# COMPARING SMALL-FOOTPRINT LIDAR AND FOREST INVENTORY DATA FOR SINGLE STRATA BIOMASS ESTIMATION – A CASE STUDY OVER A MULTI-LAYERED MEDITERRANEAN FOREST

*António Ferraz[1,2,3], Gil Gonçalves[3,4], Paula Soares[5], Margarida Tomé[5], Clément Mallet[1], Stéphane Jacquemoud[2], Frédéric Bretar[6], Luisa Pereira[7]*

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

6- CETE Normandie Centre, Laboratoire des Ponts et Chaussées, Grand Quevilly, France
7- Universidade de Aveiro, Escola Superior de Tecnologia e Gestão de Águeda, Águeda, Portugal

## ABSTRACT

Current methods for accurately estimating vegetation biomass with remote sensing data require extensive, representative and time consuming field measurements to calibrate the sensor signal. In addition, such techniques focus on the topmost vegetation canopy and thus they are of little use over multi-layered forest ecosystems where the underneath strata hold considerable amounts of biomass. This work is the first attempt to estimate biomass by remote sensing without the need for massive *in situ* measurements. Indeed, we use small-footprint airborne laser scanning (ALS) data to derive key forest metrics, which are used in allometric equations that were originally established to assess biomass using field measurements. Field- and ALS-derived biomass estimates are compared over 40 plots of a multi-layered Mediterranean forest. Linear regression models explain up to 99% of the variability associated with surface vegetation, understory, and overstory biomass.

***Index Terms***— allometric equations, airborne laser scanning, forest vertical stratification, stratum biomass estimates.

## 1. INTRODUCTION

The scientific community reports large uncertainties about $CO_2$ sources, fluxes, and sinks, which are due to terrestrial ecosystems [1]. We don't know the exact amount of biomass stored in forests and how long it will be preserved. Measuring it implies cutting down all trees over limited areas of the earth in order to weigh each component: stems, bark, branches, and leaves. Obviously, this method is not relevant to the mapping of large areas. As a result, forest biomass is commonly determined using allometric equations that establish quantitative relationships between some key plant metrics (usually easy to measure) and other plant properties (usually difficult to measure) assuming that the relative growth rates of plant variables are proportional. Many allometric equations dedicated to biomass estimation have been published in the literature: they apply to a wide range of plant species across different forest biomes [2]. Tree height ($th$) and diameter at breast height ($dbh$) proved to be the best metrics to derive plant biomass.

Consistent and accurate measurements of such metrics are generally assessed through field forest inventories. However, as far as herbaceous and shrubby vegetation is concerned, it would be tedious to measure such variables for all plants. It results that their biomass is rather appraised at the forest plot-scale as a function of the mean height and the percent cover, two variables that are easier to measure [3].

There is a growing need for producing biomass maps at different time and spatial scales because they are crucial to understanding changes and trends in terrestrial ecosystems [4]. Many studies have investigated the ability of remote sensing to estimate forest aboveground biomass (AGB, Mg/ha). Small-footprint airborne laser scanning (ALS) is one of the best-suited techniques [4]. It surpasses optical and radar imagery over dense vegetation canopies because the lidar signal does not saturate. ALS is an active method that provides distance measurements between the platform and the surface illuminated by the laser beam. Such range measurements are georeferenced using a hybrid GPS/INS system. Depending on the nature of the target, a single pulse emission may result in one to several backscattered echoes. Thereby, ALS is able to penetrate beneath forest canopies down to the ground (Fig. 1a). Many authors showed the potential of multi-echo ALS data to compute digital terrain models (DTM) over vegetated areas as well as to estimate various forest variables [5].

There are two main approaches to estimate biomass using ALS data: distribution- and allometric-based methods. The first one is a two-stage statistical method. At the first stage, forest biomass is derived from a field inventory and then used to establish regression models with ALS point cloud metrics (height percentiles, maximum height, etc.). At the second stage, the calibrated models allow to predict the biomass over the forest area that has been surveyed by the ALS. This method is easy to implement and is compatible with present forest inventories due to common reference plots. However, it requires extensive, accurate, representative, and costly field data. Moreover, over heterogeneous forest ecosystems, such as Mediterranean forests, extrapolation of these regression models seems to be problematic since they are calibrated in relatively small areas that may not be representative of the forest variability [6]. Finally, in multilayered forests, it is

necessary to extract strata in order to calculate the ALS point cloud metrics inherent to each one: overstory, understory, and surface vegetation. So far, biomass estimates have been limited to the overstory layer and vertical stratification performed empirically.

