

# Estimating tree biomass in a plant barrier from an HD LiDAR point cloud

Balthazar Callen, Emiri Misselis and Nina Moser

Toulouse National Institute of Applied Science - Department of Applied Mathematics

May 21st, 2024

---

## Abstract

Human activities increase the presence of metals and metalloids (MM) in ecosystems. MM are a significant danger for the environment and human health. Research has shown that the biomass of vegetative barriers play a major role in the absorption of MM. Therefore, calculating the Above Ground Biomass (AGB) is crucial to estimate the potential storage of MM in a vegetated area. Studies show how to compute the AGB of a vegetated area by applying segmentation and allometric equations. However, these methods cannot precisely compute the biomass of individual trees of different species and specific shapes. This study aims to accurately estimate the biomass of trees bordering an old factory in Marseille using LiDAR data point cloud. We compute the Convex Hull (CH) of each tree and compute its volume. We generate clusters of points and use voxel methods to increase the accuracy of the volume. We evaluate the AGB from the obtained volume. We compare the above estimations with the AGB computed from the initial CH. The results indicate that the accuracy of the methods is contingent upon the number of points within the point cloud of trees. The initial CH cannot be enhanced for datasets with fewer than 100 points. However, when trees contain more than 100 points, the accuracy of the volume can be improved with clustering methods, slice and adaptative algorithms. The volume has a collinear relationship with the AGB. Thus, given the linear regression model between AGB and volume, it is feasible to predict AGB based on calculated volume.

**Key words:** LiDAR, Above Ground Biomass, Volume, Convex Hull, Clustering, Voxel, Allometric equations.

# 1 Introduction

Vegetated areas bordering factories and industrial zones can be contaminated with metals and metalloids (MM). A high concentration of MM is toxic and harmful for the environment and the biodiversity [Calmon, 2022]. Thus, evaluating the quantity of MM stored by deposit by the vegetation is crucial to know the contamination level of the area. The storage capacity of the area depends on the vegetation species present and is proportional to the biomass of the zone. Each tree species stores a certain amount of MM per tonne of biomass. Therefore, computing the Above Ground Biomass (AGB) of a vegetated area is essential to evaluate the quantity of MM stored. Plant biomass refers to all the organic material produced by plants through photosynthesis, and includes woody biomass (trees, shrubs and bushes) and herbaceous biomass (grasses, herbs and vegetables). AGB is defined as the organic matter in trees expressed as oven-dry tons per unit [Torre, 2018].

The commonly used method to compute AGB is the calculation of allometric equations according to tree species. These equations give the AGB from tree metrics such as the Tree Height (TH), the Diameter at Breast Height (DBH), the Crown Projected Area (CPA) and the Crown Volume (CV) [Qiao et al., 2023]. However, field studies to collect these features are expensive, time consuming and nearly impossible to carry out in the case of large forests [Tusa et al., 2020], [Qiao et al., 2023]. Thus, LiDAR can be used to assess tree metrics from a 3D point cloud of vegetated areas.

Remote sensing LiDAR is a device that efficiently characterises forest canopy and obtains tree features such as the TH, DBH, CPA and CV needed for the allometric equations. The method is based on the obtained response time for the light to return from its reflection. A laser impulsion is emmited from the LiDAR. This emitted pulse produces a returned waveform each time it is intercepted by an obstacle. The returned waveform is recorded by the sensor. Thus, knowing the light speed and the response time, LiDAR deduces the distance from the reflected point. Therefore, the three-dimensional coordinates of every object is obtained. The

accuracy of the features extracted from the data partly depends on the number of points generated per unit of surface. In Airborne LiDAR System (ALS) the laser is mounted in an aircraft or a helicopter. This technology is sensitive to the positioning and orientation of the device but gathers a high quantity of accurate measurements of the vertical structure and optical properties.

Hence, LiDAR is useful to obtain trees metrics at a large scale. However, the allometric equations are not appropriate for trees with a particular shape (the case for bushes or trees lying in the ground for instance). Moreover, the allometric equations have not been established for all tree species yet. Therefore, LiDAR point clouds can be useful to estimate the volume of trees from their cloud point. [Zhou et al., 2022] proposes to calculate AGB from the three-dimensional green volume using the Convex Hull (CH) and voxel algorithm. Using the volume increases the accuracy of the AGB obtained. The authors suggest to take into account different tree heights and the volume in the AGB equations. When using the Convex Hull Volume (CHV),  $RSME = 45.31$ . The AGB estimation is highly improved with voxel and slice algorithms:  $RSME_{voxel} = 12.03$  and  $RSME_{slice} = 13.01$ . However the equations are specific to each species. To increase accuracy of the three-dimensional volume, [Ferraz et al., 2016] developed the AMS3D algorithm to segment points and stratify the vegetation into single layers. The AMS3D uses the mean shift vector principle to extract modes and generate clusters. However, the precision of the three-dimensional volume is subject to the density of the point cloud obtained by LiDAR and the conversion to the biomass remain specific to a tree specie.

In the case of vegetated areas with different species of trees, or data cloud with a low number of points, the allometrics equations and the 3D green volume methods remain inaccurate and difficult to adjust. Previous research mainly focused on forest, and especially single species forests. Vegetative barriers were not studied because of the disparity of types (tree, shrub, bushes) and species. These research were also conducted on larger areas. The reduced size of a vegetative barrier does not allow statistical anal-

ysis. Moreover, allometrics equations remains unavailable for the majority of species. Furthermore, previous research was not conducted in a contamination context which involves new challenges such as limited development of the species of the area, individual adaptation to the environment and individual tree structure.

In this study, our objective is to estimate the biomass in a context of contamination of a barrier of 17 trees with different shapes of 5 species : *Pistacia lentiscus*, *Pinus halepensis*, *Pittosporum tobira*, *Juniperus phoenicea* and *Quercus ilex*. We estimate this biomass with the tree volume obtained from a LiDAR point cloud. We compute the previously developed CH, voxel and slices algorithms. We use an AMS3D clustering methods that we parameter to fit the different species and data clouds. Additionally, to facilitate method comparison, we implement the DBSCAN and Kmeans clustering algorithms. We calculate the volume of the CH applied on each cluster.

The article is structured as follows. The data is described in Section 2. Section 3 outlines all the methods we establish and Section 4 presents our results. We discuss them in Section 5 and finally, Section 6 presents the conclusions which can be drawn from our results.

## 2 Material

### 2.1 Study area

The study area is situated in the southwest of Marseille (Bouches-du-Rhône, Provence-Alpes-Côte d’Azur, France) near the National Parc of Calanques (Figure 1). Industrial activity arrived in the region in the 19<sup>th</sup> century, and continued throughout the century at the Sormiou, Saména, Goudes and Callelongue industrial sites, and even into the 20<sup>th</sup> and 21<sup>st</sup> centuries at the Montredon and Escalette. Soda, sulphur, sulphuric acid and lead processing plants were among the various industrial activities of the region [Barthelemy and Hérat, 2016]. In some areas, abnormal levels of pollutants were found in soils and plants, including lead, arsenic, copper and zinc [Testiat et al., 2013].

We studied a vegetative barrier in Montredon (Figure 2). The site of Montredon is divided in two parts: an industrial part with old factories and a natural part with wooded areas where the vegetative barrier is situated. The individuals of this barrier are relevant in this study because they are at the border between the industrial and the wooded area. Hence, it is the better location to consider the absorption effect of the barrier. This barrier measures 67 meters long and is composed of 17 trees:

- 1 *Juniperus phoenicea* L. (JUPH)
- 1 *Quercus ilex* L. (QUIL)
- 2 *Pittosporum tobira* W.T. Aiton (PITO)
- 6 *Pinus halepensis* (PIHA)
- 7 *Pistacia lentiscus* L. (PILE)

We also studied the site of Roy d’Espagne as a reference area. This site is a priori non contaminated. We studied 13 trees in this site (PIHA and PILE).



