Research internship report

Semi-automatic estimation of the tree biomass using terrestrial LiDAR data

（利用陸地激光雷達數據半自動估計樹木生物質）

Thibault CHARDON

2nd year of ASTER undergraduate program
From 05/28/2023 to 08/11/2023 (112 年)

Tutor in laboratory: Pr. I-Fang SUN 孫義方
Academic tutor: Pr. Taichi KAWAMURA
Host laboratory: Center for Interdisciplinary Research on Ecology and Sustainability (CIRES)
Host university : National Dong Hwa University, Taiwan 國立東華大學
Higher education institution: Paris Cité University, France

This report is written in English, except a part of the abstract and appendixes written in Mandarin Chinese.

Submission: August 18th
Oral defense: August 28th

Acknowledgments

I first want to thank Pr. Sun for his trust in me since the beginning of our exchanges, that made me feel useful for the work that he and his assistants undertake everyday. As an undergraduate student, having to create valuable work is an incredible opportunity that I am glad to have experienced at National Dong Hwa University.
I also want to express my gratitude to all of the assistants working at CIRES – to quote 大潘，婷安，慈芳，延平，涼凱，立亘，謙諾 and 洪慈; as well as the other professors and fellows I had the chance to meet: Pr. Chen, 小跳, and many others during these three months. They all gave me so much kindness that allowed me to enjoy my journey in Shoufeng.
I also want to thank the other volunteers that were present at Fushan botanical garden, where we shared so much time climbing trees and mountains as well as eating generous dinners after hard work.

In France, Paris Cité University, I want to thank Pr. Kawamura for ensuring the continuation of the ASTER program – that is now more than ten years old – and its very unique trait consisting in gaining experience in planetary, environmental and Earth sciences while learning an oriental language.
I also want to thank the former students that transmitted the knowledge that they got from their own experiences, and by doing so contributed to give their juniors the chance to live always greater adventures during this special summer of sophomore year.

Ultimately, I would like to thank everyone else that participated in making that internship possible, as well as those that I met while travelling and having a great time all around this marvelous island and its unending landscapes.

Abstract

In tropical ecology, forests are a core topic. They are usually studied from the perspective of dynamic plots, consisting of multiple subplots of a given area, containing diverse information about trees or external conditions and their temporal evolution. While measurements can be taken on-field, some key parameters are requiring more technical methods to be assessed. This report presents a comprehensive study on the development and application of a model for estimating above-ground biomass (AGB) within a recently planted forest in eastern Taiwan, using terrestrial laser scanning (TLS) data. The primary objective is to provide AGB estimations at a subplot scale.

The LiDAR point cloud data, acquired via TLS, undergoes an intensity filtering process to extract exploitable tree-related information. It then follows a supervised machine learning classification that separates trees from terrain, based on user-provided training samples.
The model employs a combination of density-based spatial clustering of applications with noise (DBSCAN) and voxel's principal component analysis (PCA) based segmentation to separate individual trees within the remaining point cloud. A region-growing algorithm is then employed to expand and refine the segmented regions, followed by a reinforcement process to enhance tree structure using nearest neighbor assignment.
The subsequent quantitative structure modeling (QSM) iteratively reconstructs each tree using cylindrical segments, allowing for volume estimation. By knowing each tree's species and its associated density, our model can finally provide an estimation of AGB in kilograms.

The model's performance is assessed through comparison between on-field measurements and model's estimations of height and diameter at breast height (DBH). The model demonstrates coherent trends when focusing on the sets of individual species. However, challenges remain in processing taller trees and accurately estimating their AGB. Due to the lack of direct measurements of AGB, that would require destructive methods, the reliability of our model is difficult to assess. Ultimately, model's assumptions make its use quite restrictive, and asks for cautious interpretations of the results since they imply a noticeable uncertainty.
Despite these challenges, the model presents a valuable tool for assessing forest structure and dynamics. It offers insights into biomass distribution, forest health, and potential impacts of land use changes; it can play a role in informing land management policies and facilitating decision-making processes. The model's limitations underscore the need for further refinements and enhancements to provide more accurate estimations, that could be made at an individual tree scale.

在熱帶生態學中，森林是一個核心話題。通常從動態樣區（FDPs）的角度來研究，該樣區由多個給定面積的子樣區組成，包含了關於樹木或外部條件以及它們隨時間變化的多樣信息。雖然可以在現場進行測量，但一些關鍵參數需要更為專業的方法來評估。本報告介紹了一個關於台灣東部新種植森林利用地面激光掃描（TLS）數據估算地上生物量（AGB）的模型的全面研究。主要目標是在子樣區尺度上提供 AGB 估算。

通過 TLS 獲取的 LiDAR 點雲數據經過強度過濾處理，以提取與樹木相關的信息。然後，採用監督式機器學習分類方法，基於用戶提供的訓練樣本將樹木與地形分開。
該模型結合了基於密度的空間聚類應用與噪聲（DBSCAN）以及基於體素主成分分析（PCA）的分割方法，從剩餘的點雲中分離出各個樹木。隨後採用區域生長算法來擴展和細化分割區域，接著通過最近鄰分配來增強樹木結構。
隨後的定量結構建模（QSM）通過迭代重建每棵樹木，使用圓柱段進行體積估算。通過了解每棵樹的物種及其相關密度，我們的模型最終可以提供以公斤為單位的 AGB 估算。

通過比較現場測量和模型對高度和胸徑（DBH）估算的結果，評估了模型的性能。模型在關注單個物種的情況下表現出一致的趨勢。然而，在處理較高的樹木並準確估算其 AGB 方面仍存在挑戰。由於缺乏直接的 AGB 測量方法（需要破壞性方法），我們很難評估模型的可靠性。最終，模型的假設使其使用受到限制，並且對結果的解釋需要謹慎，因為它們暗示了明顯的不確定性。
儘管存在這些挑戰，該模型提供了一個有價值的工具，用於評估森林結構和動態。它可以提供有關生物量分佈、森林健康以及土地利用變化潛在影響的見解；它可以在製定土地管理政策和促進決策過程方面發揮作用。模型的局限性強調了需要進一步改進和增強，以提供更準確的估算，這些估算可以在個體樹木的尺度上進行。

Table of contents

List of figures

List of tables

Introduction

The study of forest dynamics has many applications, from mitigation of climate change by estimating fluxes such as carbon dioxide balance and its evolution, to water, food or wood supply by measuring parameters leading to assess the health of each individual tree. Since trees are the main actors in forests, focusing on their individual behavior and evolution with time can reveal general trends. In that optic, forest dynamics plots (FDPs) aim to compile extensive variables, external weather or geological events, and trees parameters; by doing so, and periodically carrying out diverse censuses, it enlightens changes happening in a temporal perspective at a forest scale. Since the interdisciplinarity has crossed ecological topics, this vision of whole ecosystems is key, because it makes it easier to interact with external actors of the forest; especially diverse human users – farmers, energy suppliers, decisionmakers, and even walkers – that have to come to terms to ensure a fair and sustainable management of the forest. This socio-ecological perspective is in its early stages today, and presents a vast realm of research exploration, as numerous interactions between environmental and social sciences come to the forefront.

During censuses, direct measurements can be taken on trees: their height and diameter at breast height (DBH) are quantifiable, and discrete parameters such as survival status, fungal presence, lianas strangling the stem, visible damages, etc. can be acquired on field. Some parameters are only quantifiable as a variation (often by using a percentage of change), for instance tree biomass cannot be precisely measured manually – at least without having to destroy the tree – though the lack of leaves of a tree and therefore part of its biomass absolute variation, can be assessed by eye. It is also possible to indirectly estimate biomass by using several measurements (Brown et al., 1989; Somogyi et al., 2007; Fu et al., 2017), although the uncertainty using such methods is made much bigger. To offer a second perspective to censuses data, as to better estimate indirectly assessable parameters such as biomass – especially woody biomass – innovative methods have been proposed in literature.

Since its use became popular and cheaper, laser imaging detection and ranging (LiDAR) has been implemented in several works to approach on-field measurements and by doing so, provide reliable extra information for each individual, such as biomass. It consists in an outgoing light ray (mostly infrared) that is then scanned after having been reflected on a nearby surface, so that the elapsed time and the intensity of the incoming ray give information about the three-dimensional space surrounding the scanner. Such methods generally need cutting edge analyses to be fully exploited, hence, diverse models are co-existing while progress is made in 3D-point cloud computing.
There are two main categories of LiDAR machinery: one is the terrestrial laser scanning (TLS), the other is aerial laser scanning (ALS). In forest and trees modeling, both play interesting roles; while TLS is efficient in acquiring trunk and thicker stems of the crown, ALS operates a clearer scanning of leaves and thinner stems. This leads to often merging both points clouds, especially for estimations needing leaves density to be calculated (Fekry et al., 2022; Paris et al., 2017). Here, we will only have as a material a TLS-scanned point cloud, thus the tree biomass is still assessable despite the estimation should carry more uncertainty. To be precise, the measurable biomass is only the one above ground, also known as above-ground biomass (AGB). We defined it here as excluding lianas, or trees that are smaller that ~1.3 meter – DBH point of measurement – or thinner than 10 centimeter at ~1.3 meter. Using this definition of AGB, we will try to build a model efficient and reliable in estimating AGB of an eastern Taiwan tropical forest plot.

Since the use of LiDAR has been popularized, approaches for estimating parameters of a tree cloud – a 3D-point cloud of an individual tree – have been widely developed with a great flexibility allowing all sorts of trees to be assessable. Quantitative structure modeling (QSM) is a commonly used method consisting of geometrical simplification of the tree's structure, that can be then described with new parameters resulting of this simplification, such as the volume of the tree, leading to estimate AGB. QSM can produce great results if the provided tree cloud is correctly differentiated from other trees, and has been efficiently noise-filtered. In other words, most of the challenge reposits on the differentiation between individual trees, and cloud denoising. It is also necessary to separate ground, or terrain, from vegetation, and to ignore too small trees and non-tree vegetation, since we want to focus on trees with a measurable DBH (so necessarily longer than 1.3 meter). Especially on dense forests, this can be very strenuous. Thus, most of the work will consist on constructing high quality tree clouds.

In this report, we will first give a glance at the material used and the studied area (1), to then propose a methodology seeking for the production of individual tree clouds, and later compute QSM on each tree (2), leading to the comparison of QSM-based measurements and on-field measurements that we will discuss on a third part (3). Finally, we will conclude by addressing the limitations of our model and exploring its interdisciplinary applications, including a brief reflection on the practical work undertaken during the internship and the related experiences that contributed to the composition of this report (4).

Material

Studied area

We will concentrate on a specific forest plot consisting in multiple subplots, utilizing one of these subplots as an exemplar in our methodology. In this context, a subplot refers to a defined forest area within the plot, characterized by parameters of its vegetation or external conditions such as temperature, humidity, or elevation gradient. The results and discuss section will consolidate findings from all the subplots that were surveyed.

Our forest plot is situated on an alluvial deposit (Lin & Chen, 2016) in the longitudinal valley, known as the Huadong valley, located in Eastern Taiwan. It benefits a temperate humid climate with hot summers (Köppen classification: Cfa).

Hualien county has the third lowest density of population in Taiwan with 70.13 people per square kilometer (National Statistics, DGBAS, 2020), principally because of the mountainous areas counted in the county's surface. The surrounding area is located in the most rural part of the county. The forest plot is near Mizhan and Beilin villages, within Fenglin township. The average age of Hualien county's population is 43.6 years, with 12.3 % of children under the age of 15, 69.6 % of active population (15 to 64 years old) and 9.7 % of people being more than 64 years old; 3.6 % of the active population in the county, which represents 8,817 people, are working in the primary sector (agriculture, forestry, fishing and animal husbandry), compared to 2.7 % in all Taiwan (*ibid.*). 32 % of the county's active population have followed or is following a cursus in higher education, and 17 % have been studying until a junior high school degree, versus respectively 43 % and 15 % at national scale (*ibid.*). National unemployment rate is established at 3.46 % with mostly young people contributing to the statistic – 11.52 % of the 19 to 24 years old population (National Statistics, DGBAS, May 2023).

In 2020 in the East district of Taiwan, the damage loss of forest area was assessed to 6.3274 ha, the felling of trees reached 3.80 ha, while reforestation was estimated to 53.64 ha. Between 2011 and 2020, the maximum damage loss, felling, and reforestation surfaces were respectively of 6.3274 ha (2020), 34.26 ha (2015), and 337.15 ha (2011) (Forestry Bureau, CoA, 2021). We can observe a decreasing trend of the reforestation during this period: it marks the end of several decades of massive afforestation (creation of new forests), mainly done in former agricultural lands, that followed a long period of urbanization with a massive logging of primary forests, which has been banned in 1989 (Chen et al., 2019).

Our forest plot is located in a former agricultural land. The first trees have been planted around 2005. There are seven different species: *bischofia javanica*, *swietenia macrophylla*, *terminalia catappa*, *fraxinus griffithii*, *alstonia scholaris*, *pterocarpus indicus* and *zelkova serrata*. It is typically a recent secondary forest, which is noticeable for estimating potential damages caused by external perturbations, since young man-planted forests generally implies a less resilient ecosystem than in primary forests (Thompson et al., 2009).