In the second approach, forest biomass is assessed like in field-based methods, i.e., by means of allometric equations. This approach is based on individual crown-segmentation methods. The latter delineate individual crowns on the ALS point cloud allowing to derive key forest metrics (tree height, crown width, etc.) required by allometric equations [7]. With the individual crown approach, one expects to determine tree and bush metrics with a consistent bias or not, so that no project- or site- specific calibration is required. Thus, these methods significantly reduce the need for field measurements. Until recently, individual crown-segmentation methods were only able to delineate the topmost tree crowns and many ALS points corresponding to the surface and understory vegetation remained unassigned. Therefore they are of limited interest over multi-layered forests since a significant amount of biomass is stored in the ground strata.

We recently developed a segmentation method, called Adaptive Mean Shift (AMS 3D), which can delineate individual crowns independently of their position in the canopy [5]. This is the first individual crown-segmentation method adapted to segment multi-layered forests. As it works directly on the 3D point cloud, it applies to all the forest strata. Moreover, it addresses vertical stratification assigning each individual crown to a forest layer (Fig. 1). Consequently, it provides the necessary forest metrics to apply allometric-based methods to complex forest ecosystems.

The goal of this study is to assess the reliability of ALS data to replace *in situ* forest inventories for biomass estimates. The forest metrics required by the allometric equations are assessed from a field forest inventory and from the AMS 3D forest maps over 40 plots of a multi-layered forest. The results of both approaches are then compared.

## 2. METHOD

### 2.1. Field experiment

The study area is located near the city of Águeda in northwest Portugal. It covers 9 km$^2$ and its altitude varies from 70 m to 220 m with gentle to steep slopes. The landscape is predominantly composed of woodlands dominated by blue gum eucalyptus (*Eucalyptus globulus*). The forest stands can be strongly populated by understory and surface vegetation: juvenile trees (eucalyptus, pine, acacia, oak), gorse bush, heath, carquesia, sistus, blackberries, brooms, ferns and herbaceous plants. A forest inventory led to the selection of 40 plots covered by eucalyptus: 30 mature (>4 years old) and 10 juvenile (1-4 years old). Each plot consists of two concentric circles, an outer (400 m$^2$) and an inner (200 m$^2$). Field measurements were performed according to a protocol approved by the Portuguese National Forest Inventory. Individual trees within the inner circle were characterized by means of *dbh* (cm), *th* (m), and crown base height (*cbh*, m). Note that, only trees higher than 2 m with a *dbh* larger than 5 cm were considered. Furthermore, within the outer circle, surface vegetation and understory were characterized by the mean height (m), the percent cover (%) and dominance (%) which is the percent cover discriminated by species. For further details on the field experiment and the forest stand characterization, please refer to [5]. 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. The digitized waveforms were converted into a 3D point cloud. Each laser pulse gave rise to 1-5 ALS points. The average point density within each plot was 9.5 pt/m$^2$ (min = 4.7 pt/m$^2$, max = 15.5 pt/m$^2$, $\sigma$ = 1.9 pt/m$^2$). The ALS

point cloud (Fig. 1a) was decomposed into 3D segments corresponding to individual vegetation features, such as shrubs or tree crowns, by means of the AMS 3D method [5]. The segments were gradually assigned to a forest layer: surface vegetation, understory, and surface vegetation (Fig. 1b). However juvenile plots displayed only two strata (surface vegetation and overstory).