Figure 1: Location of Montredon (orange) and Roy d’Espagne (yellow) in Marseille, France [Google, 2024]

### 2.2 Data

#### 2.2.1 LiDAR point cloud

Our study is based on a 3D HD point cloud of the studied area acquired by the IGN (National



Figure 2: IGN's map of Montredon - ©BD ORTHO 5cm

Institute of Geographic and Forestry Information). This point cloud is composed of 22,203,485 points. It was cleaned by L. Calmon to focus only on the vegetative barrier we studied. Hence, the entire points cloud, the point cloud of the barrier (Figure 3) and one point cloud by three (Figure 4) were available for our study.

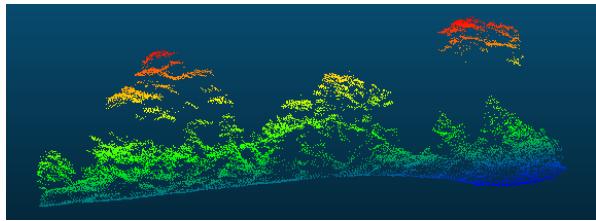


Figure 3: LiDAR HD Point Cloud of Montredon's barrier - ©LiDAR HD IGN

### 2.2.2 In-situ measurements

We also use different in-situ collected data for each individual such as Tree Height (TH) (in m), width (in m), depth (in m) and Diameter at Breast Height (DBH) (in m), which refers to the diameter of the tree trunk measured at 1.37

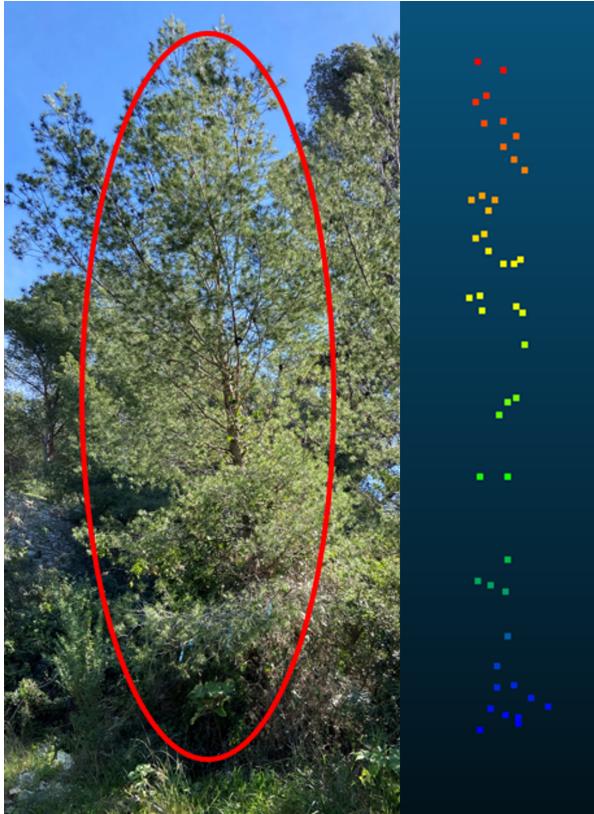


Figure 4: Picture of PIHA1 and corresponding LiDAR HD point cloud - ©LiDAR HD IGN

meters above the ground. These measurements were collected on March, 19<sup>th</sup> of 2024, using a measuring tape and a dendrometer for the height [Calmon, 2024]. These measures are summarized in Table 1 and Table 2. The barrier is composed of more individuals and species than the reference site. Due to the shape of the trees, we only have the DBH for one species (PIHA). The PIHA trees of Montredon tend to be smaller (considering the height) than the Roy d'Espagne ones (mean of 6.55 for Montredon and 9.51 for Roy d'Espagne). For the PILE individuals, we do not observe such difference (2.42 for Montredon and 2.39 for Roy d'Espagne).

We also have pictures of each tree which is useful to visualize them, particularly when they have a complex shape (Figure 4) [Calmon, 2024].

Table 1: Field measures of each individual tree at the Montredon site (in m) [Calmon, 2024]

	Height	Width	Depth	DBH
JUPH1	5.9	6.3	8.2	-
PIHA1	9.1	3.9	4.3	0.1750
PIHA2	5.3	3.2	7.6	0.1448
PIHA3	6.4	3.1	4	0.1273
PIHA4	5.4	5.2	5.9	0.2864
PIHA5	6.6	16	8.6	0.4456
PIHA6	6.5	4	5.2	0.2546
PILE1	1.45	3.8	1.3	-
PILE2	4	5.5	2.2	-
PILE3	1.5	1.6	2.2	-
PILE4	2.9	5.3	3.6	-
PILE5	2.5	5.4	3.9	-
PILE6	2	2.5	2	-
PILE7	2.6	8.4	4.5	-
PITO1	2.4	3.5	3.6	-
PITO2	2.4	5.1	1.5	-
QUIL1	3.7	2.3	3.3	0.1050

Table 2: Field measures of each individual tree at the Roy d’Espagne site (in m) [Calmon, 2024]

	Height	Width	Depth	DBH
PIHA1	9.9	7	6.8	0.5188
PIHA2	6.1	3.1	2.4	0.1368
PIHA3	7.5	3.3	3.9	0.1273
PIHA4	11.9	8.6	6.6	0.4138
PIHA5	8.1	7.4	7.9	0.2546
PIHA6	13.6	6.5	5.1	0.3660
PILE1	3.3	3	2.5	-
PILE2	2	1.9	1.6	-
PILE3	2.2	4.7	1.9	-
PILE4	3.7	5.4	4.6	-
PILE5	2	3	2	-
PILE6	1.35	1.4	1.8	-
PILE7	2.2	3.9	5.7	-

### 3 Methods

We first extract the point cloud of each tree. We clean the data set with CloudCompare comparing the point clouds and the pictures of the trees taken in the field. We generate

the CH of each cloud and compute its Volume (CHV). This volume, studied before by [Zhou et al., 2022] is a first estimation and is our reference to evaluate the accuracy of the other methods. To increase the accuracy of the volume, we carry out a cluster analysis with AMS3D, DBSCAN and Kmeans methods to separate the point cloud into three-dimensional fitted clusters. The AMS3D method based on density was carried out by [Ferraz et al., 2016] and was efficient for maritime pine and eucalyptus trees. We implement it on the species we have in the barrier and evaluate the accuracy of this algorithm. The DBSCAN is a simple algorithm also based on density that is known for statistical analysis. Thus, we compare its performances with AMS3D. Finally, Kmeans uses another method based on centroid. We choose to implement it to provide an algorithm not based on density. By using clustering and decomposing the specific shape of each single tree in these three-dimensional shapes, we aim to eliminate empty spaces in the CH and reduce the error on the fitting of the CH on the tree. We further compute the CH according to the clustered point clouds of each tree and finally calculate its volume (Figure 6). We compute the volume from the voxel methods mentionned by [Ferraz et al., 2016].

We illustrate our method with the PIHA5 tree in the data set. This tree has 2628 points in its point cloud (Figure 5). It is the individual with the highest number of points in the barrier. The points are evenly distributed throughout the data cloud. In our results, we observe that the methods do not perform as effectively with point clouds that have fewer points or uneven dispersion. We illustrate the methodology with Figure 6.

#### 3.1 Individual trees cloud extraction

We have to extract each individual cloud point from the cloud point of the entire barrier. Some algorithms exist to do so ([Straker et al., 2023]) but the complex and irregular shape of the barrier does not allow these algorithms to achieve good results. Hence, we have to extract them manually using the software CloudCompare and Qgis.

