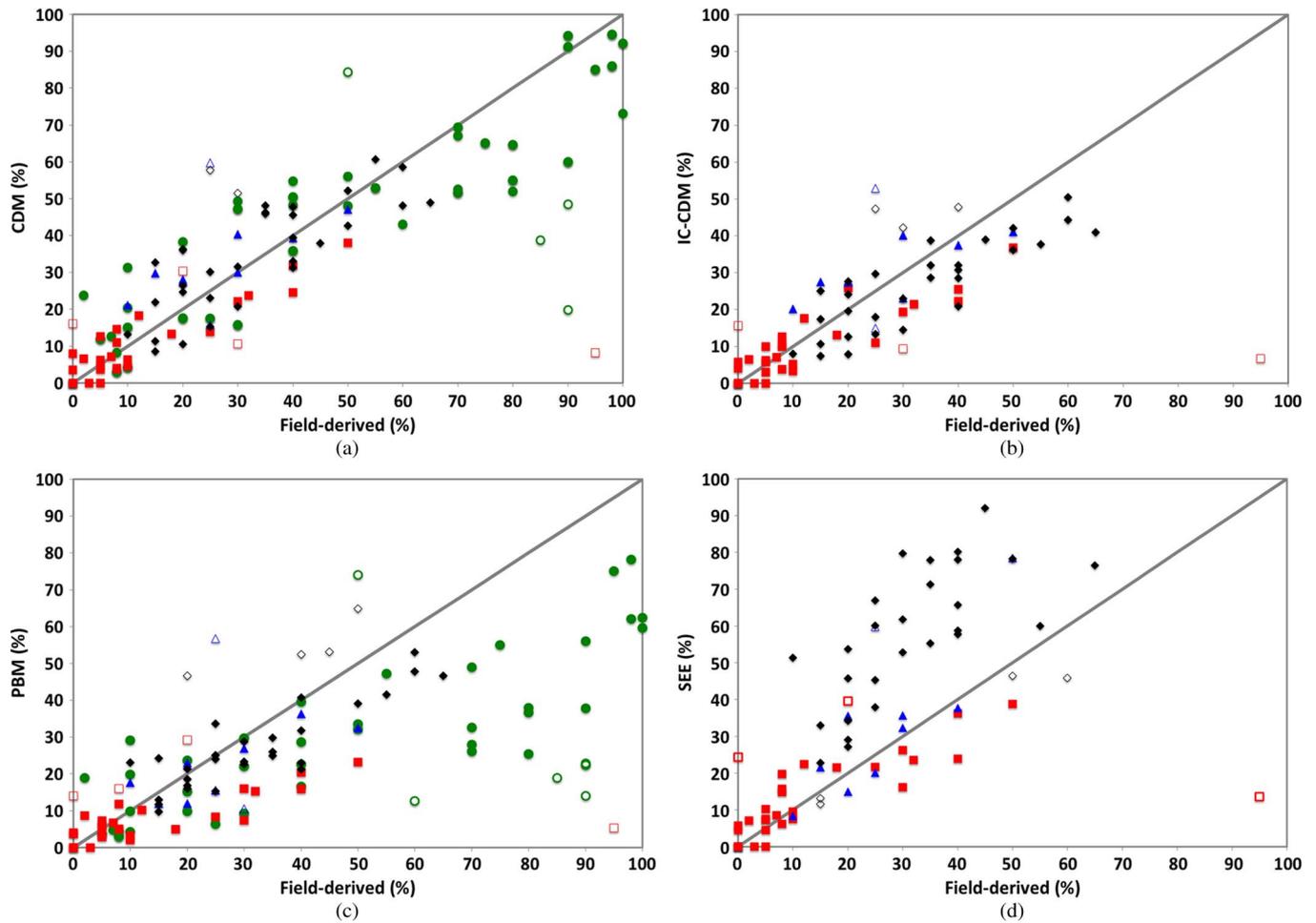


Fig. 6. ALS-derived versus field measured CrCo estimates for: (a) CDM, (b) IC-CDM, (c) PBM, and (d) SEE. Diamonds, triangles, squares, and dots correspond to mature overstory, juvenile overstory, understory, and ground vegetation, respectively. Note that the study of ground vegetation was not carried out in (b) and (d).