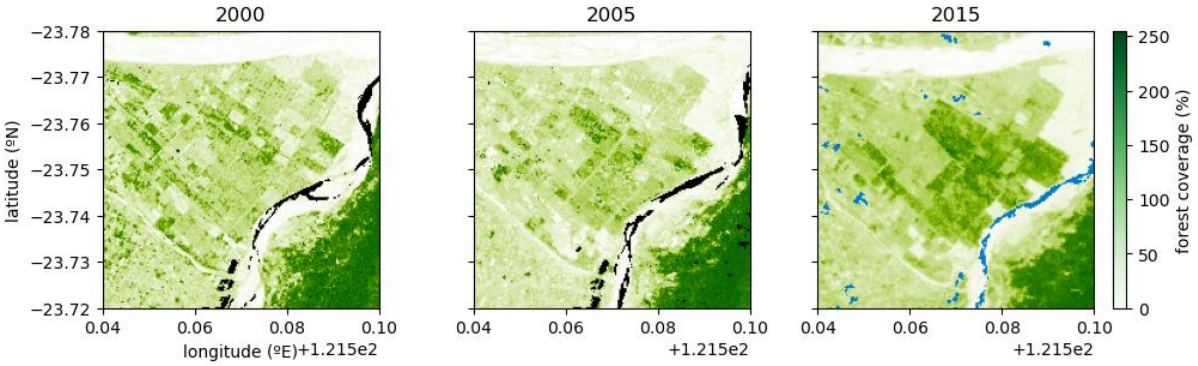


FIGURE 1: FOREST COVER EVOLUTION OF THE FOREST PLOT AREA (2000-2015)
Data from Townshend (2016), initially acquired with Landsat 5 and Landsat 7 thematic mappers (TM/ETM+). Resolution: 30 meters.

The plot is close to the active Lingding fault (Lin et al., 2022), a reverse fault which is associated with a 14.1 % probability for an earthquake of 6.5 or more of magnitude (Richter scale) to happen in the next fifty years (Central

Geological Survey, Taiwan MoEA, 2021). This earthquake hazard is associated with potential damage to living trees, leading ultimately to biomass loss (Allen et al., 2020).

The Eastern coast of Taiwan is frequently hit by typhoons, that are causing damage to trees and are not negligeable in the dynamics of a tropical forest plot (Hogan et al., 2018).

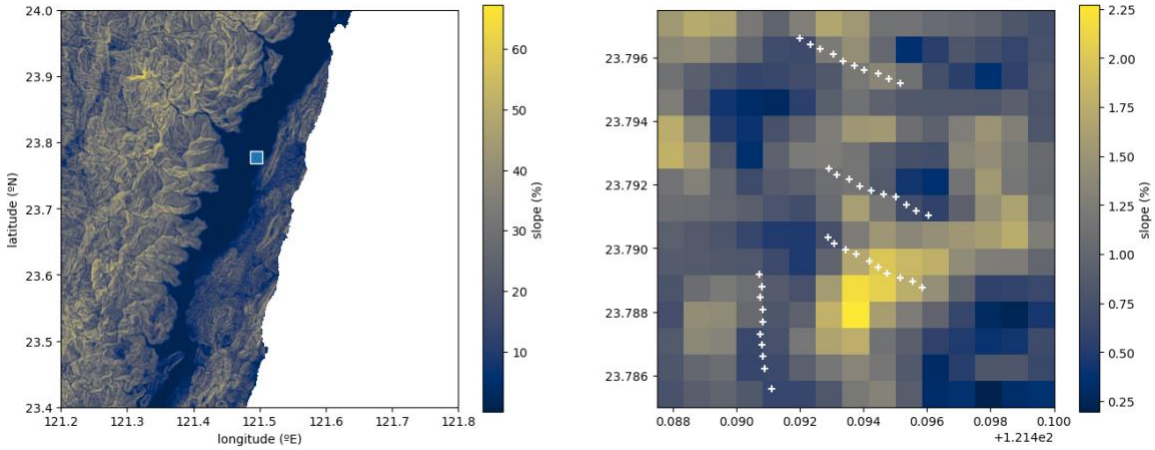


FIGURE 2: TOPOGRAPHIC MAP OF HUALIEN COUNTY AND THE FOREST SUBPLOTS
Slope data from Amatulli et al. (2020); resolution: 90 meters; distributed by OpenTopography.

The local climate is very likely to change because of climate change, inducing new conditions of temperature and precipitations. In IPCC's RCP 8.5 global scenario for 2081-2100, median annual temperature of East Asia area is predicted to rise of 5.2 ºC (relative to 1850-1900 period), and median total precipitation per day to be 6.4 % higher (IPCC, 2021). This up-shifting of medians means that extreme events – here, extreme rainfalls and heatwaves – are way more likely to happen. Eastern Taiwan is also sensible to ocean-atmospheric events such as ENSO (El Niño Southern Oscillation), that is likely to increase in amplitude with climate change (Cai et al., 2021). Thus, the forest plot that we are studying should experiment a tropical monsoon climate (Beck et al., 2018: Figure 3), similarly to the current climate in Southern Taiwan.

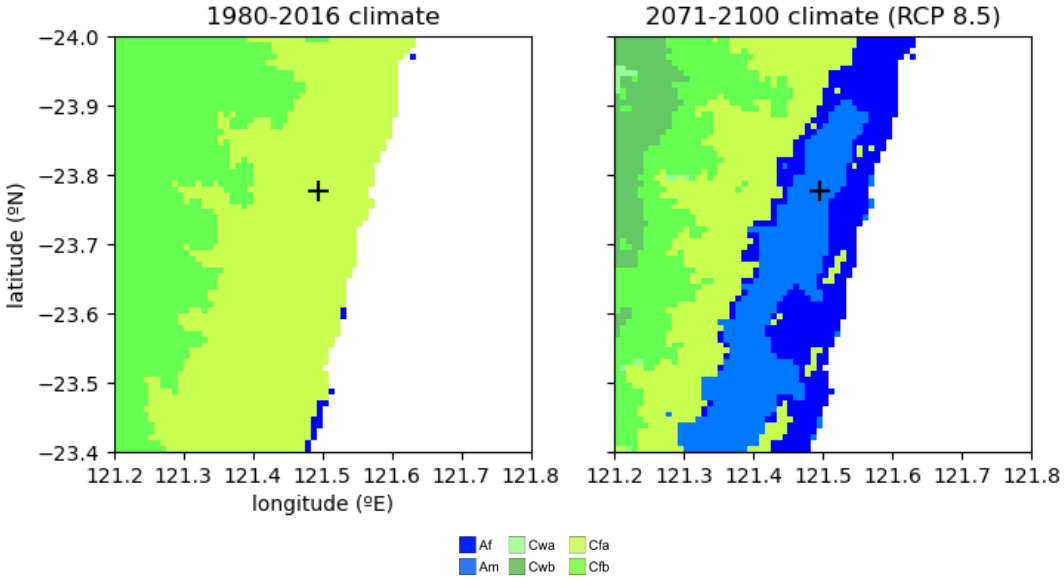


FIGURE 3: PRESENT AND FUTURE KÖPPEN CLIMATE CLASSIFICATION OF HUALIEN COUNTY
Data from Beck et al. (2018). Resolution: 1000 meters. Af: tropical, rainforest; Am: tropical, monsoon; Cwa: temperate, dry winter and hot summer; Cwb: temperate, dry winter, warm summer; Cfa: temperate, no dry season, hot summer; Cfb: temperate, no dry season, warm summer. The location of our forest plot is indicated.

LiDAR scanner

We are using a terrestrial laser scanner (TLS) that can be handheld (1.45 kg): *ZEB Horizon* from *GeoSLAM*. It has a range of 100 meters, a vertical angular resolution of 2º and a horizontal angular resolution of 0.2º. It has a frame of view of 360ºx270º and a relative accuracy of 6 millimeters. The scanner consists of a laser (class 1, wavelength of 903 nanometers) emitted from the device, and 16 sensors on this same device that will compute for each point of the scanned space the return time of the laser to determine the distance of an object and assign it to a region of the 3-dimensionnal space. The object is characterized by an intensity, being actually the relative intensity of the returning laser ray, compared to the outgoing ray: if the intensity is great, the object is assumed to be solid so that most of the laser ray has been reflected by the object, at the opposite, thinner objects may have a lower reflectance, inducing a low relative intensity (note that the computed intensities are often standardized into positive integers).

Datasets

In addition of unprocessed LiDAR data, we also have available on-field datasets (A), so that for each subplot we have information about each individual tree (species, DBH, height, survival status). Nevertheless, since we are not insured about the accurate location of each individual tree, we cannot correlate on-field measurements and our model's results at such scale – so we will compare them at a subplot scale.
We also have access to temperature and humidity data thanks to multiple *Hobo* sensors placed within the plot.

Required software and alternatives

LiDAR acquired data can be visualized as well as converted into a parsable dataset using the proprietary software of the scanner, *GeoSLAM*. For other scanners, or even cellphones with a built-in LiDAR scanner, other software should exist in the same purpose.
For further visualization and part of the processing, we will use *CloudCompare*, which is an open-source software for 3D-point cloud and mesh processing.
For computing QSM and further calculations, we will use *3D Forest*, a specialized open-source software for forest environments acquired by TLS.
Most of the computation, visualization, and statistical assessment, will be operated using *Python 3* (versions 3.8 and 3.9 have been tested). Any other high-level programming language can of course be used to produce similar results.

A GitHub repository, containing more precise hardware and software specifications, as well as the detailed scripts, should be available right here, with eventual updates made since the submission of this report: https://github.com/Thibalt-C/tree-biomass-estimator/

Methodology

Data pre-processing

After having correctly collected our data on field, we use here the *GeoSLAM* software that enables us to mark points ("control points") and then determine the spatial location of the cloud for further visualization on geographic information system (GIS) platforms. We can also at that time export the point cloud with several options of file types. We have then the possibility to down-sample our point cloud by choosing the percentage of points to keep; this is useful for too big clouds and greatly improves the computing time, while empirically keeping most of the information – despite it requires different parameters for computing well, since the global density mechanically plummets while down-sampling such data.
We will export the data as a *.las* file (for LASer) – being among the most commonly used file formats for point-cloud analysis – and test multiple down-sampled files (10 %, 50 %) in addition to a file containing all initially scanned points (100 %). If not precised, the file computed is the one containing 10 % of the initial points. We assumed that it was the most profitable according to its reasonable processing time and the results it gives.

Determination of a region of interest (ROI)

After having imported the *.las* file and converted it to a matrix of *n* by 4, with *n* points and 3 dimensions plus the recorded intensity of each point, we can look for the repartition of the points in space (Figure 4). We can then restrict manually *x* and *y* dimensions (here defined as the axes forming the horizontal plane of space), so that the zone we are focusing on becomes more homogeneous. It is often necessary because of the spherical shape of the data collected, since most of the computing methods are fitted for cube shape data. If there still are obvious outliers far from the other points, it will be simple to remove them by restricting the *z* (vertical) dimension. By doing so, we have spatially restricted our file to a region of interest (ROI) that we will focus on thereafter. Note that this selection of the ROI can be automated if we assume that the points repartition follows a normal distribution (which is enlightened in Figure 4). It is then convenient to choose a percentage of points $p \times 100$ that will be kept following the mean position $(\bar{x}, \bar{y})$ and the standard deviation (σ_x, σ_y) of the points in the (*x*,*y*) plane, by computing the probit function $\Phi^{-1}(p)$ of Bliss (1934) (1). For instance, $\Phi^{-1}(0.95) \cong 2$ (i.e. for 95 %).

$$(1)\quad \begin{pmatrix} \mathrm{ROI}_{x|min,max} \\ \mathrm{ROI}_{y|min,max} \end{pmatrix}(p) = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} \pm \Phi^{-1}(p) \times \begin{pmatrix} \sigma_x \\ \sigma_y \end{pmatrix}$$

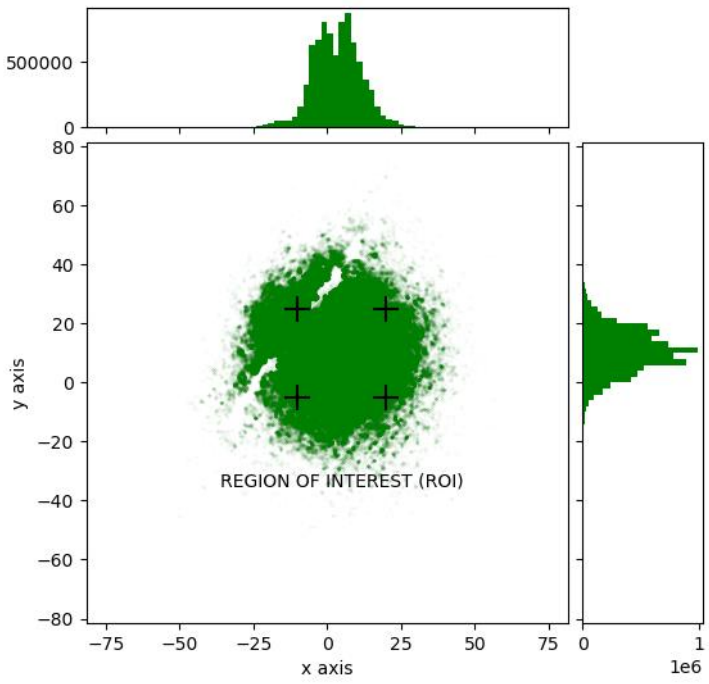


FIGURE 4: SPATIAL REPARTITION OF THE INITIAL POINT CLOUD
Up and right histograms represent local densities – i.e. number of points in a thin slice – here every 2 meters horizontally.