The AMS 3D forest maps were validated against some forest variables measured in the 40 plots [5]. 438 out of 649 trees (67.5%), hereafter called CI trees, were correctly identified by the method. The success rate varied with the tree crown social status: 98.6%, 85.2%, 61.4%, and 12.8% for dominant, codominant, dominated, and suppressed trees. As for the CI trees, the ALS-derived *th* related better to the field measurements ($R^2$ = 0.96, $RMSE$ = 0.86 m) than the ALS-derived *cbh* ($R^2$ = 0.69, $RMSE$ = 2.48 m). Nevertheless, 32.4% of the field-measured trees, hereafter called false negatives (FN), were not detected Conversely, 9.2% of the trees within the AMS 3D forest maps were false positives (FP), i.e. fictitious trees. The FP are mainly due to the over-segmentation of dominant trees: the AMS 3D identifies at least two crowns for a single tree. Linear regression models explain 70% and 68% of the variability associated with the height of surface vegetation and understory, respectively.
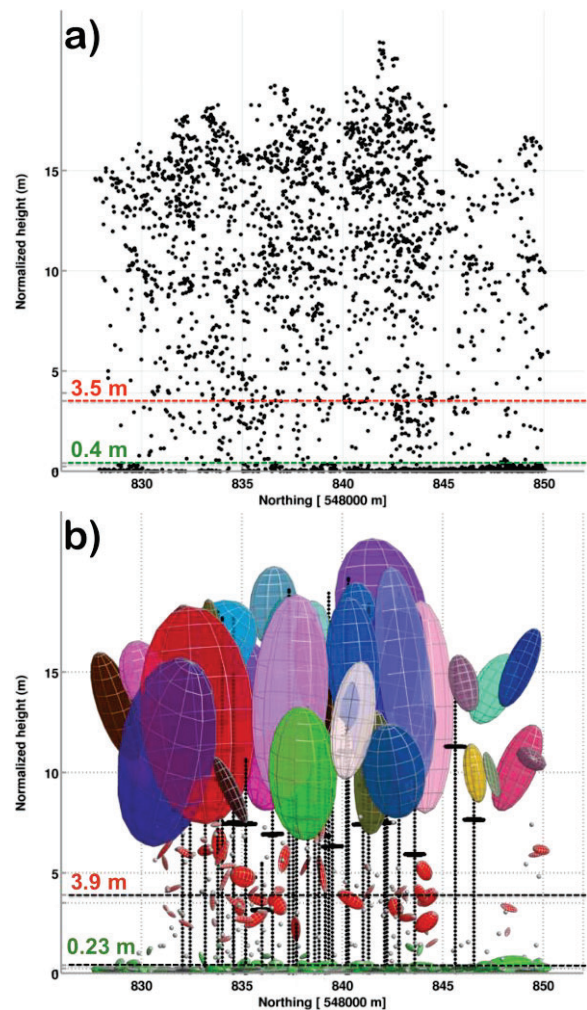


**Fig. 1. a)** ALS point cloud decomposed into **b)** 3D segments. The multicolored ellipsoids correspond to individual trees within the overstory, the red and green to surface vegetation and understory. The field-surveyed trees are marked in black. The values correspond to field-measured **(a)** and ALS-derived **(b)** mean height for understory and surface vegetation [5].

## 2.2. Field inventory derived biomass of each stratum

The biomass stored by each stratum is estimated using equations that change according to their nature or the available field measurements (Table 1). For mature plot overstory, the total biomass is the sum of single tree estimates (Equations 1-5), whereas for juvenile plot overstory, surface vegetation and understory, it is computed at the plot-scale (Equations 6). With respect to single trees, allometric equations were established for eucalyptus by using 441 specimens over 113 forest plots in Portugal [8]. Eucalyptus biomass is calculated as a function of individual tree metrics ($dbh$, $th$ and, $cbh$). In Table 1, $cl$ stands for the tree crown length (m), i.e., $cl = th - cbh$. Moreover, the allometric equations include plot specific coefficients ($aw$, $ab$, $al$, and $abr$) that depend on dominant height (mean height of the three tallest trees within a plot).