TABLE V  
LINEAR REGRESSION PARAMETERS ACCORDING TO FIG. 6. NEGATIVE VALUES MEAN UNDERESTIMATION

| Layer (n=118)                       |        | Outliers (%) | R <sup>2</sup> | RMSE (%) | Δ (%)  |
|-------------------------------------|--------|--------------|----------------|----------|--------|
| <b>Mature overstory</b><br>(n=32)   | CDM    | 6.3          | 0.66           | 9.11     | 1.14   |
|                                     | PBM    | 12.5         | 0.75           | 9.18     | -5.09  |
|                                     | IC-CDM | 9.4          | 0.73           | 10.47    | -6.69  |
|                                     | SEE    | 12.5         | 0.56           | 29.81    | 26.69  |
| <b>Juvenile overstory</b><br>(n=10) | CDM    | 20           | 0.85           | 8.54     | 6.06   |
|                                     | PBM    | 20           | 0.68           | 8.37     | -4.28  |
|                                     | IC-CDM | 20           | 0.64           | 8.71     | 3.69   |
| <b>Understory</b><br>(n=32)         | SEE    | 10           | 0.78           | 11.5     | 6.19   |
|                                     | CDM    | 12.5         | 0.85           | 6.21     | -1.89  |
|                                     | PBM    | 12.5         | 0.75           | 10.95    | -5.73  |
| <b>Ground<br/>Vegetation</b> (n=44) | IC-CDM | 9.4          | 0.74           | 7.54     | -1.78  |
|                                     | SEE    | 9.4          | 0.80           | 6.47     | -0.21  |
|                                     | CDM    | 9.1          | 0.84           | 13.76    | -2.49  |
|                                     | PBM    | 16.2         | 0.73           | 26.6     | -18.37 |

have been removed, which represents 38.7% of all the ALS points assigned to the ground vegetation. Second, contrary to the PBM, occlusion caused by the topmost vegetation is counterbalanced by the variable bandwidth in the CDM. Note that such a correction may fail in other sites or using other ALS

data sets. For instance, if the topmost vegetation is denser, the underneath vegetation could be either not or poorly observed and our empirical correction would be very limited. In such a case, a more sophisticated approach would be required. Our method could be also coupled with a physical-based model for

TABLE VI  
ADVANTAGES AND DRAWBACKS OF FOUR METHODS FOR CrCo ESTIMATES: CDMs, PBMs, IC-CDMs, AND SEE

| Method | Advantages  | Drawbacks   |
|--------|---|---|
| CDM    | <ul style="list-style-type: none"> <li>- High spatial resolution</li> <li>- Reliable for single layer estimation</li> <li>- Does not discard any ALS measurement</li> <li>- Takes into account the occlusion effects</li> </ul> | <ul style="list-style-type: none"> <li>- Variable bandwidth settings need to be validated over other sites</li> <li>- Compensation for the occlusion effect needs to be assessed over forest with denser canopies.</li> </ul>                       |
| PBM    | <ul style="list-style-type: none"> <li>- Easy to implement</li> <li>- High skills dealing with the ALS pulse density variability</li> <li>- Proven reliability over different sites</li> </ul>                                  | <ul style="list-style-type: none"> <li>- Poor spatial resolution</li> <li>- Poor results for ground vegetation</li> <li>- Discard ALS measurements (scanning angle &gt; 14°)</li> <li>- Does not take into account the occlusion effects</li> </ul> |
| IC-CDM | <ul style="list-style-type: none"> <li>- Able to model crown variability in terms of shape and size</li> </ul>  | <ul style="list-style-type: none"> <li>- Variable bandwidth settings need to be validated over other sites</li> </ul>   |
| SEE    | <ul style="list-style-type: none"> <li>- Able to model shrub and juvenile tree crowns</li> </ul>  | <ul style="list-style-type: none"> <li>- Unable to deal with crown shape variability</li> <li>- Higher computational time</li> </ul>  |