Preliminary filtering using intensity

As we look our ROI, we can observe that its density is too high for having a clear view of individual trees, due to a considerable presence of leaves, lower vegetation, and noise. By observing the point cloud with colored points based on the intensity (which can be done directly on *GEOSlam*), we can see that trees main stems have a greater intensity than the other parts of the cloud, despite this intensity also varies with the height (because the distance from the scanner is also influencing the acquired intensity of a point). Thus, we can horizontally slice our ROI, and apply an intensity filter that removes points with a too low intensity, relative to the height. We define the threshold Γ of each slice as the slice's mean intensity $\bar{\gamma}$ plus two times the standard deviation σ (2), leading to different thresholds for each slice (Figure 5, Table 1). This leads to a filtered point cloud with much of the unwanted points gone, and tending to be more efficient with a reasonable number of slices (Figure 6). Some parts of the trees, such as small stems and leaves, will have to be recovered later, for now their removal is necessary to further perform the segmentation of individual trees.

$$(2)\quad \Gamma = \bar{\gamma} + 2\sigma = \bar{\gamma} + 2\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})^2}$$

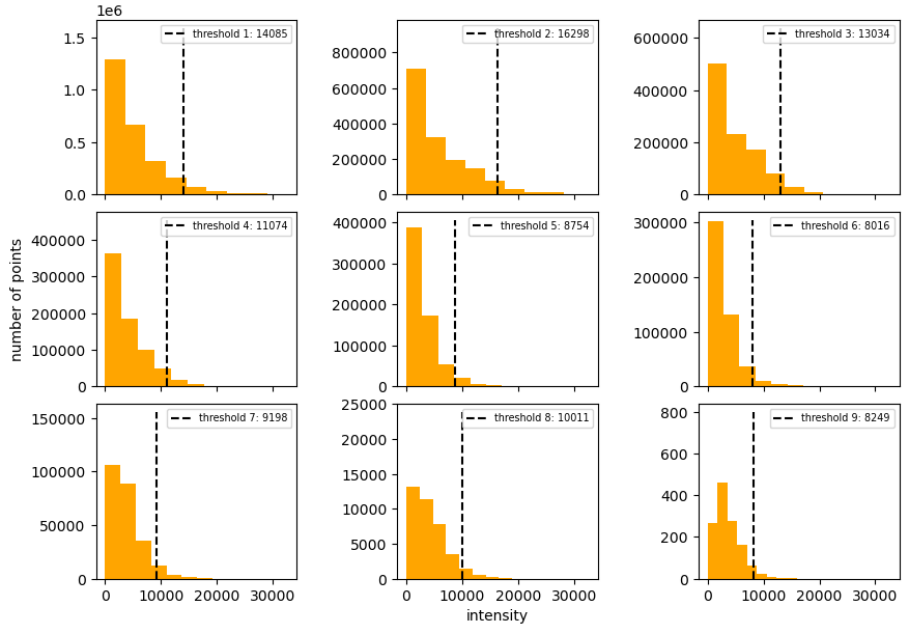


FIGURE 5: REPARTITION OF POINTS ACCORDING TO THEIR INTENSITY FOR 9 VERTICAL SLICES
The intensity is a relative value comparing for each point the strength of the returning laser, thus it does not have a unit of measure.

Height bounds of the slice (m)	Mean intensity $\bar{\gamma}$	Standard deviation σ	Intensity threshold **Γ**	Percentage of points above **Γ**
-0.7 – 0.9	5041	4522	14085	5.17 %
0.9 – 2.6	5670	5315	16298	4.97 %
2.6 – 4.3	4833	4101	13034	4.78 %
4.3 – 6.0	3993	3541	11074	4.49 %
6.0 – 7.7	3128	2813	8754	4.70 %
7.7 – 9.4	2958	2529	8016	4.22 %
9.4 – 11.1	3787	2705	9198	4.68 %
11.1 – 12.8	4200	2906	10011	4.78 %
12.8 – 14.5	3568	2341	8249	4.47 %

TABLE 1: MEAN INTENSITY, STANDARD DEVIATION AND INTENSITY THRESHOLD OF EACH VERTICAL SLICE, FOR 9 SLICES

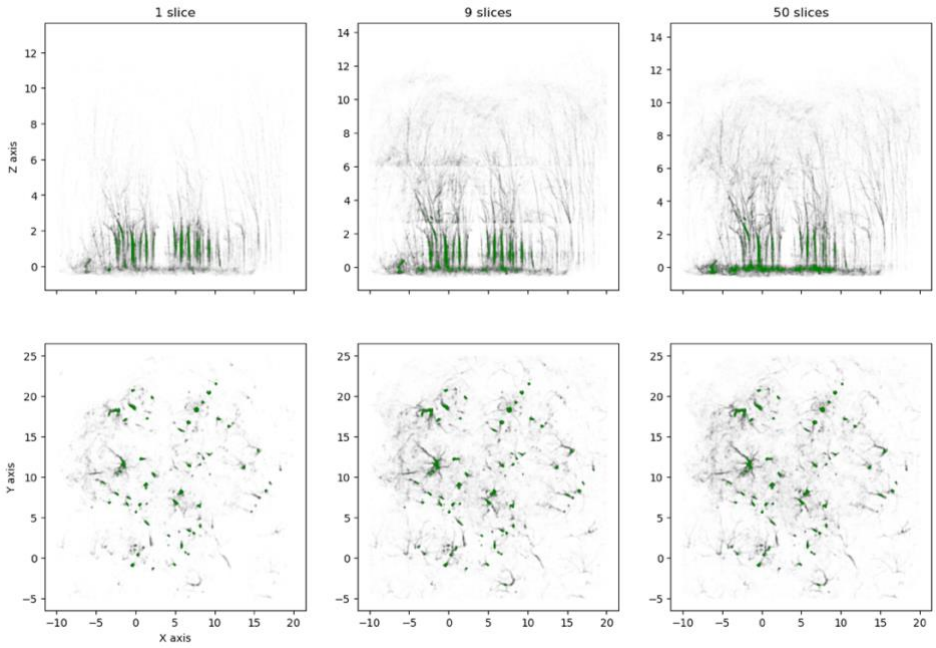


FIGURE 6: SPATIAL REPARTITION OF THE FILTERED POINT CLOUD
NB: here, the z dimension is deformed compared to x and y dimensions.

Since we can quantify the number of points above intensity thresholds Γ, we can observe the incidence of different slicing methods or quantities of initial points (100 %, 50 % and 10 %). We see that the proportion of points remaining after the filtering is more consequent with more slicing; while the quantity of points has almost no impact on the filtered point cloud with differences of ± 0.02 % for the fifty-fold sliced point cloud (Table 2). We can also observe that the initial subsampling has uniformly removed points of the cloud, so that the proportions of points in the ROI are nearly equal for each file (± 0.01 %).

Percentage of initial points	Total points	Percentage of points in the ROI	Number of slices operated	Percentage of remaining points
100 %	80,730,524	89.40 %	50	4.35 %
			9	4.32 %
			1	2.61 %
50 %	40,365,262	89.40 %	50	4.37 %
			9	4.37 %
			1	2.60 %
10 %	8,073,053	89.41 %	50	4.35 %
			9	4.36 %
			1	2.60 %

TABLE 2: REMAINING POINTS ACCORDING TO THE NUMBER OF INITIAL POINTS AND THE NUMBER OF SLICES

Terrain, non-tree vegetation and trees classification

We now export our filtered point cloud as a new *.las* file that we are going to analyze on *CloudCompare*. We want to differentiate terrain and vegetation, using the *CANUPO* plugin. This plugin has been selected for its ease of use for non-specialists, and its rapidity. It reposes on a supervised machine learning taking as an input the complete point cloud, and both sets of manually selected points for which we are confident that they all are trees, or all are ground / non-tree vegetation. The selection is easily done with the "segment" tool of *CloudCompare*. *CANUPO* works by analyzing geometries of each sample to determine the geometrical characteristics of each class of object. It then will iteratively aim to minimize the error – often quantified as mean squared error (MSE) – produced by applying parameters for the training set. The trained model then applies its calibrated parameters to classify the rest of the point cloud.
CANUPO produces quite good results if the training sets have been correctly selected. To ensure that, it is important to well include shrubs (trees that are not longer than 1.3 meters nor thicker than 10 centimeters at DBH)

in the set of non-trees, and to take wooden parts as well as leaves in the set of trees, so that the model will memorize all the range of geometries for each distinct class of object. Despite having greatly selected the training set, it is possible that the shrubs geometry is too close to the trees geometry, especially when the shrubs are quite tall. It is then necessary to put shrubs on the tree training set, allowing us to firstly separate ground and vegetation; and then to retrain and recompute the *CANUPO* plugin to separate remaining vegetation between trees and non-tree vegetation. Here, we used the first method. Figure 7 shows the training sets and the resulting classified point cloud. It is also noticeable that *CANUPO* not only outputs each point's class, but also gives their confidence level – e.g. for flat terrain points it should be near 1, while for low vegetation points it should be closer to 0.5. This gives an easy manual control to, as a last resort, remove problematic points, which should have a low confidence level, although it often subsequently partially removes higher parts of trees. Other manual adjustments, based on considering all points above an arbitrary altitude as being from trees, can be used to improve by eye the trees point cloud collected, by using again the "segment" tool of *CloudCompare*.

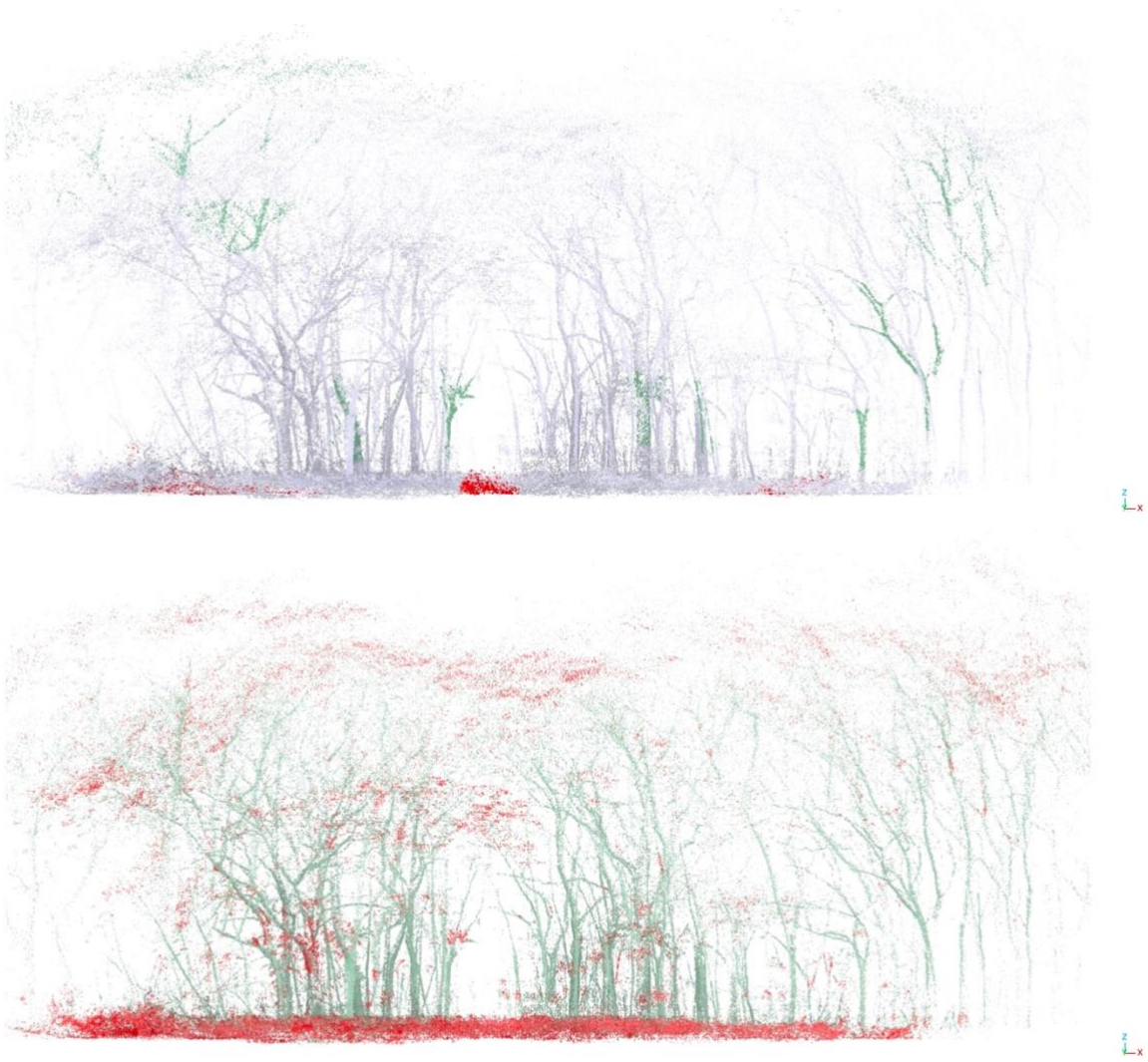


FIGURE 7: TRAINING SETS AND CANUPO CLASSIFIED POINT CLOUD

Red: non-tree (class 1); Green: trees (class 2); Blue: out of the training set (class 0). Apparent errors (red points in trees) have to be manually corrected before continuing. Fisher discriminant ratio of the model (fdr): 4.60907. Images are screenshots taken from CloudCompare.