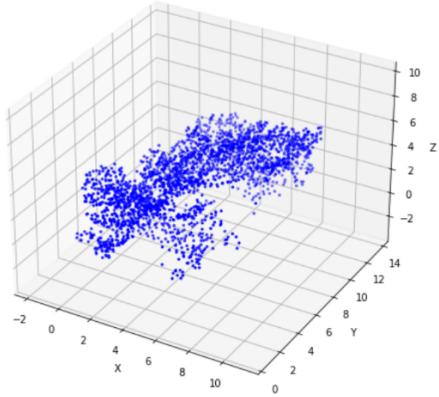


Figure 5: PIHA5 point cloud

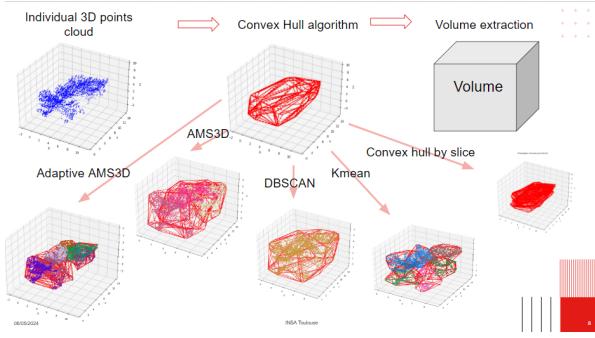


Figure 6: Methodology using CHV as a reference

First, we use Qgis to visualize the cloud point over the map of the area, allowing us to delineate each tree according to its crown. From Qgis shape, we extract each individual point cloud. Then, on CloudCompare, we clean each individual cloud based on its corresponding tree reference image.

The results of this procedure is shown in Figure 7 where each tree is colored over the barrier.

### 3.2 Convex hull algorithm

A CH is the smallest set of points that forms a polygon containing all the points from the data set. In three-dimentional space, the CH is a convex polyhedron that encloses every point of the data point cloud in its volume (Figure 8). We compute the volume of the Polyhedron to generate a first approximation of the tree volume.

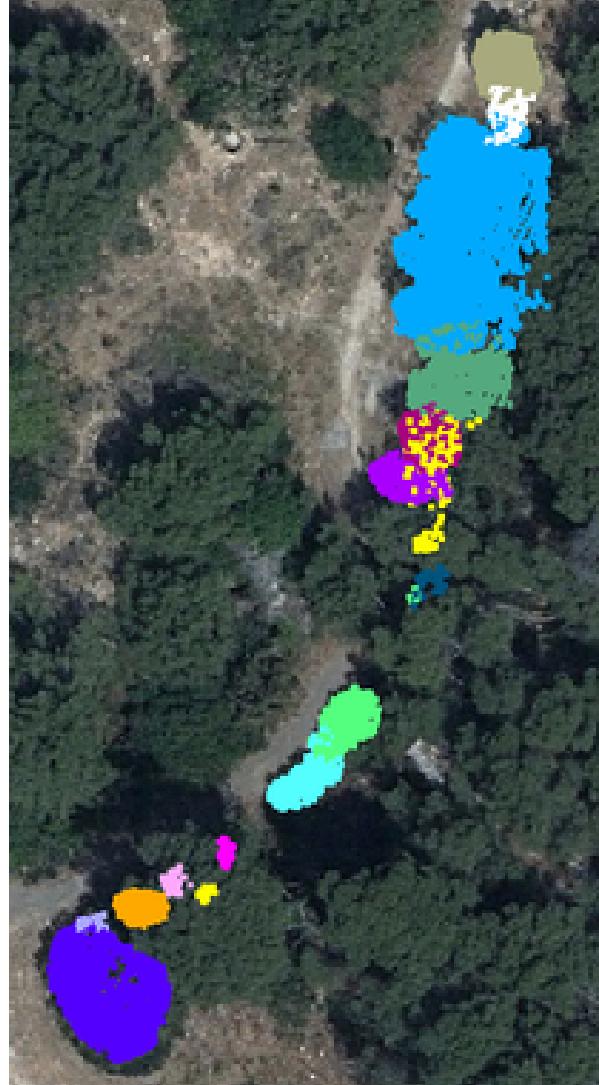


Figure 7: Visualization of the individual point cloud over the barrier in Qgis

The resulting shape is convex. However, when dealing with irregularities such as empty spaces or holes within the shape, the CH overestimates the volume. The algorithm achieves convexity by connecting two points with an edge, but this approach fails to accurately represent the empty space or narrower areas. To enhance the method's accuracy, we apply the CH to trees slices and to clusters of points.

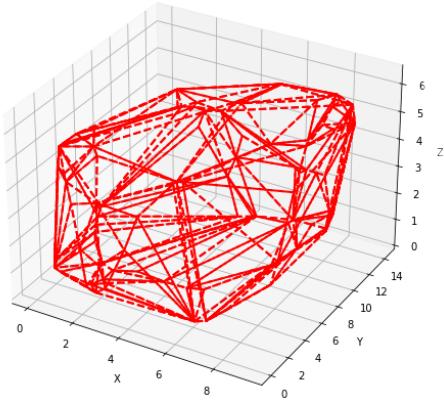


Figure 8: PIHA5 Convex hull

### 3.3 Convex hull by slice algorithm

Recognizing irregularities in our data point cloud, we opt to segment the points. This approach involves dividing the point cloud into layers (Figure 9). Hence, each layer is more homogenous in term of point density as trees density is not constant over it height. We set each slice to 10% of the tree's height. For each of these layers, we compute the CH and extract its volume. We then sum up the volume of each layer to obtain the final volume of the tree.

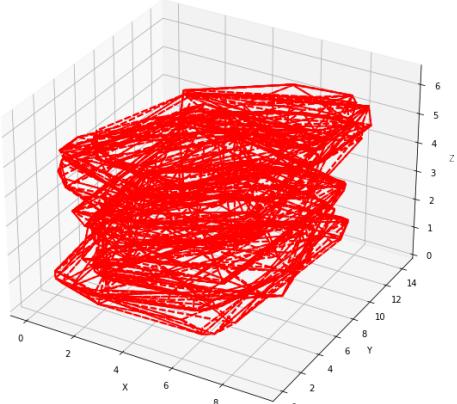


Figure 9: PIHA5 Convex hull by slice

## 3.4 Clustering with AMS3D

### 3.4.1 Background

In this section we refer to [Ferraz et al., 2012]. Adpatative Mean Shift in three-dimensional space (AMS3D) is a mean shift algorithm based on density estimation. According to the algorithm meaning, AMS3D finds the modes of the Probability Density Function (PDF) estimation of a random variable  $X$ . Mean shift vector determines how each data point should be moved to converge towards a mode of the density function. The vectors are constructed from a chosen kernel (usually the Gaussian kernel) and a defined neighborhood of a point  $x_i$ . The algorithm iteratively updates the position of the points toward the direction that increases density until convergence. In the context of clustering, the converged points are used as a cluster center and the other points are assigned to the nearest cluster center after convergence. The result of this algorithm applied on PIHA5 is shown in Figure 10.

### 3.4.2 Bandwidth

The bandwidth must be estimated in order to optimize the algorithm. The bandwith is the width of the kernel used to estimate the density of the data points. A larger bandwidth allows distant points to be part of the same cluster.

We use the function *estimate\_bandwidth* from the scikit-learn library for this purpose. This function takes the point cloud as input along with a quantile parameter. The quantile parameter controls the fraction of the smallest distances used to compute the bandwidth.

## 3.5 Adaptative AMS3D clustering

We know that our data point clouds are not uniformly structured. Hence, we use an adaptive algorithm method. Adaptive clustering refers to clustering methods that dynamically adjust their parameters based on the characteristics of the data that is used in statistical studies. [Chinrungrueng and Sequin, 1995] has developped this method for the Kmeans cluseting algorithm, we adapte this method for the AMS3D.

In the context of an adaptative AMS3D clus-

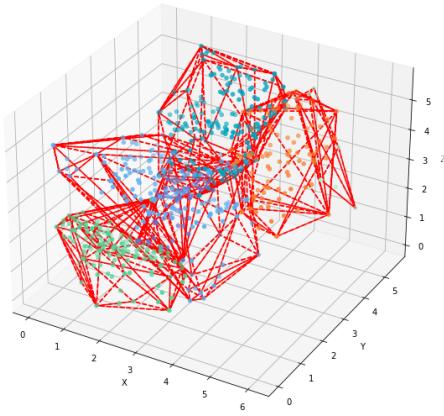


Figure 10: PIHA5 Convex hull on AMS3D clusters

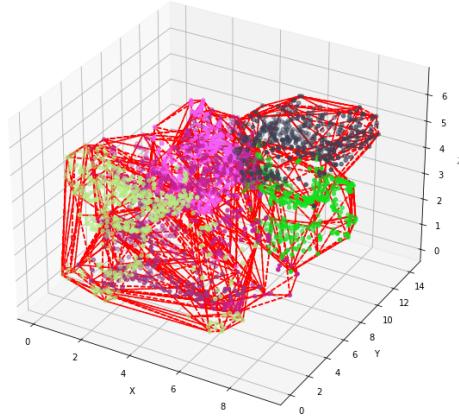


Figure 11: PIHA5 Convex hull on adaptative AMS3D clusters

tering algorithm, we develop an algorithm that adapts the bandwidth parameter to accommodate the local density of data points. This adjustment allows us to generate clusters with varying sizes and densities according to the local data characteristics.