the compensation of ALS pulse transmission loss [53]. Finally, Wing *et al.* [25] who applied the PBM metric to a Californian multilayered forest obtained better results. The point cloud generated using a discrete multireturn ALS device was not as dense as ours (6.9 pulses/m<sup>2</sup>,  $\sigma = 5.6$  pulses/m<sup>2</sup>), and they removed beforehand echoes assigned to the ground vegetation. Thus, assuming that the vertical structure of forests is similarly described in the two studies, we would not expect that a denser point cloud underestimate so much ground vegetation. We believe that vegetation characteristics strongly impact on the number of recorded echoes. In the Californian forest, this layer is covered by coarse woody debris and prominent stumps, which strongly backscatter the laser beams.

2) *Individual Crown Level*: The two methods discussed here (IC-CDM and SEE) calculate CrCo at the individual crown level [see Fig. 4(b) and (d)]. As aforementioned, crown-level estimates are summed up at the forest stand level. The RMSE is the same for IC-CDM and CDM. However, the former method tends to underestimate CrCo [see Table V and Fig. 6(a) and (b)]. This was expected because the number of points receiving lower weights is higher in the IC-CDM approach [see (7)]. Indeed, results show that the mature overstory displays the highest biases: Tree crowns are more likely to be in contact in the mature overstory than in the juvenile overstory or the understorey. As a result, isthmuses between individual crowns are much more present in the CDM, which increases CrCo estimation [see Fig. 4(b) and (d)]. The CDM produces lower RMSE than the IC-CDM: It agrees with field measurements since field operators estimated CrCo at the forest stand level rather than at the individual crown level (see Section II-B). Still, results indicate that the IC-CDM is a reliable technique to model the outmost perimeter of individual crowns independently of the layer.

As far as the SEE method is concerned, results are highly dependent on the forest layer [see Fig. 6(b)]. The RMSE increases as plants grow: 6.47%, 11.5%, and 29.81% for understorey, juvenile overstorey, and mature overstorey (see Table V). CrCo estimations show a similar trend: -0.21%, 6.19%, and 26.69%, respectively. The SEE method produces slightly more accurate results with understorey trees, indicating that a parametric approach can correctly model the crown of bushes and juvenile trees. However, it fails with adult trees, which reveals that trees

diverge from a conventional geometrical shape during their growth. Due to the canopy tree competition, the crowns within a forest stand display very different shapes and hardly follow a unique geometric pattern. As a result, a nonparametric method such as the IC-CDM is better suited to model individual crowns.

3) *Comparison of the Methods*: The main advantages and drawbacks of the four methods are summarized in Table VI. Two reasons explain the success of the PBM to estimate CrCo: It is easy to implement and deals quite well with the pulse density variability of the ALS data. It also proved to be effective over different sites. However, in this paper, it was not able to estimate ground vegetation CrCo. Finally, it mapped the CrCo at a coarser spatial resolution.

As far as the overstorey and understorey are concerned, the CDM approach produced similar results compared with the PBM. It was able to reasonably figure out ground vegetation CrCo. It dealt well with occlusion because it did not discard any ALS measurement due to the variable bandwidth. Finally, CrCo was mapped at a fine spatial resolution. Still, the variable bandwidth settings used in this paper should be validated over other sites and using different ALS data.

The SEE approach is not suited to model adult trees probably because there are too many constraints on the crown shape. In natural conditions, these shapes are highly irregular. Moreover, the computation time is higher compared with the other methods. The SEE method takes about 2 min to compute the CDM for the aforementioned 40 forest stands, whereas the remaining ones about 30 s. On the contrary, the nonparametric IC-CDM approach is better suited to model individual trees. In this case, the outmost perimeter of crowns follows the ALS echo pattern and not a standard geometrical shape.