A well-trained model should have a great "separability", that we defined similarly to Brodu and Lague (2012) by using the Fisher discriminant ratio *fdr* (3) as a quantifier of this separability – based on the mean distances to the other class, μ_1 and μ_2, and their variances ${\sigma_1}^2$ and ${\sigma_2}^2$. A plane of maximal class separability, which is the geometrical representation of the linear support vector machine (SVM) model used by *CANUPO*, helps the user to see if the point cloud has been well classified (Figure 8). It can also be modified by moving or even deforming the line delimiting both classes.

(3) $fdr = \frac{(\mu_2 - \mu_1)^2}{{\sigma_2}^2 - {\sigma_1}^2}$

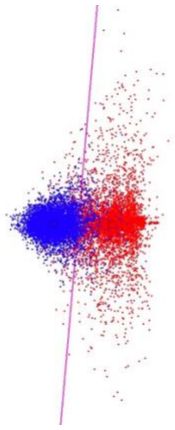


FIGURE 8: PLANE OF MAXIMAL SEPARABILITY GENERATED BY CANUPO
No scale provided. Blue: non-tree (class 1); Red: trees (class 2).

We tried several training sets (three) for each down-sampled or complete file. Results can be very fluctuating, and do not make appear a clear influence of the percentage of points on the efficiency of the model (Table 3).

Percentage of initial points	Average *fdr*	Standard deviation *fdr*	Maximum *fdr*	Average percentage of error, non-tree	Average percentage of error, tree
10 %	4.80	1.15	6.09	6.98 %	6.85 %
50 %	4.32	0.44	4.63	5.22 %	9.78 %
100 %	6.05	1.88	8.14	4.42 %	6.13 %

TABLE 3: CANUPO TRAINED MODELS ATTRIBUTES WITH THREE ATTEMPTS FOR EACH ROW
NB: all files are filtered according to the method mentioned above (2.1), using 50 slices

Segmentation of individual trees

Our ROI is now ready to be segmented. Numerous existing methods, with a similar purpose (Fekry et al., 2022; Malzer et al., 2020), are using a non-supervised machine learning model designed by McInnes et al. (2017), called HDBSCAN, which stands for hierarchical density-based spatial clustering of applications with noise. This method assesses the local density to form clusters according to an inputted minimal number of samples forming a cluster. It is based on DBSCAN, which uses the same principle but without hierarchization, leading to provide an extra parameter, precisely the "maximum distance between two samples for one to be considered as in the neighborhood of the other" (Pedregosa et al., 2011).
Using solely HDBSCAN/DBSCAN presents limits in very dense environments, since it will tend to overlap trees that are often very close at crown height; it needs to be coupled with another segmentation algorithm. To address this pitfall, we took advantage of the specialized segmentation tools available on *3D Forest*, and reinforced the consistency of our trees point clouds by operating co-registration of clusters resulting from both DBSCAN and *3D Forest* segmentation. Using a region growing algorithm, we managed to capture more crucial points to regain the whole structure of each tree, and then we reinforced the trunks as well as crowns and leafy parts, using the nearest neighbors method.

3D Forest tree segmentation

Since we previously separated trees from the rest of the cloud with *CANUPO* classification, we should first export both classes of points separately. Then, we will import these files on *3D Forest* and perform a PCA-based segmentation – PCA stands for principal component analysis.
Before computing tree segmentation, we have the possibility to acquire a mesh representing the ground/terrain, using an inverse distance weighting (IDW) algorithm on *3D Forest*. This can help us to better assess the height of the trees later. If encountering issues when performing this algorithm, it can be because the non-tree cloud contains

too much vegetation or shrubs, since IDW assumes that the cloud is only containing the terrain; it is then necessary to delete manually the highest points until the problem is solved.
We can then perform an automatic tree segmentation, using voxelization – a voxel is a volumetric pixel, i.e. a cube in three-dimensional space. Note that voxelization is a great way to accelerate computations as it reduces the number of elementary units, but is not optimal for visualization. We first define a voxel grid by giving in input a voxel size (default is 4 centimeters), then each voxel will have a value given by the points within the voxel and the selected descriptor. Here the descriptor is derived from PCA, which means that the three spatial coordinates are analyzed to output another base in $\mathbb{R}^3$ where the transformed *x* axis concentrates most of the points information, and the transformed *z* axis provides the least information – PCA has indeed a dimension reduction purpose. Each voxel descriptor contains the coordinates of the new orthogonal base, and the algorithm computes ratios of these coordinates for each axis of neighboring voxels. If they are similar – meaning that the average ratio is greater than a percentage that we arbitrary defined, then the voxels are merged. This makes sense at a voxel scale: regions of points touching each other should follow a certain continuity if they are representing the same object, and the PCA of regions with a "continuous" repartition should lead to similar results, because this sort of point continuity implies a small variation in the spatial repartition of the points in each voxel. Finally, we should define a minimum height (distance from terrain) to discriminate regions that would be too low to be trees, and a minimum number of voxels to form a region. Another parameter is setting the number of iterations for merging regions that have not been merged because of occlusions in the point cloud: to sum up, this other iterating process doubles each time the size of the voxels and recomputes the ratios with its neighbors. Simply note that a high number of iterations is associated with a point cloud containing big occlusions, while a low number of iterations is sufficient for well segmenting points clouds with minor occlusions while preventing any overlapping between regions. More details on the working of the algorithm are given by Trochta et al. (2017).
Then, the segmented trees can be exported. *3D Forest* does not allow to export clouds as *.las* files, although, since we just want the points coordinates and intensity, we can export clouds as a text file.

Clustering of trunks based on density and co-registration

After having reuploaded our ROI, which is now a cloud that should contain only the trees that we want to separate; and converted it to a *n'* by 4 matrix – *n'* is the number of remaining points on the trees point cloud – we are going to take a slice of it at trunk level. While the thickness should stay around 0.3-1.0 meter, the height of the slicing depends on the species of tree: the goal is to find a region where each tree that we want to analyze has only one stem, being actually its trunk. Nevertheless, stems separating at very low-level nay under the ground can be eventually merged later. Here, we selected a slice located between 0.70 and 1.30 meter.
Since this ROI slice is a region where clusters should be far from each other (in absence of tree crowns) and of nearly identical density (cf: 2.1.2), we used DBSCAN; it performs faster than HDBSCAN, and leaves more freedom to the user while tuning the model to find the best parameters (Table 4).
To assess our quality of clustering and well tune the parameters, we can observe by eye on *CloudCompare* the superposed ROI and clusters (easy to visualize by setting different colors for the ROI and the clusters, playing with the opacity as well). A trunk represented by several clusters is not a problem for further computations that will merge them if close enough to each other. Nevertheless, the less clusters are used to represent a single trunk, the most accurate the co-registration processed later will be. Indeed, having multiple clusters can imply, especially for large trunks, to increase our maximum distance between clusters that we want to merge so that these separated clusters can be detected as a single region; this can lead to subsequently merge clusters of noise. An optimal configuration has to be selected, with the help of visual information (Figure 9).

Model	Maximum space between core points *eps*	Minimum number of core points *n_samples*	Number of clusters detected	Percentage of noise
DBSCAN 1	0.1 m	10	201	77 %
DBSCAN 2	0.1 m	20	37	87 %
DBSCAN 3	0.15 m	10	370	60 %
DBSCAN 4	0.15 m	20	71	81 %
DBSCAN 5	0.2 m	10	478	49 %
DBSCAN 6	0.2 m	20	162	70 %

TABLE 4: DBSCAN PARAMETERS AND OUTPUT RESULTS
The computed vegetation cloud is stemming from the 10 % subsampled point cloud, after having separated trees from the rest of the cloud using CANUPO.

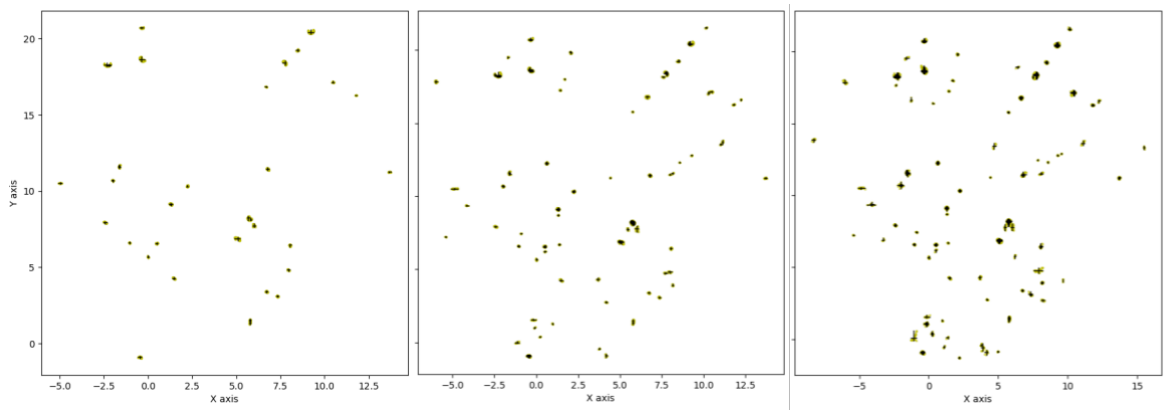


FIGURE 9: DBSCAN-ACQUIRED CLUSTERS WITH DIFFERENT PARAMETERS

Left plot is using DBSCAN 2 model (the most restrictive), middle plot is using DBSCAN 1 model (the most convenient), right plot is using DBSCAN 5 model (the less restrictive).

After having selected the most optimal DBSCAN model – i.e. pair of parameters – we can perform co-registration. It reposits on centroids of each cluster $\mathcal{C}_v$, and centroids of each region's slice in the same *z* axis zone than clusters $\mathcal{R}_{u,restr}$. Note that the results of the co-registration of each model can also give information about its efficiency (Table 5). Centroids are simply defined here as the average coordinates in the (*x*,*y*) plane (4, 5). If clusters are more numerous than regions, the algorithm will iterate for each region and merge clusters that are close enough; if regions are more numerous than clusters, then the algorithm will iterate for each cluster and find the nearest regions using the same condition. The second option can be suitable for homogeneous trees point clouds having almost no occlusions or irregularities, while the first option corresponds to our current cloud. This merging algorithm takes only one variable as an input, being the maximum distance d between clusters/regions to merge them into a single region $\mathcal{M}_p$ (6). Figure 10 allows us to verify if the resulting centroids of the merged clusters/regions at trunk level are coherent.

Model for clustering	Number of clusters	Number of clusters/regions merged	Percentage of regions kept	Average number of clusters in one merged region
DBSCAN 1	201	31	88.9 %	5.58
DBSCAN 2	37	27	77.1 %	1.59
DBSCAN 3	370	31	88.9 %	10.55
DBSCAN 4	71	29	82.9 %	2.28
DBSCAN 5	478	32	91.4 %	13.03
DBSCAN 6	162	29	82.9 %	5.34

TABLE 5: RESULTS OF THE CO-REGISTRATION OF CLUSTERS AND REGIONS

The models are referring to the previous table which described each model's parameters. Maximum merging distance has been set to 0.2 meters.

$$(4)\quad \mathcal{C}_u = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_m\} \in \mathbb{R}^3\ ,\ \mathcal{R}_{v,restr} = \{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_{m'}\} \in \mathbb{R}^3\ /\ \mathcal{R}_{v,restr|z} \in [\min(\mathcal{C}_{u|z}) : \max(\mathcal{C}_{u|z})]$$

$$(5)\quad \vec{\Omega}_{u|x,y} = \frac{1}{n} \times \sum_{k=0}^{n} \overrightarrow{M_{k|x,y}}\ \forall\ \overrightarrow{M_k} \in \mathcal{C}_u\ ,\ \vec{\Omega}_{v|x,y} = \frac{1}{n} \times \sum_{k=0}^{n} \overrightarrow{M_{k|x,y}}\ \forall\ \overrightarrow{M_k} \in \mathcal{R}_{v,restr}$$

$$(6)\quad \mathcal{M}_p = \mathcal{C}_u \cup \mathcal{R}_{v,restr}\ /\ \left\|\vec{\Omega}_{v|x,y} - \vec{\Omega}_{u|x,y}\right\| \leq\ d$$

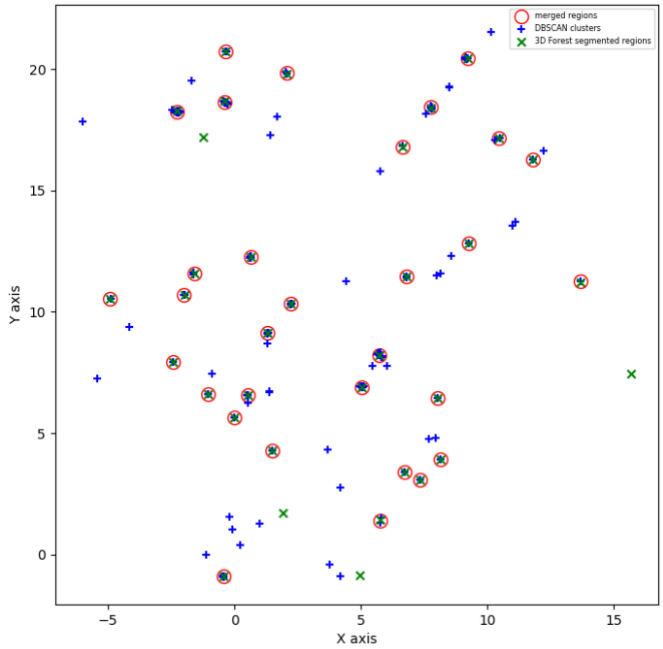


FIGURE 10: CLUSTERS AND REGIONS CO-REGISTRATION RESULTS

We used DBSCAN 1 model for clustering, and maximum merging distance d=0.2 m.