The biomass of each stratum is computed as a function of the mean height ($m$), percent cover (%), area ($m^2$), and mean bulk density ($kgm^{-3}$), namely $\bar{h}, pc, a$ and $\overline{bd}$ in Table 1. $\overline{bd}$ is a weighted mean of the bulk densities within each plot. The bulk density, defined as the biomass per unit volume, depends on the species. In [9] is given the reference bulk densities of the main plant species in Portugal. Finally, the weight corresponds to the species dominance (%) within each stratum.

| Biomass (kg) | | | |
|---|---|---|---|
| Trees | Stem | $w_w = 0.0104 \times dbh^{aw} \times th^{1.35942}$ | 1) |
| | Bark | $w_b = 0.00061 \times dbh^{ab} \times th^{1.08249}$ | 2) |
| | Leaves | $w_l = 0.04579 \times dbh^{al} \times cl^{0.82504}$ | 3) |
| | Branches | $w_{br} = 0.02478 \times dbh^{abr} \times cl^{0.02478}$ | 4) |
| | Total | $w_t = w_w + w_b + w_l + w_{br}$ | 5) |
| Strata | | $w_{stratum} = \bar{h} \times pc \times a \times \overline{bd}$ | 6) |
| DBH (cm) | | | |
| Trees | | $dbh = \dfrac{0.6073 \times th}{1 - 0.0116 \times th}$ | 7) |

Table 1. Equations for biomass and $dbh$ estimates.

## 2.3. ALS data derived biomass of each stratum

The biomass of each stratum is also assessed using allometric equations (Table 1). As for the individual trees, we take advantage of the fact that only eucalyptus live in the overstory. Otherwise, it would be necessary to discriminate the tree species in order to select the adequate allometric equation. The AMS 3D provides all the unknown variables of the allometric equations except the $dbh$, the size of which is incompatible with the actual ALS spatial resolution employed for forestry applications, i.e., from 4 pt/m$^2$ to 20 pt/m$^2$. As a consequence, the $dbh$ is generally derived from $th$ measurements via allometric equations [7]. For the eucalyptus in our study area we apply the equation established by [10] (Equation 7, Table 1).

In each forest stratum, the AMS 3D only provides $\bar{h}$ and $a$, i.e., neither $pc$ nor $\overline{bd}$ are assessed. In this work, $pc$ was set in each plot to values provided by the field inventory, whereas $\overline{bd}$ was assigned to an averaged value that applies to the whole study area: 2.46 $kgm^{-3}$ and 2.21 $kgm^{-3}$ for surface vegetation and understory, respectively. They correspond to the reference bulk densities average [9] of the plant species present in each stratum (Section 2.1). In fact, the extraction of both $pc$ and $\overline{bd}$ in the ground strata remains a scientific issue that would require further investigation and that is discussed in Section 3.

The average field-derived biomass in the 40 plots equals 7.74 Mg (surface vegetation), 7.41 Mg (understory), 6.18 Mg (juvenile plot overstory) and 34.97 Mg (mature plots overstory). The average ALS-derived biomass represents 93.3% (7.22 Mg), 104.8% (7.76 Mg), 99.2% (6.14 Mg), and 95.4% (33.35 Mg) of the field-derived biomass, respectively.

Linear regression was used to investigate the strength of the relationship between ALS- and field-derived biomass at the plot-scale (Table 2 and Fig. 2). The estimates for mature plots overstory, calculated as the sum of the individual tree biomass, was highly correlated ($R^2 = 0.99$) with the field-derived values. However, these results must be handled with caution. The $dbh$ of correctly identified trees (CI) was computed by means of Equation 7 as a function of the ALS-derived $th$ with $R^2 = 0.78$ and $RMSE = 2.1$ cm. Then, the biomass of such trees was calculated using Equations 1-5 with $R^2 = 0.81$ and $RMSE = 22.63$ kg. These results show that the field-derived biomass correlates better with the ALS-derived biomass at the plot-scale ($R^2 = 0.99$) than with the estimates at the individual tree-scale ($R^2 = 0.81$). Table 2 and Fig. 2 display the result at the plot-scale: here the biomass of the CI and FP trees was aggregated over the forest plot. It seems that the fictitious trees generated by the AMS 3D (FP) compensate the trees undetected by the ALS (FN). Nevertheless, it does not explain the strong correlation at the plot-scale since the percentage of FN (32.4%) is much higher than that of FP (9.2%). Consequently, on can also conclude that the biomass stored in the FN trees is not so important at the plot-scale. Indeed, most of the FN trees in our study area are suppressed trees that store little biomass [5].