To determine the appropriate bandwidth, we consider  $k$  the number of closest neighbors to take into consideration in the estimation of the local density. We employ the *NearestNeighbors* function to compute the  $k$  closest neighbors for the dataset and then calculate the mean distance between these neighbors. The adaptive bandwidth is subsequently determined as the bandwidth value that covers 90% of this mean distance, adjusting the width based on local density characteristics. The methods generates different CH as shown in Figure 11.

### 3.5.1 Neighbors for adaptative AMS3D

Specifically, if the total number of points exceeds 1000, we set  $k$  to 10% of the total number of points. Otherwise,  $k$  is set to 20%. We have to impose this condition because if the number of points is small, 10% is not sufficient to obtain an accurate model. According to observations on the data set, we imposed the limit as 1000 points.

## 3.6 Clustering with DBSCAN

The Density Based Spacial Clustering on Application and Noise (DBSCAN) algorithm is a clus-

tering method based on the estimation of the local density. The algorithm relies on two parameters. According to *sklearn.cluster.DBSCAN*, epsilon is the largest radius to form a cluster, it defines the neighborhood of each point. The minimum number of points in a cluster is minPts. At each iteration, DBSCAN selects an unvisited point  $x_i$ . If  $x_i$  is a core point, a new cluster  $C_i$  is created. Next, the algorithm finds all points in the epsilon space, these are added to the cluster. For each added point  $x_j$ , it checks if it finds a core point. In this case all  $x_j$  neighborhood are added to the cluster. The process continues until no point can be added to a cluster. The points density of PIHA5 is regular and does not allow the algorithm to separate the points into clusters (Figure 12).

## 3.7 Clustering with Kmeans

Kmeans is a simple clustering algorithm that generate  $K$  groups. The methods is known and developed in mathematical studies [Chinrungrueng and Sequin, 1995]. The algorithm initializes  $K$  centroids and computes the Euclidean distance to each cluster centroid for each data point. Kmeans assigns the data point to the closest cluster and updates the cluster centroid as the mean of the data allocated in the cluster. It iterates until convergence. The Kmeans algorithm converges but not necessarily to the global minimum. Hence, the number  $K$  has to be well chosen to guaranty convergence toward the best clusters (Figure 13).

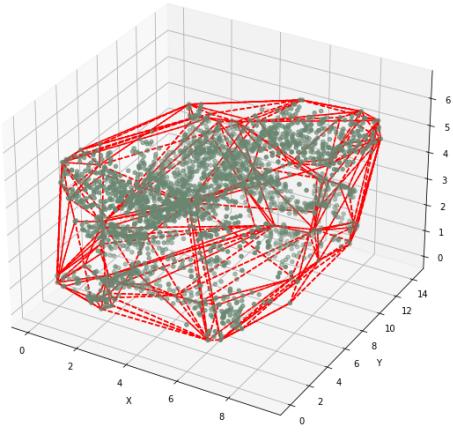


Figure 12: PIHA5 Convex hull on DBSCAN clusters

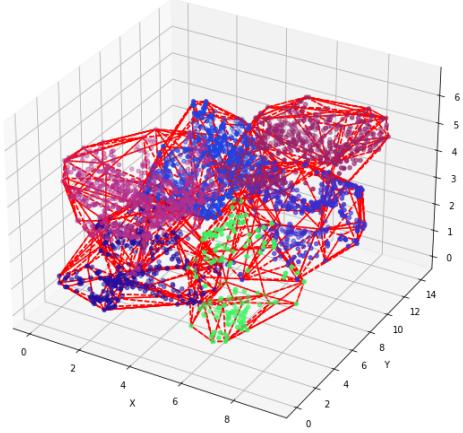


Figure 13: PIHA5 Convex hull on Kmeans clusters

### 3.8 Voxel Algorithm

A voxel defines a volumetric pixel. It represents a discrete element of volume in a three-dimensional space. The Voxel algorithm is a computational method that uses a three-dimensional grid to partition the data set allocating each point of the point cloud to a voxel [Zhou et al., 2022]. We compute the size of each voxel dividing each dimension length by the voxel resolution. The voxel resolution is the dimension of the cube. We split the initial space into voxels and allocate each point to a voxel. The final grid is the selection of voxels with more than one point. We compute the volume occupied by the voxels of the final

grid to calculate the tree volume. The final grid of PIHA5 is shown in Figure 14

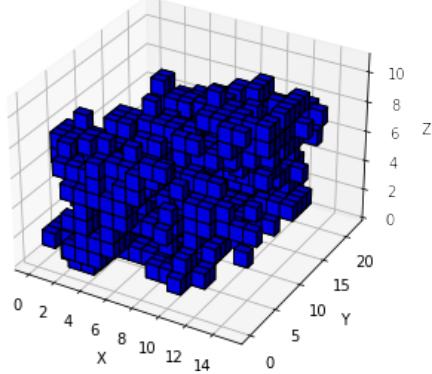


Figure 14: PIHA5 Voxel method

### 3.9 Adaptative Voxel Algorithm with AMS3D

As previously mentioned, we employ an adaptive method to address the irregular distribution of points within the dataset. The adaptive voxel algorithm is designed to compute voxel resolution dynamically based on clusters generated by the AMS3D algorithm and their associated densities. We compute the density of each cluster. This density represents the concentration of points within the cluster. Each cluster's density is normalized by dividing it by the maximum density among all clusters. This normalization process scales the densities to a common range for comparison. For each cluster we create a grid of voxels. The resolution of this grid is set equal to the normalized density of the cluster. Higher densities result in finer grids (smaller voxels), while lower densities correspond to coarser grids (larger voxels). We allocate each point of the cluster to the corresponding voxel within the grid based on its spatial location. We compute the volume of each voxel based on its size and dimensions within the grid. We sum the volumes of all occupied voxels across all clusters, the algorithm establishes the total volume representing the tree's volume.

## 4 Results

### 4.1 Accuracy of trees extraction

We proceed to validation of the segmentation, as the trees have singular shape and are surrounded by other individuals. We extract from each point cloud its height (with CloudCompare), width and depth (Qgis). We compare them to the fields measures and we calculate the relative error. The lower is the error, the better is the segmentation and the individual point clouds represent well the trees. Results are summarize in Table 3. We see that the segmentation over the height of the individuals is rather good, with a mean of 5%. However, the precision over the width and depth is poor, with a mean around 26%. These results can be explained because of the characteristics of the individuals. They form a vegetative barrier so each individual is surrounded by many others, leading to a segmentation hard to perform.

Table 3: Errors on the segmentation (%)

	Height	Width	Depth
JUPH1	1	4	3
PIHA1	15	66	58
PIHA2	3	9	34
PIHA3	6	42	5
PIHA4	5	8	5
PIHA5	1	6	1
PIHA6	9	15	11
PILE1	7	55	31
PILE2	3	42	14
PILE3	5	6	23
PILE4	10	9	19
PILE5	4	7	18
PILE6	9	60	60
PILE7	4	3	13
PITO1	3	28	44
PITO2	0	31	46
QUIL1	2	47	70
Mean	5	25.7	26.7

In the next part, we compare the volume obtained by each method against the CHV. Our goal is to define the method's accuracy by reducing its value, while ensuring that the reduction

does not exceed 50%. We established that less than 50% of the CHV is an underestimation. This decision is based on observations from tree plots.

### 4.2 Choice of the method parameters

To assess the accuracy of the methods based on the chosen parameters, we generate three categories. Each one presents different parameters for each method. We apply the algorithms set with the different parameters on our data and establish the ones that generate the best results. Table 4 presents the parameters of each methods in the different categories.