Upper extension of the tree by region growing

Even though the co-registration of clusters and regions looks to have been efficient for eliminating low quality regions, this hypothesis is only true for the lower parts of the trees. Remaining regions can have been cut, especially if they are close to the limits of the ROI. It is also possible that the higher parts of the trees are too deteriorated to be identified as part of the corresponding regions. To manage this issue, we will perform a region growing algorithm that uses each region's highest point (*z* axis) as a starting point, and looks for the nearest point's distance: if the distance is close enough (we have to define a threshold), the point will be added to the region and the next iteration will use this point to compute again the nearest point distance (Figure 11). This algorithm, because it processes point by point (oppositely to voxel by voxel), takes a lot of time to perform (more than an hour for our testing data). It is important to well choose the threshold for adding new points to a region to avoid overlapping, and to take into account that for some regions that are touching each other, overlapping will be difficult to avoid (manual segmentation will have to be processed further). After region growing, regions with qualitative higher parts should have grown while the regions with too few points or occlusions may recover less nay none parts. We will highlight the most qualitative trees later, for now we can keep all of the regions.

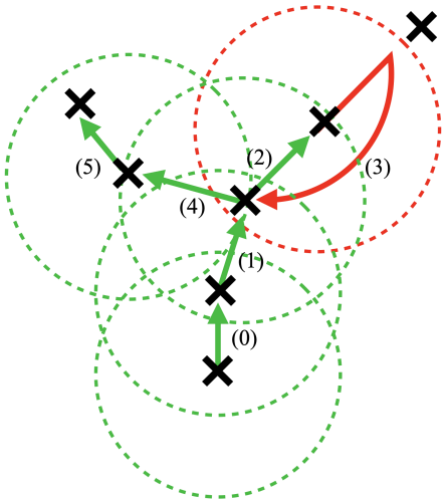


FIGURE 11: WORKING PRINCIPLE OF OUR REGION GROWING ALGORITHM

If the selected point's closest neighbor is within the circle of radius d, the iteration goes upward (0,1,2,4,5). If not, the iteration goes backward (3) and the previous selected point is tested again to check if its new closest neighbor is within the circle of radius d.

Reinforcement of the trees structures by nearest neighbor assignment

To recover some parts of the regions that may have been occluded, so that the further computations will work well, we re-used the initial point cloud without intensity filtering to assign points that are very close to each region (a nice threshold should be 0.05 to 0.10 m).
We also permitted fusion of further points depending on height, by defining a second threshold (0.1 to 0.5 m) and a height from which the threshold to merge free points with the region linearly grows from the first threshold to the second one (this gradient is 1.0 meter high).
Using the original point cloud is possible because we now have – at least we assume that we have – the main structure of each region. This assignment is using unsupervised nearest neighbors learning, that can take a few minutes to perform.

A last step before individual tree analysis is discriminating too small trees, that we will assume to have been wrongly processed regarding to the smallest height recorded on the on-field measurements dataset. The threshold is defined by the user, so that trees leaning angles can be taken into account while performing this step – because the on-field measurements dataset contains lengths, that are different from heights.

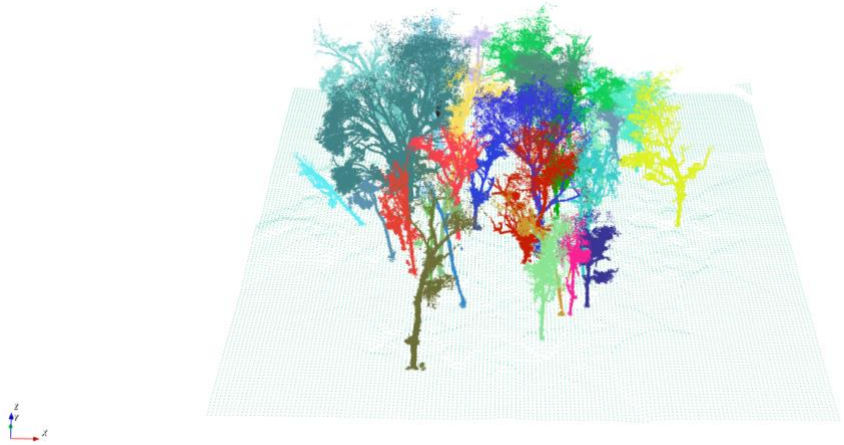


FIGURE 12: INDIVIDUAL TREE CLOUDS ON IDW-ACQUIRED TERRAIN
Screenshot taken from 3D Forest.

Quantitative structure modelling and tree parameters assessment

Point-cloud-based direct estimations

Successfully processed individual tree clouds can be analyzed using *3D Forest*. The software will use the terrain that we provided previously from *CANUPO* classification, and determine for each tree its base position. Visual check is convenient for each tree variable estimation, and if one measurement is not conclusive, it is possible to adjust the parameters of the computations made. Here are some variables that we can assess:

- DBH can be estimated via two methods: randomized Hough transformation (RHT) tries to fit an ellipse by randomly selecting points located between 1.25 and 1.35 meters and outputs the most probable ellipse after a given number of epochs (one epoch is a complete cycle of iterations on a dataset); least square regression (LSR) will aim to fit an ellipse minimizing the error with the actual position of the points located between 1.25 and 1.35 meters (because of that, it is sensible to the presence of outliers).
- Length and height are respectively the distance between the most distant points of each tree and the distance in the *z* axis between the highest point and the base point of the tree; note that the length computed by *3D Forest* is different from the length recorded on the on-field measurements dataset, that is by extension the length of the main stem.
- Convex/concave planar projections are obtained by making an orthogonal projection of the tree in the (*x*,*y*) plane. Combined with the tree height, they can be great variables to assess the access to direct light for each individual tree.

For each individual tree, we will estimate its DBH (using RHT), its height and its convex plane. Then we will perform quantitative structure modelling (QSM) to retrieve new approximations of the DBH as well as a volume estimation that we will use to assess the wooden biomass of the tree.

QSM is performed by randomized Hough transformation, similarly to DBH, except that it iterates for several slices of identical height, producing finally a stack of cylinders representing the tree. With this method, we can detect a branch separation by comparing fittings of one, two, or more cylinders. The height of individual cylinders is chosen by the user, and several parameters such as the minimum length or diameter of a branch can be set. Then, we can estimate the total volume of the QSM reconstructed tree by summing up the volumes of the cylinders, knowing each's diameter. Note that QSM can be affected if the leaves of the tree are detected as branches, creating stems that are not actually existing. It is possible to classify leaves and stems to remove leaves without losing the small stems, for instance with the tool *LeWoS* from Wang et al. (2020). Still, in our case, we assumed that the uncertainty about the result was already bigger than the calibration that can be brought by such methods, so we kept the leaves for each individual tree cloud here. For our focusing subplot, we observed quite satisfying results of the tree reconstruction by QSM, despite some stems tended to be forgotten by QSM (Figure 13).

Although we will not use this method here, we can also acquire the volume via voxelization: each voxel containing points will be counted as part of the tree, resulting in summing the not empty voxels volumes.

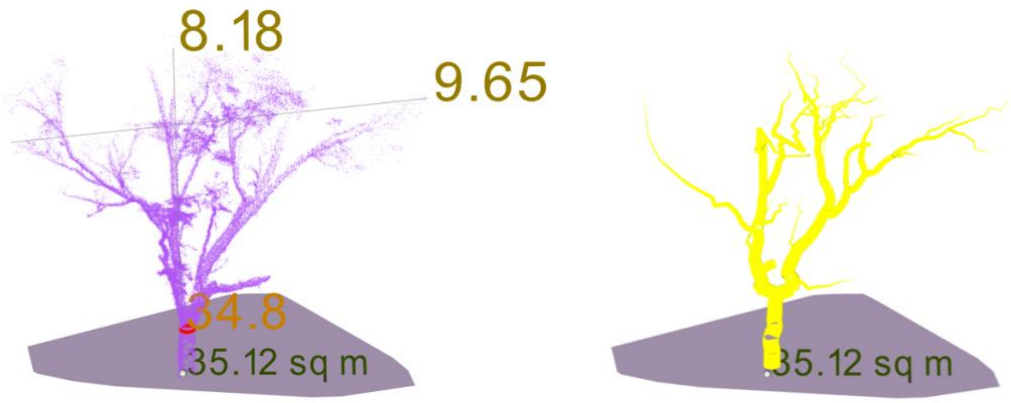


FIGURE 13: EXAMPLE OF A TREE ASSESSED AND RECONSTRUCTED USING QSM
Height: 8.18 meters; length: 9.65 meters (here, length is typically different from the on-field definition); DBH: 34.8 centimeters; convex plane: 35.12 squared meters. Cylinders of 10 centimeters each.

Above-ground biomass (AGB) estimation from previously assessed variables

Having the volume of the QSM reconstructed tree as well as the volume by voxel, we can obtain the woody biomass by assuming that the density of a tree is homogeneous, so that if we know the species of the tree and its associated density (B), we can retrieve its biomass in kilograms. Since leaves are very light elements of the tree, we can here make the simplification that the AGB is corresponding to the woody biomass of the trees that we measured. We also assumed that AGB is represented by the biggest trees by not paying attention to smaller trees. And finally, since we lost several trees during our process because their quality was too low to recover them, we should interpret the outputted AGB with caution. We did so by always giving the number of trees actually measured within the plot.

To prevent uncoherent results, we insured of the following:

- Each computed tree has to have a difference between its RHT-based DBH and its QSM-based DBH of less than 30 %, if not LSR-based DBH is tested instead of RHT-based DBH;
- The volume of the tree has to be in the same order of magnitude than estimations from Brown's equation (explained further) using RHT-based DBH and estimated height;
- The user has to ensure that the estimations are fitting well with the observed point cloud: if QSM creates too thick branches at crown level (often because of a too high leave density, or overlapping of other trees), the tree has to be removed.

Results and discuss

The interpretation of our results is limited because of both large scale of comparison and absence of reference measurements. The average AGB is to contrast with the average DBH and height, as well as the percentage of trees detected, that can give information about the reliability of our analyzed sample. In that perspective, it is clear that our model needs on-field measurements to provide useful information. In addition, the total AGB of a subplot presents the biggest uncertainty because of its cumulative aspect.

Comparison with on-field data

We tested our model on four different samples, consisting of four different subplots; two (T2S8, T3S8) were containing *bischofia javanicae* whereas the two others (T3S2, T3S4) contained *swietenia macrophyllae*. To firstly check if the trees that have been processed are similar to the trees that have been identified on-field, we compared their DBH and height through on-field measurements and cloud-based estimations. We cannot proceed tree by tree, so we observed trends given by all the samples, as well as separated trends for each species. It revealed a certain coherence – especially when focusing on each species – even though the repartitions of cloud-based estimations were more dispersed: their correlation coefficients were around 32-36 % while measured values had 71-85 % of correlation (Figure 14). Hence, while DBH has been on average under-estimated for *bischofia javanicae* and over-estimated for *swietenia macrophyllae*, the average estimated height was always lower than the measured height, probably because the tallest trees have not been fully recovered by our model – especially for *swietenia macrophyllae* (Table 6).

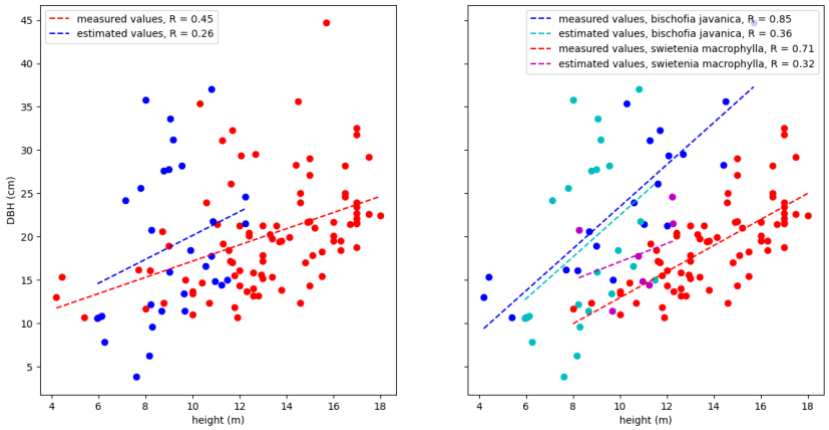


FIGURE 14: DBH AND HEIGHT ESTIMATIONS OF OUR MODEL VERSUS ON-FIELD MEASUREMENTS

Subplot	Sp.	Number of eligible trees	Actual number of trees	Average estimated DBH (cm)	Average measured DBH (cm)	Average estimated height (m)	Average measured height (m)
T2S8	*b.j.*	15	13	17.3	20.7	8.6	9.1
T3S2	*s.m.*	2	29	22.7	18.9	10.2	13.6
T3S4	*s.m.*	5	41	16.0	19.1	11.0	14.3
T3S8	*b.j.*	8	8	23.1	28.2	8.9	11.7

TABLE 6: COMPARISON OF AVERAGE ESTIMATIONS AND MEASUREMENTS OF EACH SUBPLOT
B.j.: bischofia javanica; s.m.: swietenia macrophylla. Eligible trees are QSM-processed trees giving fair results, defined as DBH and tree reconstruction visually corresponding to expectations, less than 30% of difference between RHT DBH and QSM DBH, plus a ratio between Brown's equation result and AGB estimation of more than 0.1 (one scale of magnitude). Note that the measured height is actually obtained from the living length and the leaning angle of each tree.