The results obtained with surface vegetation, understory and juvenile overstory by means of Equation 6 are promising. We found strong correlations and satisfactory RMSE. The mean height of the ground forest strata provided by the AMS 3D is a suitable measurement for biomass estimates. In addition, the averaged value assigned to $\overline{bd}$ for each forest stratum is a reliable proxy for biomass estimates. However, these results were computed using the $pc$ values provided by the field inventory. In order to reduce the site-dependence, new methods to extract single strata $pc$ from ALS data must be investigated. As above-mentioned, the extraction of $pc$ and $\overline{bd}$ remains a scientific challenge. The difficulty and expectation regarding the extraction of these variables are also quite different.

| Forest stratum | n | $R^2$ | RMSE ($Mg/ha$) | Δ ($Mg/ha$) |
|---|---|---|---|---|
| Mature overstory | 30 | 0.99 | 18.02 | 5.84 |
| Juvenile overstory | 10 | 0.79 | 6.81 | -0.65 |
| Understory | 30 | 0.93 | 3.11 | -0.68 |
| Surface vegetation | 40 | 0.80 | 3.08 | 0.31 |

**Table 2.** Linear regression parameters for field- versus ALS-derived biomass by forest strata. n corresponds to the number of plots and positive Δ means a biomass under-estimation.

Many authors claim that the $pc$ - together with the canopy height - is the forest variable that can be the most easily assessed by ALS data. Roughly speaking, $pc$ is computed as the fraction of ALS vegetation hits over the total number of ALS hits [11]. It has been successfully calculated over single layered forests, but it fails in multi-layered forests. On the one hand, in complex forest structures, it is necessary to stratify the vegetation to determine the strata-specific statistics, i.e., the number of ALS echoes returned by the overstory, the understory, and surface vegetation [12]. On the other hand, the shading effect introduced by the overstory on

the number of ALS hits returned by surface vegetation and understory must be corrected. The denser the overstory the lower the probabilities of a laser beam reaching the underneath vegetation. This correction is a common practice in forest studies using large-footprint ALS. However, no solution has been yet proposed for small-footprint ALS point clouds [13, 14].

The computation of $\overline{bd}$ implies knowledge about the species present within each forest plot and their dominance (Section 2.2). Species identification from ALS data has been limited to the overstory layer. To our knowledge, the discrimination of the species that compose surface vegetation and understory has not been studied. Some constraints actually limit the development of such a technique. Compared to the overstory, shrubby and herbaceous formations are generally more diverse, which greatly increases the complexity of the approach. In addition, the geometric structure of underneath plants is frequently poorly represented in the point cloud due to shading effects. For shrubby and herbaceous formations, we believe that $\overline{bd}$ must be assessed from ancillary data. We suggest establishing averaged values well adapted to be applied to a specific forest area or eco-region. The plant species that are expected within an ecosystem can be assessed either by prior knowledge about the forest or derived from ecological studies such as biodiversity mapping or biogeographical distribution of plant species [15, 16].
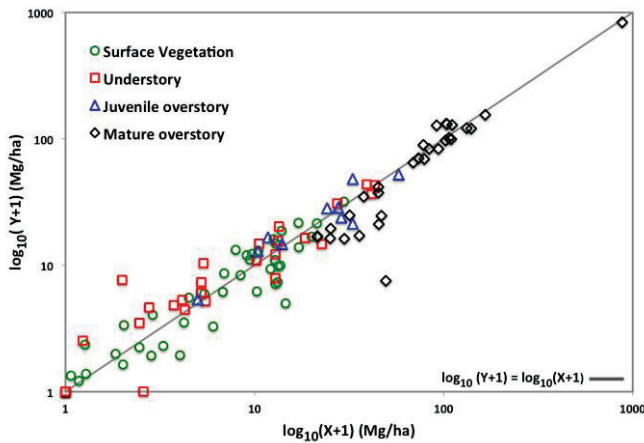


**Fig. 2.** ALS-derived biomass compared to field-derived biomass. X and Y correspond to the field- and ALS-derived biomass, respectivly.

### 4. CONCLUSION

Previous methods for biomass estimates over multi-layered forests only focus on the overstory layer. However, the ground strata can contain a considerable amount of biomass. For example in our study, about 27% of the aboveground biomass is stored either in surface vegetation (13.7%) or in understory (13.1%). That information is also crucial for other forest applications such as fuel mapping and animal or vegetation biodiversity assessment.