## V. CONCLUSION

We have proposed a new ALS-derived product, i.e., the CDM, dedicated to vegetation density and cover characterization. Its reliability was assessed in terms of estimating the CrCo maps. The latter were outlined from the CDM, assuming that the vegetation density is highest in areas within the outmost perimeter of plant crowns than in areas between crowns. Our nonparametric approach produce realistic CrCo maps where the edges of vegetation follow the ALS echo distribution pattern.

Results indicate that the CDM-based approach is robust to the variations in ALS pulse angle and able to deal with the variability in the spatial sampling rate at which the ALS measures a forest. The optimization of the degree of smoothing is still a major issue in the field of surface interpolation from ALS measurements. For instance, the widely used canopy height models (CHMs) have been interpolated using a survey- and application-dependent degree of smoothing. Most so often, a single degree has been applied to the entire study site. In this paper, we have introduced for the first time a multiscale smoothing filter designed to locally accommodate for the pulse density variability, i.e., to the heterogeneity of sampling rate inherent to every ALS survey. Our approach deals well with the sampling rate variability in both directions: horizontal (e.g., due to the number of overlapping strips) and vertical ones (due to the occlusion effect induced by the topmost vegetation). However, a validation over different study sites and using field estimations provided by more reliable techniques is still necessary to assess the robustness and scalability of the method. Indeed, it stills unclear whether the default parameters set in this experiment with respect to the variable bandwidth initialization and the density threshold are optimal to apply to other data sets. Nevertheless, similar to other ALS-based methods and techniques, such as the extraction of individual trees from a CHM, calibration of parameters may be needed to apply to different forest types.

Finally, we believe that the CDM can be used jointly with the CHM to improve ALS forestry. In the near future, we intend to investigate the complementarity of both products in segmenting individual trees by applying image processing techniques to a joint feature space.