Finally, it seems like the model detected more easily *bischofia javanicae*, while having a lot of failures for *swietenia macrophyllae* especially when it comes to performing QSM: most of the trees had too poor density at DBH level, leading to a high difference between RHT/LSR-based DBH and QSM-based DBH or even inability to

estimate DBH or to reconstruct the tree using QSM. We also noticed that the reinforcement based on nearest neighbor assignment was sometimes responsible for overlapping on other trees or adding up too many leaves: in these cases, it could have been interesting to use instead the tree clouds obtained after performing region growing.

Comparison with another non-destructive method

We did not destroy any subplot to compare our estimated AGB with a reference measurement of the trees composing it. This implies that we cannot reposit on a direct comparison to assess the quality of our estimation. Hence, we regarded another method proposed in literature to assess the AGB, so that we can figure out indirectly the effectiveness of our model, by knowing the characteristics and performances of the compared model.
We compared our results with estimations made by Brown et al. (1989) using a regression equation needing to know the DBH in centimeters, the height h in meters, and the density of the tree ρ in grams per centimeters cubed (7).

$$(7) \quad AGB_{Brown} = \exp\left(-2.4090 + 0.9522 \times \ln\left(DBH^2 \times h \times \rho\right)\right)$$

We only used the trees that have been successfully processed by our model, so that we compare methods without taking into account the inability of our model to take into account all of the trees. We used the height and the DBH estimated from the tree cloud since we cannot assign on-field measurements to the processed individual tree clouds. We observed mostly higher estimations from our model compared to Brown's equation estimations, that tended to become higher for great AGB – i.e. great DBH and/or height (Table 7, Figure 15).

Subplot	Number of trees detected	Average estimated DBH (cm)	Average estimated height (m)	Average estimated AGB (kg)	Average estimated $\boldsymbol{AGB_{Brown}}$ (kg)
T2S8	15	17.3	8.6	385.9	133.3
T3S2	2	22.7	10.2	298.6	165.3
T3S4	5	16.0	11.0	85.4	93.1
T3S8	8	23.1	8.9	202.3	181.5

TABLE 7: ESTIMATIONS OF OUR MODEL AND COMPARISON WITH S. BROWN'S REGRESSION EQUATION

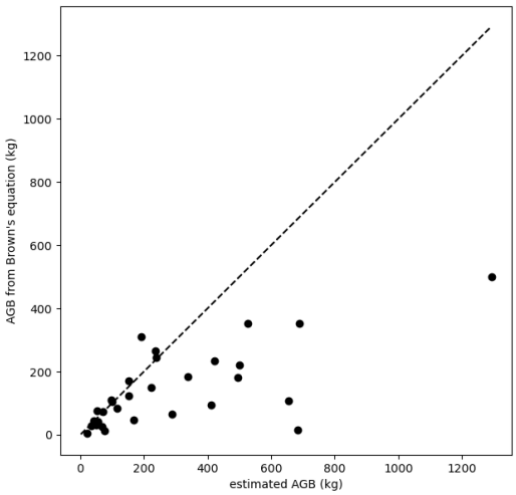


FIGURE 15: COMPARISON OF INDIVIDUAL TREES AGB FROM OUR MODEL AND S. BROWN'S REGRESSION EQUATION

Conclusion

Our model has been built to estimate above ground biomass (AGB) within a recently planted forest of eastern Taiwan, using a terrestrial laser scanner (TLS). It first aims to filter the LiDAR-acquired data by deleting less important points of the cloud (*.las* file), and differentiating measurable trees from the rest of the point cloud. Then, it uses two segmentation models to acquire regions corresponding to individual trees, and performs a complementary segmentation algorithm that will make the regions grow in size, while finally adding up points of the initial point cloud that are close enough to the grown regions to reinforce each tree's structure. Finally, quantitative structure modelling (QSM), as well as other estimations based on geometric calculus, are performed on each individual tree point cloud, leading to know its diameter at breast height (DBH), its height, and its volume. By knowing the species of trees present in each scanned area and its associated density, we can retrieve its AGB.

We tested the model on several areas of the forest plot, leading to observe various results. Subplots containing quite small trees (T2S8) were more easily analyzed by the model, leading to provide trees looking complete and without overlapping of over trees; while subplots with taller trees (T3S2) provided not fully grown regions and tended to merge parts of surrounding tree crowns after nearest neighbor reinforcement – these parts are probably responsible for getting the highest parts hidden from the scanner – leading to provide poorly exploitable QSM data.

Our model presents several limitations, resulting in a high incertitude of the estimated AGB and restrictive rules for its use. We can sum up them as the following:

- Due to the influence of the distance of the scanner while acquiring the intensity of a point, we had to rescale our cloud to a region of interest (ROI) which cuts some trees that will not be processed well.
- The DBSCAN clustering will not work well for a subplot with a high elevational gradient, since it works by taking a thin horizontal slice of the point cloud; a solution would be to take a thicker slice to perform the clustering, despite the results should be less reliable thenceforth.
- Our model uses a region growing algorithm that iterates point by point, costing a noticeable amount of time while it would have been possible – even though more complex – to proceed using voxels (volumetric pixels).
- The reinforcement of the tree clouds using nearest neighbors assignment can lead to overlapping other trees at crown level or to take into account great amounts of leaves that can affect QSM performance.
- Our model uses tools that consider trees to grow straight, while their actual length can be different than their height/altitude, leading to more inaccuracy in the estimation of DBH for leaning trees.
- While users of the model have the possibility to make manual modifications between each step of the process, these corrections have been almost everytime necessary, representing an extra task for the user.
- The results are assessed at a subplot scale, which can reveal interesting trends but remains unprecise when it comes to assess the efficiency of our model depending on individual tree's characteristics; however this can be solved by adding to the model a co-registration of scanned trees and labelled trees if we have the precise coordinates of each individual tree, as well as an accurately calibrated geographic extent of our scanned area.
- The thinner the trees of the scanned area are, the less well our model performs, meaning that in young forests where the AGB is mostly contained in shrubs, our model might produce more inaccurate results; in other words, our model assumes that the AGB is contained in the thicker trees.
- We observed a noticeable difference between maximum estimated height and maximum measured height, meaning that our model has not been able to fully recover bigger trees, mostly because their distance to the TLS was already too big.
- We cannot compare our estimations of AGB with neutral values, that would require destructive means which are not suitable to our case (the forest plot is not big enough to destroy a part of it in that single purpose).

This estimation of AGB can help to better understand the dynamics leading a forest. For instance, it offers a complementary approach to mortality censuses by assessing the evolution of the AGB in a subplot, ultimately giving a relation between damage to living trees and biomass loss within the forest plot, similarly to Zuleta et al. (2023).
It can also orient local land management policies, by estimating the impact of diverse uses of the forest. In our case, the private company *Xinguang* who owns the studied area since the 70's, plans to implement new photoelectric facilities – initially 437 of the total 718 hectares. 66 hectares have been already probated by the local district's agricultural comity, with regard to existing elements about the chosen area's agricultural productivity

(Events in Focus, 2021). Our model can track the evolution of the AGB in diverse subplots to unscramble gradients in biomass production within the plot, moreover it can help to observe if the existing photoelectric implantation has an impact on the surrounding forest. Combined to the local socio-economic situation described higher, it contributes to enlightened decision-making. Such interdisciplinary application requires in the other hand adequate decisionmakers representing all the actors of this area to be effective: the company looking for profit, the government concerned about transitioning to renewable energy, the local inhabitants potentially impacted by the change of land use (employment, landscape…), as well as the natural entities that are more and more considered as active actors since they provide limited services and resources that need a certain pace to regenerate and become sustainable (Latour, 2018).

Global framework of the internship

This internship allowed me to explore in detail the specific topic of tropical ecology. Even though most of my work consisted in programming and testing a model on a computer, I also had many opportunities to explore tropical forests and assist the conduct of censuses. Going to the actual forest plot that we studied here was of course very helpful for the development of my model. I was able to observe the occupation of space by diverse vegetation, which gave me clues about the main challenges to separate trees from other vegetation in computations, the imponderable errors (e.g. small lianas strangling the trees that are almost impossible to remove from individual tree clouds), the on-field conditions for scanning well the trees, as well as the time and resources needed to acquire the LiDAR data.
I also participated in fifth global census of Fushan experimental forest, that took place in a mountainous old-grown subtropical rainforest of northern Taiwan. During my work for this initiative conducted by Taiwan Forestry Bureau, the Smithsonian institute and several universities of Taiwan, I was able to see the techniques, required resources and limitations in following individual tree's evolution (growth, survival, leaning or rooting to the ground, new branches, etc.). This also inspired me to make my model compatible with more diverse forests, even though I have not been able to acquire data there. Some presentations from researchers in diverse sub-topics in tropical ecology during this census have also been a source of inspiration for my report.

In addition of skills in tropical ecology, I learned a lot about technical skills that were closely related to my work, such as TLS, segmentation, or machine learning. This autonomous learning was maybe the main challenge of this internship in my point of view, because it required me to select methods that were affordable to my current knowledge and time, and then to find ways to check by myself if I fairly used the concepts that I selected.

During my internship and more generally during my stay in Taiwan, I have been able to communicate mostly in Chinese, and thereby I finish this internship with the feeling to have notably enhanced my fluency in speaking as well as my understanding of spoken or written – traditional – Chinese. Most people have been very benevolent to me so that I wasn't fearing to make mistakes while speaking out or writing out messages. Especially, English fluent people were comprehensive about my will of using Chinese instead of English as soon as it was possible, and helped me a lot in acquiring new vocabulary. I used English mostly to explain my work's big lines in a first time, and for discussion demanding very technical vocabulary that I obviously had not yet when I arrived. Henceforth, I feel much more comfortable with using Chinese every day and already look forward to other opportunities to stay in Taiwan in the future.

As a first internship of the kind, it was important to me to smartly manage my available time, by considering feasible tasks to make my work doable within about 300 hours. Finding more and more related papers about TLS in the literature during my research, I realized that the topic was endless and experienced very frequent updates – since I began my research, new releases of *CloudCompare* or *scikit-learn* came out offering new tools for tree segmentation and HDBSCAN. If I had more time available, for sure, I would have proceeded differently. Thus, I am satisfied of my organization in terms of schedules, that let me enjoy my stay without having to worry about unfeasible deadlines.

I have been able to finish the work that I was initially assigned. I then had to explain the working of the model I created to the other assistants of the laboratory, so that they can use it as a new tool to analyze the forest plot. In consequence I managed to provide a convenient support that can be automatically generated for each use, summing up all the important results (C), in addition of a small user manual (D). At the end, this model aims to be frequently used to acquire new AGB data within the framework of a project conducted by Pr. Sun to assess the impact of diverse land uses on nearby forests.

BIBLIOGRAPHY

Allen RB., MacKenzie DI., Bellingham PJ., et al. (2020). "Tree survival and growth responses in the aftermath of a strong earthquake", 108, 107-121. https://doi.org/10.1111/1365-2745.13238

Amatulli, G., McInerney, D., Sethi, T., Strobl, P., Domisch, S. (2020). *Geomorpho90m - Global High-Resolution Geomorphometry Layers*. Distributed by OpenTopography. https://doi.org/10.5069/G91R6NPX

Beck H., Zimmermann N., McVicar T. et al. (2018). "Present and future Köppen-Geiger climate classification maps at 1-km resolution", *Sci Data*, 5, 180214. https://doi.org/10.1038/sdata.2018.214

Bliss C-I. (1934) "The method of probit", *Science*, 79 (2037), 38-39.

Brodu N., Lague D. (2012). "3D Terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology", *ISPRS Journal of Photogrammetry and Remote Sensing*, 68, 121-134, ISSN 0924-2716, https://doi.org/10.1016/j.isprsjprs.2012.01.006

Brown S., Gillespie A., Lugo A. (1989). "Biomass Estimation Methods for Tropical Forests with Applications to Forest Inventory Data", *Forest Science*, 35, 881-902.