Table 4: Parameters for each categories

	1 <sup>st</sup> cat	2 <sup>nd</sup> cat	3 <sup>rd</sup> cat
Quantile (AMS3D)	0.5	0.3	0.1
Epsilon (DBSCAN)	1	0.5	0.2
k (Kmeans)	3	4	5
Resolution (Voxel)	0.2	0.1	0.3

We choose these parameters arbitrarily based on their definitions. Our objective is to organize trees into distinct group and identify the most effective method for each group. To achieve this, we calculate the CHV for each tree to delineate these groups. We observe that trees with a higher number of points present a larger volume. Specifically, we categorize the trees into three groups: low (less than 100 points), med (between 100 and 300 points), and high (more than 300 points).

Table 5: Trees with higher CHV (m<sup>3</sup>)

Tree	Points	Height	Width	Depth	CHV
JUPH1	1145	5.95	7.16	8.55	173.94
PIHA2	342	5.47	4.73	3.15	42.50
PIHA3	457	6.02	3.56	3.99	49.83
PIHA4	703	5.65	6.10	5.62	123.01
PIHA5	2628	6.54	9.30	14.29	537.53
PIHA6	505	5.90	3.94	4.44	52.63
PILE7	303	2.71	4.44	8.21	46.58

Table 5, Table 6 and Table 7 present the trees of each group and their metrics.

Table 6: Trees with medium CHV ( $\text{m}^3$ )

Tree	Points	Heighth	Width	Depth	CHV
PILE2	134	4.10	3.13	2.08	13.56
PILE4	293	3.19	4.38	5.61	19.48
PILE5	251	2.60	4.07	4.63	15.84
PITO2	222	2.41	2.84	3.45	11.19

Table 7: Trees with smaller CHV ( $\text{m}^3$ )

Tree	Points	Heighth	Width	Depth	CHV
PIHA1	46	7.73	0.80	1.65	4.96
PILE1	28	1.55	2.27	1.25	1.32
PILE3	27	1.57	1.64	1.65	1.19
PILE6	8	2.18	0.43	1.04	0.12
PITO1	29	2.46	1.98	2.37	3.54
QUIL1	16	3.62	0.96	0.92	1.07

For each method, we plot the obtained volume (y-axis) against the CHV (x-axis). We draw a red line following the equation  $x = y$  (CHV equal volume of the methods) and a blue line following the equation  $y = 0.5x$  (volume of the method equal 50% of the CHV). The goal is to ensure that our points fall within these lines for each method.

#### 4.2.1 Bandwidth for AMS3D

We choose to use quantiles equal to 0.1, 0.3 and 0.5. In fact, we try to use 10%, 30% and 50% of our smallest distance to estimate the bandwidth.

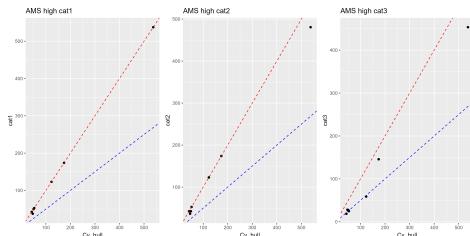


Figure 15: AMS3D parameter test on 'high' trees

Figure 15 shows that the third method,

$quantile = 0.1$ , allows a reduction of the volume for trees in the 'high' group. This parameters satisfy our conditions.

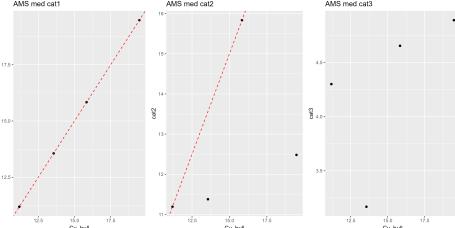


Figure 16: AMS3D parameter test on 'med' trees

According to Figure 16, the third category underestimates the volume. The first one does not improve the CHV. The second category and  $quantile = 0.3$  fits the best the points clouds.

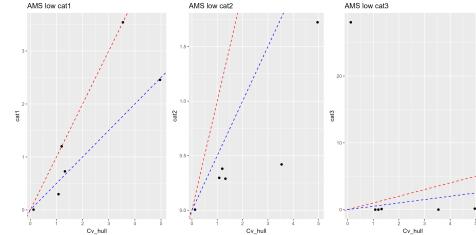


Figure 17: AMS3D parameter test on 'low' trees

Figure 17 shows that the volume accuracy hasn't been improved for the 'low' group trees with all parameters.

#### 4.2.2 Parameters for DBSCAN

The selection of parameters is crucial. We decide to compare the model with  $\epsilon$  set at 1, 0.5, 0.2. We impose 3 as the minimum number of points in a cluster to form a CH.

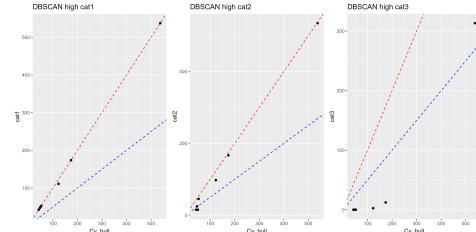


Figure 18: DBSCAN parameter test on 'high' trees

According to Figure 18, the second methods,  $\epsilon = 0.2$ , reduces the volume.

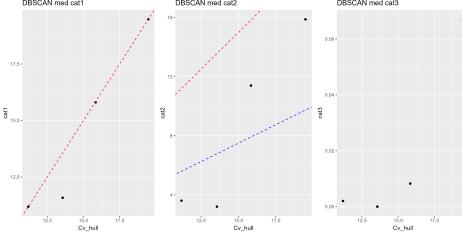


Figure 19: DBSCAN parameter test on 'med' trees

Figure 19 shows that only the first category allow the model to follow the constraints. The improvement is not conceiving, only one volume is different than the CHV.

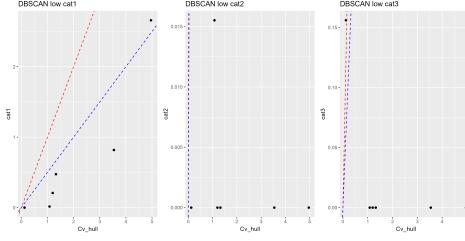


Figure 20: DBSCAN parameter test on 'low' trees

According to Figure 20, DBSCAN method is not efficient for trees of the category 'low'.

#### 4.2.3 Choice of the parameters K for Kmeans

We arbitrarily choose  $K = 3$ ,  $K = 4$ ,  $K = 5$  for our clustering approach. These choices aim to strike a balance and generate enough clusters to approximate our dataset accurately, while avoiding an excessive number of clusters that could lead to underestimation.

According to Figure 21, the second method with  $k = 4$  generates volumes that align well with our criteria. The third method also shows effectiveness, but we prefer the second method due to its more consistent and regular results.

Figure 22 show that the method is not efficient with all parameters for 'medium' volume trees.

According to Figure 23, the method is not accurate for 'low' volume trees

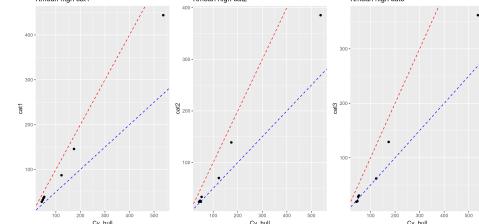


Figure 21: Kmeans parameter test on 'high' trees

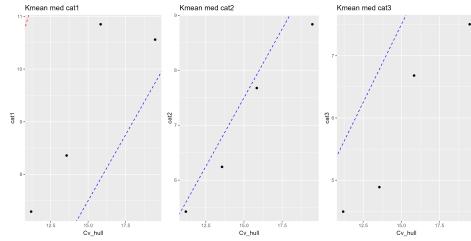


Figure 22: Kmeans parameter test on 'med' trees

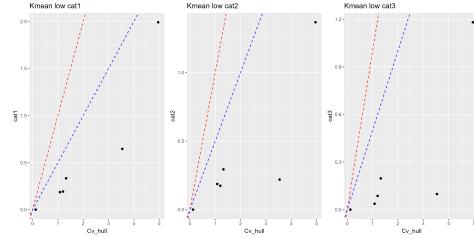


Figure 23: Kmeans parameter test on 'low' trees

#### 4.2.4 Voxel resolution

We set the voxel resolution to 0.1, 0.2, and 0.3. The voxel size needs to be small to accurately evaluate the volume, but if the size is too small, there is a risk of underestimating its value. We test three different values to assess this risk.