## REFERENCES

- [1] F. G. Hall *et al.*, "Characterizing 3D vegetation structure from space: Mission requirements," *Remote Sens. Environ.*, vol. 115, no. 15, pp. 2753–2775, Nov. 2011.
- [2] T. R. Rambo and M. P. North, "Canopy microclimate response to pattern and density of thinning in a Sierra Nevada forest," *Forest Ecol. Manage.*, vol. 257, no. 2, pp. 435–442, Jan. 2009.
- [3] H. Niska *et al.*, "Neural networks for the prediction of species-specific plot volumes using airborne laser scanning and aerial photographs," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1076–1085, Mar. 2010.
- [4] L. A. Venier and J. L. Pearce, "Boreal forest landbirds in relation to forest composition, structure, and landscape: Implications for forest management," *Can. J. Forest Res.*, vol. 37, no. 7, pp. 1214–1226, Jul. 2007.
- [5] S. Martinuzzi *et al.*, "Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability," *Remote Sens. Environ.*, vol. 113, no. 12, pp. 2533–2546, Dec. 2009.
- [6] M. Vastaranta *et al.*, "TerraSAR-X Stereo radargrammetry and airborne scanning LiDAR height metrics in imputation of forest aboveground biomass and stem volume," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1197–1204, Feb. 2014.
- [7] G. Patenaude, R. Milne, and T. Dawson, "Synthesis of remote sensing approaches for forest carbon estimation: Reporting to the Kyoto Protocol," *Environ. Sci. Policy*, vol. 8, no. 2, pp. 161–178, Apr. 2005.
- [8] W. Ni-Meister, D. L. B. Jupp, and R. Dubayah, "Modeling lidar waveforms in heterogeneous and discrete canopies," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 9, pp. 1943–1958, Sep. 2001.
- [9] S. Youngkeun *et al.*, "Spectral correction for the effect of crown shape at the single-tree level: An approach using a Lidar-derived digital surface model for broad-leaved canopy," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 755–764, Feb. 2013.
- [10] D. V. Sandberg, R. D. Ottmar, and G. H. Cushon, "Characterizing fuels in the 21st century," *Int. J. Wildland Fire*, vol. 10, no. 4, pp. 381–387, 2001.
- [11] F. Morsdorf *et al.*, "Discrimination of vegetation strata in a multilayered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning," *Remote Sens. Environ.*, vol. 114, no. 7, pp. 1403–1415, Jul. 2010.
- [12] M. Garcia, D. Riaño, E. Chuvieco, J. Salas, and F. M. Danson, "Multispectral and LiDAR data fusion for fuel type mapping using Support Vector Machine and decision rules," *Remote Sens. Environ.*, vol. 115, no. 6, pp. 1369–1379, Jun. 2011.
- [13] M. del Rosario Iglesias, A. Barchuk, and M. P. Grilli, "Carbon storage, community structure and canopy cover: A comparison along a precipitation gradient," *Forest Ecol. Manage.*, vol. 265, pp. 218–229, Feb. 2012.
- [14] A. C. S. Fiala, S. L. Garman, and A. N. Gray, "Comparison of five canopy cover estimation techniques in the western Oregon Cascades," *Forest Ecol. Manage.*, vol. 232, no. 1–3, pp. 188–197, Aug. 2006.
- [15] T. N. Carlson and D. A. Ripley, "On the relation between NDVI, fractional vegetation cover, and leaf area index," *Remote Sens. Environ.*, vol. 62, no. 3, pp. 241–252, Dec. 1997.
- [16] M. Schlerf and C. Atzberger, "Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data," *Remote Sens. Environ.*, vol. 100, no. 3, pp. 281–294, Feb. 2006.
- [17] J. Hyyppä, O. Kelle, M. Lehtikoinen, and M. Inkinen, "A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 5, pp. 969–975, May 2001.
- [18] C. S. Lo and C. Lin, "Growth-competition-based stem diameter and volume modeling for tree-level forest inventory using airborne LiDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2216–2226, Apr. 2013.
- [19] C. Mallet and F. Bretar, "Full-waveform topographic lidar: State-of-the-art," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 1, pp. 1–16, Jan. 2009.
- [20] S. B. Jennings, N. D. Brown, and D. Sheil, "Assessing forest canopies and understory illumination: Canopy closure, canopy cover and other measures," *Forestry*, vol. 72, no. 1, pp. 59–74, 1999.
- [21] 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 Sens. Environ.*, vol. 115, no. 4, pp. 1065–1080, Apr. 2011.
- [22] A. Gonsamo, P. D'odorico, and P. Pellikka, "Measuring fractional forest canopy element cover and openness—Definitions and methodologies revisited," *Oikos*, vol. 122, no. 9, pp. 1283–1291, Sep. 2013.
- [23] C. Hopkinson, "The influence of flying altitude, beam divergence, and pulse repetition frequency on laser pulse return intensity and canopy frequency distribution," *Can. J. Remote Sens.*, vol. 33, no. 4, pp. 312–324, Aug. 2007.
- [24] S. Solberg, "Mapping gap fraction, LAI and defoliation using various ALS penetration metrics," *Int. J. Remote Sens.*, vol. 31, no. 5, pp. 1227–1244, Feb. 2010.
- [25] B. M. Wing *et al.*, "Prediction of understory vegetation cover with airborne lidar in an interior ponderosa pine forest," *Remote Sens. Environ.*, vol. 124, pp. 730–741, Sep. 2012.
- [26] M. A. Wulder *et al.*, "Lidar sampling for large-area forest characterization: A review," *Remote Sens. Environ.*, vol. 121, pp. 196–209, Jun. 2012.
- [27] 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 Sens. Environ.*, vol. 104, no. 1, pp. 50–61, Sep. 2006.
- [28] Y. Wang, H. Weinacker, and B. Koch, "A lidar point cloud based procedure for vertical canopy structure analysis and 3D single tree modelling in forest," *Sensors*, vol. 8, no. 3938, pp. 3938–3951, Jun. 2008.
- [29] A. C. Lee and R. M. Lucas, "A LiDAR-derived canopy density model for tree stem and crown mapping in Australian forests," *Remote Sens. Environ.*, vol. 111, no. 4, pp. 493–518, Dec. 2007.
- [30] J. P. Serra, *Image Analysis and Mathematical Morphology*, vol. 1. London, U.K.: Academic, 1982.
- [31] Instruções para o trabalho de campo do Inventário Florestal Nacional, Divisão para a Intervenção Florestal, Direção de Unidade de Gestão Florestal, Divisão para a Intervenção Florestal, Autoridade Florestal Nacional, Lisboa, Portugal, 2009.
- [32] A. Ferraz *et al.*, "3-D mapping of a multilayered Mediterranean forest using ALS data," *Remote Sens. Environ.*, vol. 121, pp. 210–223, Jun. 2012.
- [33] J. Stokes, C. Ashmore, L. Rawlins, and L. Sirois, "Glossary of terms used in timber harvesting and forest engineering," U.S. Dept. Agric., Forest Serv., Southern Forest Experiment Station, New Orleans, LA, USA, Gen. Tech. Rep. SO-73, 1989.