Cai W., Santoso A., Collins M. et al. (2021). "Changing El Niño–Southern Oscillation in a warming climate", *Nat Rev Earth Environ*, 2, 628-644. https://doi.org/10.1038/s43017-021-00199-z

Central Geological Survey, Ministry of Economic Affairs, RoC. (2021). *"Probabilities Map" for Active Faults in Taiwan*. https://www.moeacgs.gov.tw/eng/News/news_more?id=85a463a206de4f06a9cd3c46ca9d9b79

Chen Y-Y., Huang W., Wang W-H., Juang J-Y., Hong J-S., Kato T., Luyssaert S. (2019). "Reconstructing Taiwan's land cover changes between 1904 and 2015 from historical maps and satellite images", *Scientific reports*, 9 (1), 3643. https://doi.org/10.1038/s41598-019-40063-1

Events in Focus (2021). *花蓮鳳林兆豐農場　欲闢 500 公頃光電場　內政部先通過 66 公頃* [press article]. https://eventsinfocus.org/news/7146104

Fekry R., Yao W., Cao L., Shen X. (2022). "Ground-based/UAV-LiDAR data fusion for quantitative structure modeling and tree parameter retrieval in subtropical planted forest", *Forest Ecosystems*, 9, ISSN 2197-5620, https://doi.org/10.1016/j.fecs.2022.100065

Forestry Bureau of the Council of Agriculture (2021). *Forestry Statistics, Yearbook 2020*. https://www.forest.gov.tw/EN/0004346

Fu L-Y., Zeng W-S., Tang S-Z. (2017). "Individual Tree Biomass Models to Estimate Forest Biomass for Large Spatial Regions Developed Using Four Pine Species in China", *Forest Science*, 63 (3), 241-249. https://doi.org/10.5849/FS-2016-055

Gonzalez de Tanago J., Lau A., Bartholomeus, H. et al. (2018). "Estimation of above-ground biomass of large tropical trees with terrestrial LiDAR", *Methods Ecol Evol*, 9, 223-234. https://doi.org/10.1111/2041-210X.12904

Gutiérrez, J.M., R.G. Jones, G.T. Narisma, L.M. Alves, M. Amjad, I.V. Gorodetskaya, M. Grose, N.A.B. Klutse, S. Krakovska, J. Li, D. Martínez-Castro, L.O. Mearns, S.H. Mernild, T. Ngo-Duc, B. van den Hurk, and J.-H. Yoon. (2021). "Atlas". In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L.Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K.Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press. http://interactive-atlas.ipcc.ch/

Hogan J-A., Zimmerman J-K., Thompson J., Uriarte M., Swenson N-G., Condit R., Hubbell S., Johnson D-J., Sun I-F., Chang-Yang C-H., et al. (2018). "The Frequency of Cyclonic Wind Storms Shapes Tropical Forest Dynamism and Functional Trait Dispersion", *Forests*, 9 (7), 404. https://doi.org/10.3390/f9070404

Iturbide M., Fernández J., Gutiérrez J.M., Bedia J., Cimadevilla E., Díez-Sierra J., Manzanas R., Casanueva A., Baño-Medina J., Milovac J., Herrera S., Cofiño A.S., San Martín D., García-Díez M., Hauser M., Huard D., Yelekci Ö. (2021). *Repository supporting the implementation of FAIR principles in the IPCC-WG1 Atlas*. Zenodo. https://github.com/IPCC-WG1/Atlas

Kellner, J.R., Armston, J., Birrer, M. et al. (2019). "New Opportunities for Forest Remote Sensing Through Ultra-High-Density Drone Lidar", *Surv Geophys*, 40, 959–977. https://doi.org/10.1007/s10712-019-09529-9

Krůček M., Trochta J., Cibulka M., Král K. (2019). "Beyond the cones: How crown shape plasticity alters aboveground competition for space and light—Evidence from terrestrial laser scanning", *Agricultural and Forest Meteorology*, 264, 188-199, ISSN 0168-1923. https://doi.org/10.1016/j.agrformet.2018.09.016

Krůček M., Král K., Cushman K-C., Missarov A., Kellner J-R., (2020). "Supervised Segmentation of Ultra-High-Density Drone Lidar for Large-Area Mapping of Individual Trees", *Remote Sensing*, 12 (3260), 2-16. https://doi.org/10.3390/rs12193260

Latour, B. (2018). "Esquisse d'un Parlement des choses", *Écologie & politique*, 56, 47-64. https://doi.org/10.3917/ecopo1.056.0047

Liang X. et al. (2018). "International benchmarking of terrestrial laser scanning approaches for forest inventories", *ISPRS Journal of Photogrammetry and Remote Sensing*, 144, 137-179, ISSN 0924-2716. https://doi.org/10.1016/j.isprsjprs.2018.06.021

Lin C-W., Chen W-S. (2016). *Geologic Map of Taiwan.* https://www.researchgate.net/publication/329415855_Geologic_Map_of_Taiwan

Lin C-W., Liu Y-C., Chou P-S., Lin Y-H. (2022). *ACTIVE FAULT MAP OF TAIWAN, 2021.* https://www.researchgate.net/publication/362326131_ACTIVE_FAULT_MAP_OF_TAIWAN_2021

Malzer C., Baum M. (2020). "A Hybrid Approach To Hierarchical Density-based Cluster Selection", in: IEEE, *International Conference on Multisensor Fusion and Integration for Intelligent Systems*. https://doi.org/10.1109%2Fmfi49285.2020.9235263

McInnes L., Healy J., Astels S. (2017). "hdbscan: Hierarchical density-based clustering", *Journal of Open Source Software*, 2 (11), 205, https://doi.org/10.21105/joss.00205

National Statistics of the Directorate-General of Budget, Accounting and Statistics (2020). *2020 Statistical tables.* https://eng.stat.gov.tw/News.aspx?n=2401&sms=10889

National Statistics of the Directorate-General of Budget, Accounting and Statistics (2023). *Manpower Survey Results in May 2023.* https://eng.stat.gov.tw/News_Content.aspx?n=2369&s=231585

Paris C., Kelbe D., Van Aardt J., Bruzzone L. (2017). "A novel automatic method for the fusion of ALS and TLS LiDAR data for robust assessment of tree crown structure", *IEEE Trans. Geosci. Rem. Sens.* 55, 3679-3693. https://doi.org/10.1109/TGRS.2017.2675963

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. (2011). "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, 12 (85), 2825-2830.

Saglam A., Bilgehan Makineci H., Akhan Baykan N., Kaan Baykan O. (2020). "Boundary constrained voxel segmentation for 3D point clouds using local geometric differences", *Expert Systems with Applications*, 157, ISSN 0957-4174. https://doi.org/10.1016/j.eswa.2020.113439

Somogyi Z., Cienciala E., Mäkipää R., Muukkonen P., Lehtonen A., Weiss P. (2007). "Indirect methods of large-scale forest biomass estimation"**,** *Eur. J. For. Res*, 126 (2), 197-207. https://doi.org/10.1007/s10342-006-0125-7

Thompson I., Mackey B., Mcnulty S., Mosseler A. (2009). *Forest Resilience, Biodiversity, and Climate Change. A Synthesis of the Biodiversity/Resilience/Stability Relationship in Forest Ecosystems*. https://www.researchgate.net/publication/232129022_Forest_Resilience_Biodiversity_and_Climate_Change_A_Synthesis_of_the_BiodiversityResilienceStability_Relationship_in_Forest_Ecosystems

Townshend, J. (2016). *Global Forest Cover Change (GFCC) Tree Cover Multi-Year Global 30 m V003* [Dataset]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MEaSUREs/GFCC/GFCC30TC.003

Trochta J., Krucek M., Vrska T., Kral K. (2017). "3D Forest: An application for descriptions of three-dimensional forest structures using terrestrial LiDAR", *PLoS ONE*, 12. https://doi.org/10.1371/journal.pone.0176871

Wang, D, Momo Takoudjou, S, Casella, E. LeWoS (2020). "A universal leaf-wood classification method to facilitate the 3D modelling of large tropical trees using terrestrial LiDAR", *Methods Ecol Evol*, 11, 376– 389. https://doi.org/10.1111/2041-210X.13342

Zuleta D., Arellano G., McMahon S-M., Aguilar S., Bunyavejchewin S., Castaño N., Chang-Yang C-H., Duque A., Mitre D., Nasardin M., Pérez R., Sun I.-F., Yao T. L., Valencia R., Krishna Moorthy S-M., Verbeeck H., Davies S-J. (2023). "Damage to living trees contributes to almost half of the biomass losses in tropical forests", *Global Change Biology*, 29, 3409-3420. https://doi.org/10.1111/gcb.16687

DECEMBER 2022 MORTALITY CENSUS DATASET

The initial data was giving tree by tree information, here we compiled information by subplot. Subplots used in this report are highlighted.

Subplot	Number of trees	Species	Average DBH (cm)	Average living length (m)	Maximum leaning (º)
T1S10	3	*fraxinus griffithii, zelkova serrata*	19.8	9.8	0
T1S2	2	*fraxinus griffithii, alstonia scholaris*	25.5	14.8	0
T1S3	1	*terminalia catappa*	19.3	12.7	0
T1S4	4	*alstonia scholaris*	26.9	13.0	0
T1S5	1	*terminalia catappa*	20.0	12.2	30
T1S6	13	*fraxinus griffithii, alstonia scholaris*	17.9	13.3	15
T1S7	6	*fraxinus griffithii, alstonia scholaris, terminalia catappa*	24.4	13.9	0
T1S8	9	*alstonia scholaris, terminalia catappa*	29.2	13.4	30
T1S9	2	*fraxinus griffithii*	13.4	9.8	0
T2S1	17	*bischofia javanica*	20.4	10.4	0
T2S10	11	*fraxinus griffithii, bischofia javanica*	16.1	11.3	0
T2S2	20	*bischofia javanica*	19.0	10.0	0
T2S3	22	*bischofia javanica*	21.6	10.7	0
T2S4	19	*bischofia javanica*	20.8	10.5	0
T2S5	25	*bischofia javanica*	18.9	9.6	0
T2S6	21	*bischofia javanica*	21.5	9.9	0
T2S7	20	*bischofia javanica*	19.3	9.7	0
T2S8	**13**	***bischofia javanica***	**20.7**	**9.1**	**0**
T2S9	5	*fraxinus griffithii, bischofia javanica, zelkova serrata*	17.9	11.6	15
T3S1	30	*swietenia macrophylla*	20.3	13.6	0
T3S10	6	*bischofia javanica*	23.3	9.7	0
T3S2	**29**	***swietenia macrophylla***	**18.9**	**13.7**	**15**
T3S3	26	*swietenia macrophylla*	19.0	14.3	0
T3S4	**42**	***swietenia macrophylla***	**19.0**	**14.0**	**0**
T3S5	41	*swietenia macrophylla, bischofia javanica*	18.6	13.3	75
T3S6	8	*bischofia javanica*	21.2	8.7	0
T3S7	19	*bischofia javanica*	25.9	11.4	0
T3S8	**8**	***bischofia javanica***	**28.2**	**11.7**	**0**
T3S9	7	*bischofia javanica*	26.5	9.2	0
T4S1	16	*pterocarpus indicus, swietenia macrophylla, terminalia catappa*	25.0	14.1	30
T4S10	20	*swietenia macrophylla*	24.8	13.1	90
T4S2	5	*terminalia catappa*	18.9	11.5	15
T4S3	11	*terminalia catappa*	22.3	12.3	15
T4S4	12	*pterocarpus indicus, swietenia macrophylla, terminalia catappa*	19.2	12.0	15
T4S5	2	*terminalia catappa*	16.2	12.0	0
T4S6	7	*terminalia catappa*	23.5	11.2	45
T4S7	4	*terminalia catappa*	20.9	11.4	45
T4S8	5	*terminalia catappa*	20.6	11.2	0
T4S9	11	*swietenia macrophylla, bischofia javanica, terminalia catappa*	20.4	11.2	0

PLOT'S SPECIES ASSOCIATED DENSITIES

Species	Number of trees within the plot	Density (g/cm^3 or T/m^3)[1]
bischofia javanica	210	0.58
swietenia macrophylla	191	0.49
terminalia catappa	74	0.52
fraxinus griffithii	29	N/A
alstonia scholaris	10	0.36
pterocarpus indicus	6	0.52
zelkova serrata	3	N/A

[1] Brown, S. (1997). "Estimating Biomass and Biomass Change of Tropical Forests: A Primer", *FAO Forestry Paper*, 134. https://www.fao.org/3/W4095E/w4095e0c.htm

EXAMPLE OF GENERATED REPORT

AGB report: 2023-8-7

***AGB**: above-ground biomass, counting only trees longer than 1.30 meter and having a DBH (diameter at breast height) thicker than 10 centimeters. Check for the precautions of use in the associated report[2] and GitHub repository[3].*

***QSM**: quantitative structure modelling, tree reconstruction processing by performing for vertical slices randomized Hough transformation (ellipse fitting) to rebuild the tree as a stack of cylinders.*

Location of the starting point: 121.495364ºE, 23.791372ºN
Subplot(s): T2S8
Area of the analyzed region of interest: 30.0 by 30.0 meters (900.0 m^2)
Species of tree: *bischofia javanica*

Number of trees within the subplot(s): 13
Number of trees detected by the model after co-registration: 30 (231.0 %)
Number of trees ready for QSM after removal of too small trees: 27 (208.0 %)
Number of trees having successfully passed QSM: 15 (115.0 %)

(by tree)	**Estimated DBH (cm)**	**Measured DBH (cm)**	**Estimated total height (m)**	**Measured living height (m)**
Average	17.3	20.7	8.6	9.1
Standard deviation	10.6	8.4	1.7	2.8
Maximum	37.0	35.6	11.5	14.5
Minimum	3.8	10.7	6.0	4.2

TABLE 1: MEASUREMENTS ACCURACY (DBH, HEIGHT)
Measured living height is approximated from measured living length using its leaning angle. Note that we compare the estimated total height with the actual living height (≠ actual total height).

(by tree)	**Estimated AGB (kg)**	**AGB from Brown's equation[4] (kg)**	**Estimated AGB, surfaced (kg/m^2)**
TOTAL	5788.5	1999.9	6.4
Average	385.9	133.3	/
Standard deviation	340.2	147.9	/
Maximum	1292.8	500.1	/
Minimum	20.1	4.7	/

TABLE 2: AGB ESTIMATIONS

[1] Chardon T. (2023). *Semi-automatic estimation of the tree biomass using terrestrial LiDAR data* [internship report]
[2] *Ibid.*
[3] https://github.com/Thibalt-C/tree-biomass-estimator/
[4] Brown S., Gillespie A., Lugo A. (1989). "Biomass Estimation Methods for Tropical Forests with Applications to Forest Inventory Data", *Forest Science*, 35, 881-902.