According to Figure 24, the method is not accurate for all trees.

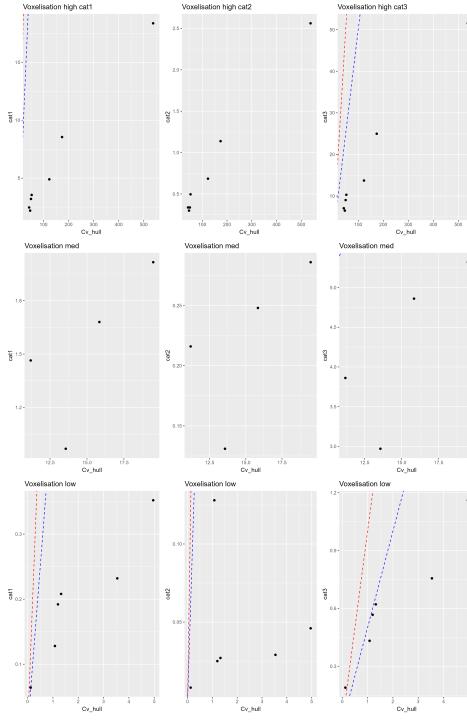


Figure 24: Voxel parameter test on each groups

#### 4.2.5 Summary of chosen parameters

Table 8 summarizes the final categories. We use a cross (X) to denote the methods that do not align with our criteria.

Table 8: Final parameters for each method

Method	High	Medium	Low
AMS3D	0.1	0.3	X
DBSCAN	0.5	1	X
Kmeans	4	X	X
Voxel	X	X	X

### 4.3 Results of convex hull by slice

According to Figure 25, the CH by the slice method is only accurate for the 'high' number of points group.

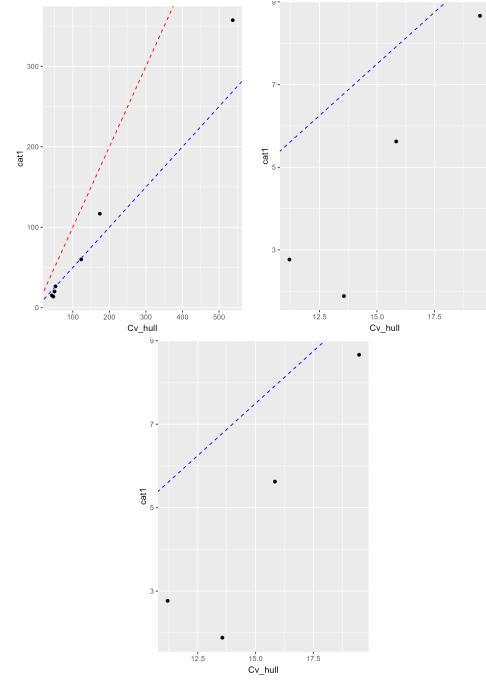


Figure 25: Convex hull by slice on each group

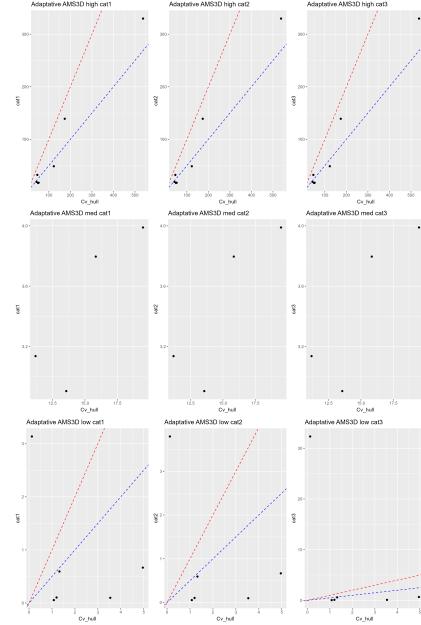


Figure 26: Adaptative clustering with AMS3D on each group

## 4.4 Results of adaptive clustering with AMS3D

Figure 26 highlights that the adaptative method with AMS3D is efficient for the 'high' number of points group.

## 4.5 Final methods

Finally, for the group of trees with a 'low' number of points, the only method that remain valid is the CH.

The volume of the group of trees with a medium number of points can be evaluated by computing the CH on points clustered using AMS3D and DBSCAN algorithms.

CH by slice, AMS3D, DBSCAN, Kmeans and adaptive AMS3D are accurate methods to evaluate the volume of trees with a 'high' number of points.

To compare these final methods, we define the reference volume as 70% of the CHV. [Zhou et al., 2022] refered that the CH overestimate the volume. We choose arbitrarily 70% due to the lack of a reference volume. We calculate the difference between the volume obtained by each method and the reference volume. This difference is then divided by the aimed volume to obtain a percentage error relative to the "reference" volume (Table 9).

Table 9: Relative error to the reference volume for each method in Montredon (%)

Group	$A_{AMS3D}$	$A_{AMS3D}$	$DBSCAN$	Kmeans	CH Slice
Low	X	X	X	X	X
Medium	-0.43	X	-0.37	X	X
High	-0.39	0.26	-0.52	-0.36	0.31

Results are similar for all methods. For trees with fewer than 100 points, the CH method proves to be the most accurate. DBSCAN demonstrates efficiency for trees with more than 100 points but fewer than 300, yielding a volume approximately 37% higher than the reference. The Adaptive AMS3D method is the most

accurate for trees with more than 300 points, achieving a volume reduction of around 25% relative to the reference volume.

## 4.6 Roy d'Espagne result comparison

The trees from the Roy d'Espagne site are used to confirm or refute the reproducibility of the methods. There are 13 trees in this area but we only have the point clouds of 10 of them. We did not obtain usable results for the segmentation of the 3 others. 9 trees of this area have more than 300 points. One tree has 106 points. We use the parameters established before.

Table 10: Relative error to the reference volume for each method in Roy d'Espagne (%)

Group	$A_{AMS3D}$	$A_{AMS3D}$	$DBSCAN$	Kmeans	CH Slice
Medium	0.08	X	-0.25	X	X
High	0.12	0.25	-0.25	-0.02	0.26

We obtain the results in Table 10 according to the reference method set at 70% of the CHV.

Here, the Kmeans method proves to be the most accurate in evaluating the tree volume for trees with more than 300 points in their point cloud. Additionally, applying the CH to AMS3D clustered points yields the best results for 'medium' size point cloud. Here only one cloud is in that group, this result is hardly interpretable.

This observation highlights the inadequacy of relying solely on a specific method. The results of each methods after selection are close to one another. The number of observations is not sufficient to conclude on a specific method.

However, the Roy d'Espagne observation indicates that a higher number of points significantly improves the accuracy of volume estimation, especially when employing clustering methods, which have demonstrated to be the most effective in this context.

## 4.7 Final volume

We summarize the final volume in Table 11. We refer to the volume obtained with the previous selected method as the Method Volume (MV). The MV is compared with the CHV calculated initially.

Table 11: CHV and final MV for each tree ( $m^3$ )

	CHV	MV
JUPH1	173.94	139.12
PIHA1	4.96	4.96
PIHA2	42.50	19.14
PIHA3	49.83	16.96
PIHA4	123.01	48.77
PIHA5	537.53	329.78
PIHA6	52.63	17.49
PILE1	1.32	1.32
PILE2	13.56	0.47
PILE3	1.20	1.20
PILE4	19.48	19.48
PILE5	15.83	15.80
PILE6	0.13	0.12
PILE7	46.58	32.32
PITO1	3.55	3.54
PITO2	11.19	11.19
QUIL1	1.07	1.07
PIHA1R	274.70	224.75
PIHA2R	22.49	16.77
PIHA3R	29.81	22.06
PIHA4R	423.52	357.80
PIHA5R	211.75	176.53
PIHA6R	455.69	394.10
PILE1R	8.07	2.58
PILE4R	44.28	28.48
PILE6R	1.22	1.07

## 4.8 Estimating AGB using tree features extracted from LiDAR measurements

We compute the AGB of *Pistacia lentiscus* using the formula provided by [Armand et al., 1993]. The formula takes into account the width, height, and length of each tree. We compare the obtained AGB using lidar features extraction with the AGB using field measurements. We plot the equation  $x = y$  with a red line to

evaluate the comparison in Figure 27.