- [34] L. Pereira and G. Gonçalves, "Accuracy of a DTM derived from full-waveform laser scanning data under unstructured eucalypt forest: A case study," presented at the International Federation of Surveyors Congress, Sydney, NSW, Australia, 2010.
- [35] G. Gonçalves and L. Pereira, "A thorough accuracy estimation of DTM produced from airborne full-waveform laser scanning data of unmanaged Eucalyptus plantations," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3256–3266, Aug. 2012.
- [36] RiANALYZE, Horn, Austria, RIEGL, 2011.
- [37] A. Soininen, TerraScan User's Guide. Helsinki, Finland: TerraSolid, 2011. (Accessed: 6/07/2011), ed. [Online]. Available: [http://www.terrasolid.fi/system/files/tscan\\_2.pdf](http://www.terrasolid.fi/system/files/tscan_2.pdf)
- [38] E. Baltsavias, "Airborne laser scanning: Basic relations and formulas," *ISPRS J. Photogramm. Remote Sens.*, vol. 54, no. 2/3, pp. 199–214, Jul. 1999.
- [39] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.
- [40] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background detection," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, May 2006.
- [41] G. R. Terrel and D. W. Scott, "Variable kernel density estimation," *Ann. Stat.*, vol. 20, no. 3, pp. 1236–1265, Sep. 1992.
- [42] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *J. Amer. Stat. Assoc.*, vol. 91, no. 433, pp. 401–407, Mar. 1996.
- [43] D. Comaniciu, "An algorithm for data-driven bandwidth selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 281–288, Feb. 2003.
- [44] M. Singh and N. Ahua, "Regression based bandwidth selection for segmentation using Parzen windows," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Los Alamitos, CA, USA, 2003, pp. 2–9.
- [45] W. M. Getz and D. Saltz, "A framework for generating and analyzing movement paths on ecological landscapes," *Proc. Nat. Acad. Sci.*, vol. 105, no. 49, pp. 19066–19071, Dec. 2008.
- [46] S. Steiniger and A. J. S. Hunter, "A scaled line-based kernel density estimator for the retrieval of utilization distributions and home ranges from GPS movement tracks," *Ecol. Inf.*, vol. 13, pp. 1–8, Jan. 2013.
- [47] N. Koutsias, K. D. Kalabokidis, and B. Allgöwer, "Fire occurrence patterns at the landscape level: Beyond positional accuracy of ignition points with kernel density methods," *Nat. Resour. Modeling*, vol. 17, no. 4, pp. 359–375, Dec. 2004.
- [48] M. Danese, M. Lazzari, and B. Murgante, "Kernel density estimation methods for a geostatistical approach in seismic risk analysis: The case study of potenza hilltop town (Southern Italy)," in *Computational Science and Its Applications—ICCSA 2008*, vol. 5072, O. Gervasi, B. Murgante, A. Lagana, D. Taniar, Y. Mun, and M. Gavrilova, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 415–429.
- [49] T. Nakaya and K. Yano, "Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics," *Trans. in GIS*, vol. 14, no. 3, pp. 223–239, Jan. 2010.
- [50] T. K. Anderson, "Kernel density estimation and K-means clustering to profile road accident hotspots," *Accid. Anal. Prev.*, vol. 41, no. 3, pp. 359–364, May 2009.
- [51] N.-B. Heidenreich, A. Schindler, and S. Sperlich, "Bandwidth selection for kernel density estimation: A review of fully automatic selectors," *ASTA Adv. Stat. Anal.*, vol. 97, no. 4, pp. 403–433, Oct. 2013.
- [52] D. Riaño, E. Meier, B. Allgöwer, E. Chuvieco, and S. Ustin, "Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling," *Remote Sens. Environ.*, vol. 86, no. 2, pp. 177–186, Jul. 2003.
- [53] I. Korpela, A. Hovi, and F. Morsdorf, "Understorey trees in airborne LiDAR data—Selective mapping due to transmission losses and echo-triggering mechanisms," *Remote Sens. Environ.*, vol. 119, pp. 92–104, Apr. 2012.
- [54] P. J. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 1981.
- [55] A. Ferraz *et al.*, "Comparing small-footprint ALS and forest inventories data for single strata biomass estimations—A case study over a multilayered mediterranean forest," in *Proc. IEEE IGARSS*, Munich, Germany, Jul. 2012, pp. 6384–6387.
- [56] J. E. Means *et al.*, "Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western cascades of Oregon," *Remote Sens. Environ.*, vol. 67, no. 3, pp. 298–308, Mar. 1999.
- [57] M. A. Lefsky *et al.*, "Lidar remote sensing of the canopy structure and biophysical properties of Douglas-Fir Western Hemlock forests," *Remote Sens. Environ.*, vol. 70, no. 3, pp. 339–361, Dec. 1999.