This study is the first attempt to estimate forest biomass using optical remote sensing without calibration. To reduce the dependence on *in situ* observations, further research is needed to establish a method able to assess the percent cover by forest strata. These preliminary results are very encouraging and we expect to develop such an approach in the near future.

Conversely, adult overstory biomass is assessed without the need for *in situ* measurements and the results are quite satisfactory. In fact, one expects that the overstory plots biomass be under-estimated since the ALS can detect only 67.5% of the total number of trees. However, results show that the undetected trees, which are mainly suppressed trees, are not so significant in terms of plot-scale biomass. Indeed, the fictitious trees generated by the AMS 3D (9.2%) improve the plot-scale forest biomass, i.e., they compensate for the biomass stored in the trees that the instrument was not able to detect (32.4%).

### 5. REFERENCES

[1] P. Friedlingstein, R. Houghton, G. Marland, J. Hackler, T. Boden, T. Conway, J. Canadell, M. Raupach, P. Ciais, and C. Le Quéré, "Update on $CO_2$emissions", *Nature Geoscience, 3*, pp. 811-812, 2010.

[2] J. Jenkins, D. Chojnacky, D. Heath, and A. Bridsey, "National scale biomass estimators for United States tree species", *Forest Science, 49*, pp. 12-35, 2003.

[3] P. Muukkonen, R. Mäkipää, R. Laiho, K. Minkkinen, H. Vasander, and L. Finér, "Relationship between biomass and percentage cover in understorey vegetation of boreal coniferous forests", *Silva Fennica, 40*, pp. 231-245, 2006.

[4] F.Hall, K. Bergen, J. Blair, R. Dubayah, R. Houghton, G. Hurtt, J. Kellndorfer, M. Lefsky, J. Ranson, S. Saatchi, H.Shugart, and D. Wickland, "Characterizing 3D vegetation structure from space: Mission requirements", *Remote Sensing of Environment, 115*, pp. 2753-2775, 2011.

[5] 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.

[6] M. García, D. Riaño, E. Chuvieco, and F. Danson, "Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data", *Remote Sensing of Environment, 114*, pp. 816-830, 2010.

[7] S.Popescu, "Estimating biomass of individual pine trees using airborne lidar", *Biomass and Bioenergy, 31*, pp. 646-655, 2007.

[8] N. António, M. Tomé, P. Soares, P. Tomé, and J. Fontes, "Effect of tree, stand, and site variables on the allometry of *Eucalyptus globulus* tree biomass", *Canadian Journal of Forest Research, 37*, pp. 895-902, 2007.

[9] S. Simões, "Expansão ao Alentejo e Algarve de uma curva de acumulação pós-fogo para a biomassa arbustiva", M.Sc. thesis, Universidade Técnica de Lisboa, Instituto Superior de Agronomia, Lisboa, Portugal, 2006 (in portuguese).

[10] P. Soares, and M. Tomé, "Airborne laser scanning technologies – need to estimate tree variables normally obtained in traditional forest inventory", IUFRO conference on mixed and pure forest in a changing world, Vila Real, Portugal, 6-8 October, 2010.

[11] 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.

[12] D. Jaskierniak, P. Lane, A. Robinson, and A. Lucieer, "Extracting LiDAR indices to characterize multilayered forest structure using mixture distribution functions", *Remote Sensing of Environment*, 115, pp. 573-585, 2011.

[13] M. Lefsky, D. Harding, W. Cohen, G. Parker, and H. Shugart, "Surface lidar remote sensing of basal area biomass in deciduous forests of eastern Maryland, USA", *Remote Sensing of Environment*, 67, pp. 83-98, 1999.

[14] J. Means, S. Acker, D. Harding, J. Blair, M. Lefsky, W. Cohen, M. Harmon, and W. McKee, "Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon", *Remote Sensing of Environment*, 67, pp. 298-308, 1999.

[15] S. Gachet, E. Vela, T. Tatoni, "BASECO: a floristic and ecological database of Mediterranean French flora", *Biodiversity and Conservation,* 14, pp. 1023-1034, 2005.

[16] H. Hazen, "Biodiversity Mapping", In *International Encyclopedia of Human Geography*, Elsevier, Oxford, pp. 314-319, 2009.