Figure 27 underscores the effectiveness of estimating AGB through LiDAR feature extraction. AGB using LiDAR features tends to be higher than the AGB computed from field features. Field features extraction can be biased due to the inability to evaluate tree shape from above. Conversely, LiDAR feature extraction can be inaccurate due to erroneous data points. However, the results are satisfying, and with a larger dataset, we could more precisely estimate other tree features and potentially eliminate the need for field measurements. This would enable the application of various allometric equations, many of which are based on DBH.

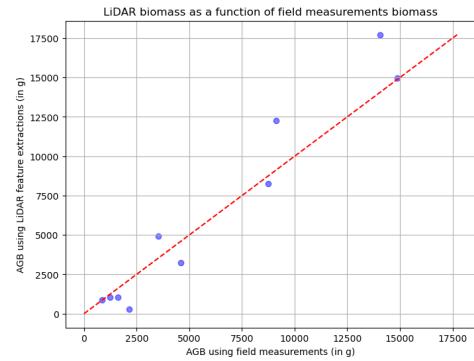


Figure 27: AGB using LiDAR feature extraction according to AGB using field measurements for *Pistacia lentiscus*

## 4.9 Converting volume to AGB using linear regression

### 4.9.1 For *Pinus halepensis* (PIHA)

To compute the AGB of *Pinus halepensis* in Montredon and Roy d'Espagne, we use the formula given by [Mitsopoulos and Dimitrakopoulos, 2007] and compute the AGB as

$$\ln(\text{foliar AGB}) = -3.766 + 1.811 \ln(\text{DBH}).$$

DBH is a field measurement. The aim is to establish a link between biomass and volume. We plot the AGB against the volume.

We compare the AGB to the volume from each methods : CHV, MV, AMS3D, DBSCAN,

Kmean, A-AMS3D and CV by slice (CVS). We plot the linear regression line, which is the line that minimizes the distance to the points. The volume appears to be proportional to the biomass as shown in Figure 28.

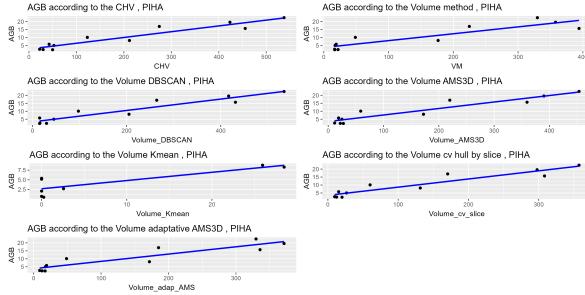


Figure 28: AGB according to the Volume, PIHA

We compute the Mean Squared Error (MSE). The MSE is the mean of the squared differences between the predicted AGB (the points on the linear regression line) and the actual AGB. We also compute the  $R^2$  coefficient.  $R^2$  is the proportion of AGB variance that is predicted by the volume variable. A score close to one indicates a good representation of the AGB by the volume (Table 12).

Table 12: MSE and  $R^2$  for each methods, 'PIHA'

Methods	MSE	$R^2$
CHV	5.17	0.90
MV	9.2	0.82
AMS3D	5.22	0.89
DBSCAN	5.76	0.88
Kmean	7.05	0.86
A-AMS3D	5.92	0.88
CVS	8.16	0.83

The results reveal that for PIHA, the volume obtained with CH shows a more consistent proportionality with the AGB compared with the volume extracted from other methods. The AGB is highly (90%) representable by the CHV. The methods used for PIHA include half Kmeans and half A-AMS3D volume. The results highlight their lack of accuracy in computing the AGB. As expressed before, we cannot establish a consistent method due to the limited number of points.

#### 4.9.2 For *Pistachia lentiscus* (PILE)

To evaluate the AGB of the *Pistachia lentiscus* we use the [Armand et al., 1993] formula:

$$AGB = 2.76 \times volume^{0.85}.$$

We carry out the same experiment for the PILE as for the PIHA.

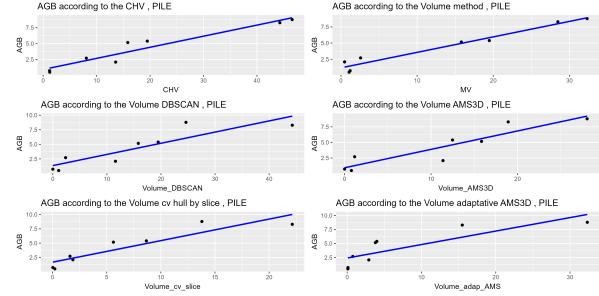


Figure 29: Volume according to the AGB, 'PILE'

We observe in Figure 29 that the AGB of the 'PILE' closely correlates with the volume, the line closely fits the points. We compute the MSE and  $R^2$  in Table 13.

Table 13: MSE and  $R^2$  for each methods, PILE

Methods	MSE	$R^2$
CHV	0.71	0.92
MV	0.41	0.95
DBSCAN	1.87	0.79
AMS3D	1.44	0.84
Slice	1.45	0.84
Adapt	2.72	0.70

In the case of *Pistacia lentiscus* trees, results of AGB estimation are close to the theoretical AGB. The method we established to compute the volume increases the precision of the AGB. The volume represents more than 95% of the variance, the information of the AGB.

### 4.10 Prediction of PIHA volume

The PIHA species is the most prevalent in our dataset. We predict the AGB only for these trees, as the number of trees in other species is low. We establish a linear regression model using data from 9 randomly selected trees, we test the

model on an independent set of 3 trees. However, due to the limited number of trees, the predictive accuracy can significantly enhance with a larger dataset. The linear model derived from the training trees yields  $R^2 = 0.87$  and  $MSE = 5.77$ . Using this model for predicting AGB, we compute the MSE between the predicted and true AGB, resulting in  $MSE = 2.43$ . These findings demonstrate the efficacy of estimating AGB from volume, as it can yield accurate results using a linear model trained on a larger dataset.

## 5 Discussion

The previously obtained volume results highlights that cloud point quality has a significant influence on the volume computation. In the scope of our study, the quality of our point clouds is problematic. Consequently, some individual point clouds contain too few points or have a density of points that is too low, especially trees with less than 100 points. As a result, distinguishing one tree from its neighbor during the extraction of individual point clouds is challenging and distinguishing the trunk from the foliage proves impossible. In these cases, the segmented individual point cloud can turn out to be coarse. However, for biomass calculation, studies as [Qiao et al., 2023] have developed allometric equations based on tree height and DBH. While the quality of our point clouds is sufficient to obtain height, it does not allow us to extract DBH. The crown often surrounds the trunk, so a measurement at 1.37 meters on CloudCompare gives the crown diameter at that height rather than the DBH. If the quality of our point clouds were higher, it might have been possible to differentiate these elements in the 3D cloud.

Furthermore, correctly evaluating the volume when the point cloud is filled with empty spaces is difficult. The CH method tends to overestimate the volume because it can not differentiate narrower tree shapes. Besides, clustering methods underestimate the volume, particularly because of areas where there are no points but where the tree exists, leading to inaccuracies in volume computation. We choose the volume of the CH of the point cloud as the reference volume and compare the results from

different clustering algorithms to this reference. However, this choice of reference is arbitrary, so a field reference volume could allow us to discuss more precisely the best method to apply. A reference volume for each species could also be conducted on more substantial datasets.

We tested three parameters for each method. For more advanced parameterization, exploring a wider range of parameters and employing model accuracy evaluation methods such as cross-validation could enhance the robustness of the analysis.