**António Ferraz** was born in Antas de Penedono, Portugal. He received the M.S. degree in geomatics engineering from Universidade de Coimbra, Coimbra, Portugal, in 2007 and the Ph.D. degree in geophysics from Institute de Physique du Globe de Paris, Paris, France, in 2012.

From November 2007 to January 2014, he was with the French National Institute for Geographic Information and Forestry, Paris. Since May 2014, he has been with the NASA Postdoctoral Program at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. His main research interests include remote sensing of vegetation, optical image and 3-D point cloud processing, direct retrieval of forest variables using airborne laser scanning systems, forest 3-D modeling, and carbon and biomass estimation over tropical environments.



**Clément Mallet** received the Master of engineering degree from Ecole Nationale des Sciences Géographiques, Marne-la-Vallée, France, and the M.Sc. degree in physics in remote sensing from Université Pierre et Marie Curie-Paris 6, Paris, France, in 2005 and the Ph.D. degree in image and signal processing from Telecom ParisTech, Paris, in 2010.

He is currently a Senior Researcher with the French National Institute for Geographic Information and Forestry, Paris. His research interests include point cloud and optical image processing, land-cover classification, and more generally multimodal remote sensing.

Dr. Mallet serves as the Chair of the ISPRS Working Group on Image Sequence Analysis for 2012–2016. He served as the Editor-in-Chief for the *French Journal of Photogrammetry and Remote Sensing* between 2011 and 2015.



**Stéphane Jacquemoud** was born in 1965. He received the Diploma in agricultural engineering from Montpellier SupAgro, Montpellier, France, in 1988 and the Postgraduate Diploma in remote sensing and the Ph.D. degree in physics, in the field of hyperspectral remote sensing applied to vegetation studies, from Paris Diderot University, Paris, in 1989 and 1992, respectively.

From 1989 to 1992, he worked with INRA Avignon, Avignon, France. Then, from 1992 to 1994, he worked as a Visiting Scientist with the European Commission Joint Research Center, Ispra, Italy. From 1994 to 1995, he was a Visiting Postdoctoral Fellow with the University of California, Davis, CA, USA. In 1995, he joined the Department of Physics, Paris Diderot University, as an Assistant Professor, and in 2003, he moved to the Department of Earth and Environmental and Planetary Sciences as a Full Professor. He is currently with Institut de Physique du Globe de Paris, Paris, where he studies natural surfaces using remote sensing. His main research interests include near-surface geophysics, environment, defense, precision agriculture, and exobiology.