We developed our methods based on a limited number of individuals. Thus, we did not conduct a statistical study on increased sample sizes to validate or select the best method to evaluate each volume. Studying a forest predominantly populated by a given species would allow for a more substantial dataset, establish more reliable statistical models for each species, and finally apply the findings to our target individual trees.

Extracting each point cloud from the complete cloud is very time-consuming and tedious, requiring comparison with field photographs, which can constrain the applicability and repeatability of the proposed methods, thus diminishing the advantages of satellite data extraction. If the objective is to estimate biomass for a large number of individuals or over a very extensive area, one can understand the relevance of studies focusing on stratification by complete vegetation barriers rather than individuals [Ferraz et al., 2012].

Typically, when data on the scale of individual trees are accessible, using biomass models tailored to such data, often relying on measurements such as DBH [Brown, 2002], [Di Cosmo et al., 2016] is preferable. The limited literature on simple relations between volume and biomass made it difficult to establish straightforward guidelines for estimating tree biomass from the obtained volumes.

Allometric equations involving a simple relationship between apparent volume and AGB do not necessarily exist for all tree species. Since the complete volume of the tree is taken, no

distinction is made between the trunk, branches, and leaves, which nonetheless have different densities and AGB per  $m^3$ . This type of simple relation is more commonly used to quantify biomass per area. Thus, [Di Cosmo et al., 2016] developed a model to predict AGB density ( $Mg\ ha^{-1}$ ) from growing stock volume density ( $m^3\ ha^{-1}$ ) by forest category. Furthermore, if they exist, linear regression formulas between volume and biomass can vary from one study to another.

## 6 Conclusion

To evaluate the AGB of a tree we used a green volume approach based on an ALS point cloud. This study experiments a CH method applied on different data points. Data points were clustered with AMS3D, DBSCAN, Kmeans and slice algorithm. Voxels methods were unaccurate. The results show that the input parameters of the different methods significantly affects the volume estimation. We found that the parameters depend on the data cloud number of points. We delineated 3 groups and established results on each of them. We only had 21 trees and were not able to establish statistical analyses on the obtained results. Thus, we set a reference volume as 70% of the CH volume. From that, we obtained that trees with less than a 100 points are not eligible for clustering methods because they underestimate the volume. We used the volume given by the CH for these trees. The volume of the group with more than 100 points but less than 300 shows to be more accurate when the CH is used on points clustered with AMS3D and DBSCAN. AMS3D computes a higher volume than the referenced volume from 37%, we obtain 42% for DBSCAN. We can not confirm the repeatability of the method due to the few number of trees. Finally, trees with more than 300 points have accurate results with all methods. Adaptive AMS3D and CH by slice are the top performers for evaluating the volume these trees. We obtained respectively 25% and 30% higher volume than the reference value. To evaluate the methods on unknown trees, we used the Roy d'Espagne data set that has 10 trees. The points clouds of these trees have a higher number of points than de Montredon trees with an average of 2241 points. We showed that a high number of points increases the accuracy of the volume

estimation. Adaptative AMS3D and CH by slices are more stable than the clustering methods. The volume they provide is about 25% to 30% different from the reference volume for both data set. With AMS3D, DBSCAN and Kmeans the accuracy of the volume is respectively increased by 20%, 30% and 30% on the tree with a high number of points. We computed the AGB from tree features such as DBH or height. We showed that the obtained AGB is linearly correlated with the volume. This result allows us to predict the AGB from the volume using a linear regression model.

## Acknowledgment

We would like to thank our two tutors, Sophie Fabre and Lucie Calmon, for their support and assistance throughout the entire project. Thank you for introducing us to your research topic and for allowing us to participate in various aspects. Lucie, we wish you the best in the pursuit of your thesis and will be attentive to its progress.

Finally, we would like to thank our English teacher, Rebecca Coles, for initiating us to scientific English. Thank you for all your feedback on our article and presentation slides. The entire project was very motivating, and having your feedback was particularly helpful.

## References

- [Armand et al., 1993] Armand, D., Etienne, M., Legrand, C., Marechal, J., and Valette, J. (1993). Phytovolume, phytomasse et relations structurales chez quelques arbustes méditerranéens. *Ann SCi For.*
- [Barthelemy and Hérat, 2016] Barthelemy, C. and Hérat, A. (2016). Le beau et le dangereux, le protégé et le pollué. Dissonances paysagères aux abords du massif des Calanques. *Les Carnets du paysage.*
- [Brown, 2002] Brown, S. (2002). Measuring carbon in forests: current status and future challenges. *Environmental Pollution*, 116(3):363–372.
- [Calmon, 2022] Calmon, L. (2022). La végétation native méditerranéenne peut-elle jouer un

rôle de barrière à l'envol des particules contenées sur une friche industrielle ? *Rapport de stage*.

[Calmon, 2024] Calmon, L. (2024). In-situ measurement on montredon and roy d'espagne area.

[Chinrungrueng and Sequin, 1995]

Chinrungrueng, C. and Sequin, C. (1995). Optimal adaptive k-means algorithm with dynamic adjustment of learning rate. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 6:210–223.

[Di Cosmo et al., 2016] Di Cosmo, L., Gasparini, P., and Tabacchi, G. (2016). A national-scale, stand-level model to predict total above-ground tree biomass from growing stock volume. *Forest Ecology and Management*, 361:269–276.

[Ferraz et al., 2012] Ferraz, A., Bretar, F., Jacquemoud, S., Gonçalves, G., Pereira, L., Tomé, M., and Soares, P. (2012). 3-d mapping of a multi-layered mediterranean forest using als data. *Remote Sensing of Environment*, 121:210–223.

[Ferraz et al., 2016] Ferraz, A., Saatchi, S., Mallet, C., Jacquemoud, S., Gonçalves, G., Silva, C. A., Soares, P., Tomé, M., and Pereira, L. (2016). Airborne lidar estimation of above-ground forest biomass in the absence of field inventory. *Remote Sensing*, 8(8):653.

[Google, 2024] Google (2024). Location of marseille in france, and montredon and roy d'espagne in marseille. <https://www.google.com/maps/place/Marseille>. Consulted on march, 2024.

[Mitsopoulos and Dimitrakopoulos, 2007]

Mitsopoulos, I. and Dimitrakopoulos, A. (2007). Allometric equations for crown fuel biomass of aleppo pine ( *pinus halepensis* mill.) in greece. *International Journal of Wildland Fire - INT J WILDLAND FIRE*, 16.

[Qiao et al., 2023] Qiao, Y., Zheng, G., Du, Z., Ma, X., Li, J., and Moskal, L. M. (2023). Tree-species classification and individual-tree-biomass model construction based on hyperspectral and lidar data. *Remote Sensing*, 15(5):1341.

[Straker et al., 2023] Straker, A., Puliti, S., Breidenbach, J., Kleinn, C., Pearse, G., Astrup, R., and Magdon, P. (2023). Instance segmentation of individual tree crowns with yolov5: A comparison of approaches using the forinst instance benchmark lidar dataset. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 9:100045.

[Testiati et al., 2013] Testiati, E., Parinet, J., Massiani, C., Laffont-Schwob, I., Rabier, J., Pfeifer, H.-R., Lenoble, V., Masotti, V., and Prudent, P. (2013). Trace metal and metalloid contamination levels in soils and in two native plant species of a former industrial site: evaluation of the phytostabilization potential. *Journal of Hazardous Materials*, 248:131–141.

[Torre, 2018] Torre, L. (2018). Biomass estimation using lidar data. *International Journal of Sustainable Energy Planning and Management*, 17:79–90.

[Tusa et al., 2020] Tusa, E., Monnet, J.-M., Barré, J.-B., Dalla Mura, M., Dalponte, M., and Chanussot, J. (2020). Individual tree segmentation based on mean shift and crown shape model for temperate forest. *IEEE Geoscience and Remote Sensing Letters*, 18(12):2052–2056.

[Zhou et al., 2022] Zhou, L., Li, X., Zhang, B., Xuan, J., Gong, Y., Tan, C., Huang, H., and Du, H. (2022). Estimating 3d green volume and aboveground biomass of urban forest trees by uav-lidar. *Remote Sensing*, 14(20):5211.