**Gil Rito Gonçalves** received the B.S. degree in survey engineering and the M.S. degree in photogrammetry from the University of Coimbra, Coimbra, Portugal, in 1989 and in 1992, respectively, and the Ph.D. degree in geographic information science from the University of Marne-la-Vallée, Champs-sur-Marne, France.

Since 1989, he has been working with the Faculty of Science and Technology, University of Coimbra, where he is currently an Assistant Professor. His research interests include digital terrain modeling, terrestrial and airborne laser scanning systems, digital photogrammetry, and remote sensing image processing for coastal monitoring.



**Margarida Tomé** received the B.S. and Ph.D. degrees from the University of Lisbon, Lisbon, Portugal, in 1977 and 1989, respectively.

She is currently a Professor with the School of Agronomy, University of Lisbon, Lisbon, Portugal, where she teaches forest resources inventory, forest modeling, and forest management. She has supervised 12 Ph.D. and several M.Sc. theses. She coordinates the Forest Resources Inventory and Modeling Group (ForChange), one of the research groups of the Forest Research Centre that has the

aim to develop scientifically sound methods that can be used for Atlantic and Mediterranean forest ecosystem management by the public administration, nonindustrial forest owners, industry, nongovernmental organizations, and concerned citizens. The group has a strong focus on the transfer of knowledge to end-users by developing technological applications that might integrate and make useful the research results—Planted forests, with exotic or native tree species, have been an important research topic of the group. She is the author of more than 300 published papers, from which 23% have been referenced in ISI. She is a Coeditor of the Springer series *Managing Forest Ecosystems* with 33 books published since 2001. She is a coauthor of *Modelling Forest Trees and Stands* (Springer, 2012).

Ms. Tomé has been strongly involved with the European Forest Institute (EFI), namely, with EFIATLANTIC and EFIMED regional offices, and with the International Union of Forest Research Organizations, where she was Division IV coordinator for the last two terms.



**Paula Soares** received the M.Sc. degree in plant productivity and the Ph.D. degree in forestry from the University of Lisbon, Lisbon, Portugal, in 1995 and 2000, respectively.

Since 1996, she has been working with the University of Lisbon. She is teaching modules of silviculture and forest resources inventory and she has been involved in several European and National forest projects. She has a strong connection to the society, coordinating several studies related with evaluation and monitoring of carbon stocks in forest

areas and with the availability of maritime pine and eucalyptus wood in the future, which involves modeling tasks and adaptation and use of forest simulators. She has several publications referenced in ISI and supervised several M.Sc. theses. Her research interests include modeling of production forests and use of unmanned aerial vehicles in forest inventory.



**Luísa Gomes Pereira** received the B.S. degree in survey engineering from Porto University in Portugal, in 1983; the M.S. degree in digital photogrammetry from the International Institute for Aerospace Survey and Earth Sciences, Enschede, the Netherlands, in 1988; and the Ph.D. degree in digital photogrammetry from the Technical University of Delft, Delft, the Netherlands, in 1996.

In 1983, she joined the Department of Mathematics, University of Porto, Porto, Portugal, as an Assistant. In 1996, she became a Researcher with the

Ministry of Transport, Public Works and Water Management Rijkswaterstaat, the Netherlands. Since 1998, she has been with Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro, where he is currently an Associate Professor.



**Frédéric Bretar** received the M.S. degree in remote sensing from the Paris Diderot University, Paris, France, in 2000 and the Ph.D. degree in image processing from École des Ponts ParisTech, Télécom ParisTech, Paris, France, in 2006. He also obtained an accreditation to supervise research (HDR) from Paris Diderot University.

He was a Principal Investigator of the Lidar Remote Sensing Group, French National Geographic Institute, before heading the laboratory of Earth sciences in Rouen, where he developed photogrammetric methodologies for geophysics. Since 2012, he has been a Vice-Consul General with the Consulate General of France in Shanghai, Shanghai, China, where he is in charge of science and technology.