FIGURE 1: LOCATION OF THE SCANNED AREA

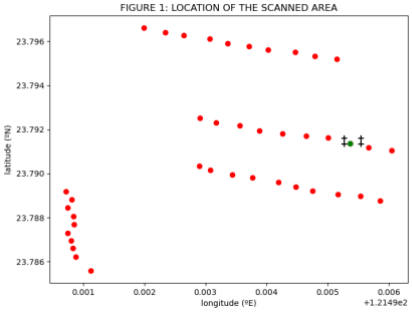


FIGURE 2.1: ESTIMATED AGB ACCORDING TO TREE DBH

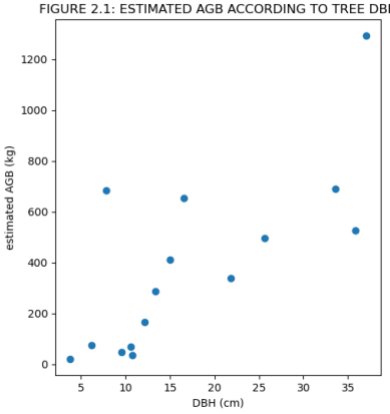


FIGURE 2.2: ESTIMATED AGB ACCORDING TO TREE HEIGHT

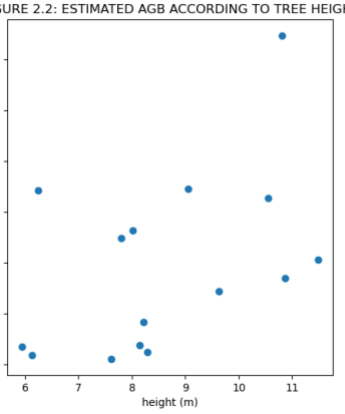


FIGURE 3.1: SPATIAL REPARTITION, PLANAR VIEW

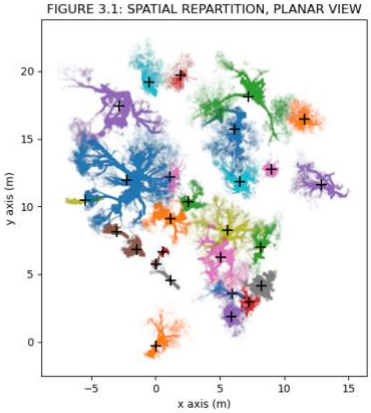


FIGURE 3.2: SPATIAL REPARTITION, SIDE VIEW

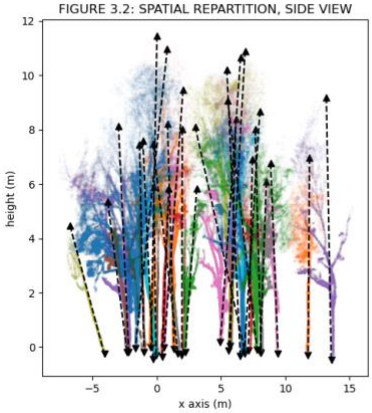


FIGURE 4: DBH AND HEIGHT ESTIMATIONS OF OUR MODEL VERSUS ON-FIELD MEASUREMENTS

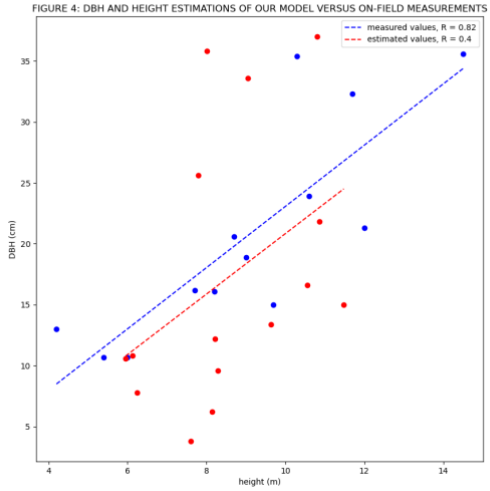
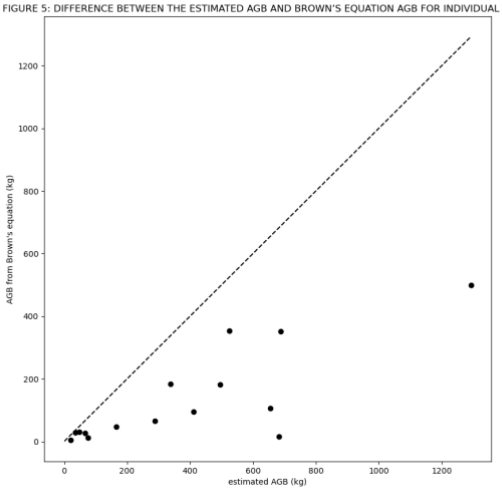


FIGURE 5: DIFFERENCE BETWEEN THE ESTIMATED AGB AND BROWN'S EQUATION AGB FOR INDIVIDUAL TREES



CHEAT SHEET FOR MODEL'S EVERYDAY USE

利用陸地激光雷達數據半自動估計樹木生物質

第一步 ~3分

3D點雲收藏·按強度的過濾

打開GeoSlamHub，連接激光雷達，將數據以.las文件格式導入。同時，需進行下採樣（保留原文件的約10%）。

執行'函數們.ipynb'檔案（請使用Jupyter Lab，可在Anaconda軟體中找到）。你也可以在終端中執行以下命令："python {文件路徑}/函數們.py"

現在可以使用以下函數，請記得調整輸入參數：
"強度過濾(文件,錢片,RDI)"
（注意，RDI是一個列表：【xmin,xmax,ymin,ymax】，也可以使用"auto"）

在GITHUB可以找到更多的信息（英文）

第二步 ~20分

跟CANUPO的數據過濾

開啟CloudCompare，導入經過過濾的文件。
下"1"鍵，使用"segment"工具。選擇一些區域，其中它含樹木和非樹木。使用"split cloud"功能，將點雲分成三個部分（class 0，class 1，class 2）。
選擇"CANUPO：train classifier"，並為class 1和class 2設置一個新模型（請務必勾選"Use original cloud for descriptors"）。
設定足夠的"fdr"值，建議設置為4或更高，然後執行"CANUPO：classify"。
再次使用"segment"工具和"split cloud"功能，將樹木和非樹木部分分開。

第三步 ~1小時

DBSCAN聚類和3DForest分割的共同註冊，地區生長算法

請注意，在每一步驟中都要仔細檢查潛在問題，可以使用CloudCompare進行手動編輯。

使用函數"DBSCAN聚類(trees_file,RDIzmin,RDIzmax,eps,min_samples)"，以找到樹幹的聚類。

打開3DForest，導入含有樹木的文件，使用"Automatic Tree Segmentation"功能將樹木分開，最後導出每個新的文件。

使用函數"共同註冊(dist_to_merge,RDIzmin,RDIzmax)"進行地區融合。

使用函數"地區生長算法(trees_file,dist_to_add_upper,dist_to_add_lower,RDIzmin,RDIzmax)"擴展樹木的結構。這個過程可能需要較長時間。

使用函數"用最近的鄰居方法的加強(original_file, general_threshold, crown_threshold, height_for_crown_threshold,min_height)"以加強最終的樹木分割。

第四步 ~2小時

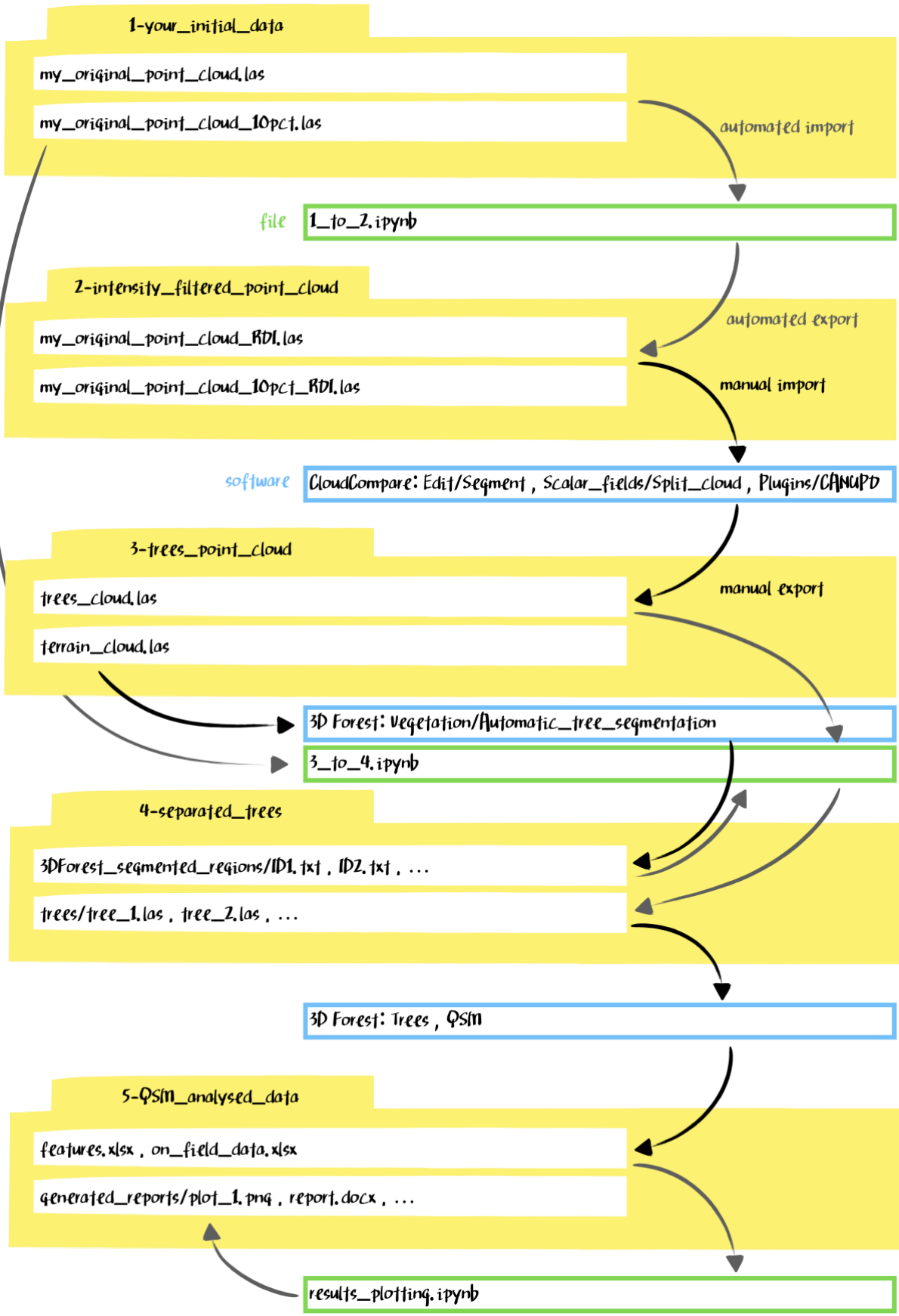
QSM·最終數據

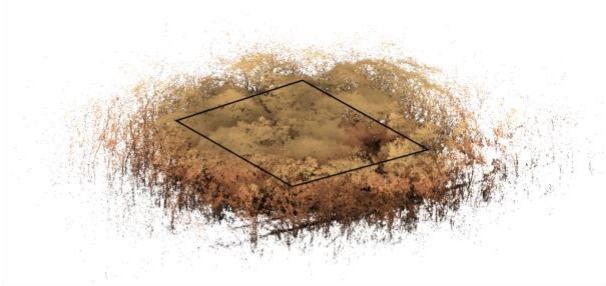
打開3DForest，導入每個樹木的文件，然後依次使用以下功能：
- Tree base position by lowest point
- DBH by randomized Hough transformation (RHT)
- Tree height
- Tree reconstruction (QSM)
- volume by QSM

數據將顯示在一個數據表中。你可以將每個數字複製並粘貼到"features.xlsx"文件中。如果有紅色的儲存格：
- DBH：使用"DBH by least square regression (LSR)"
- AGB：再嘗試其它"volume by QSM"的參數

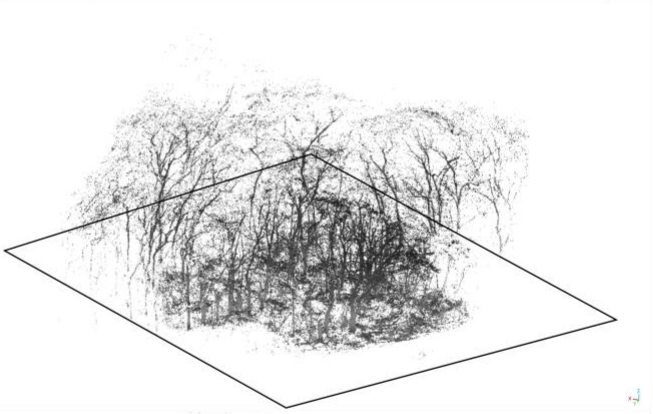
如果還有紅色的儲存格，必須消除它們（全線）。也可以消除少於10cm的表線。

最後，使用"結果繪圖(start_location)"為了顯示結果。

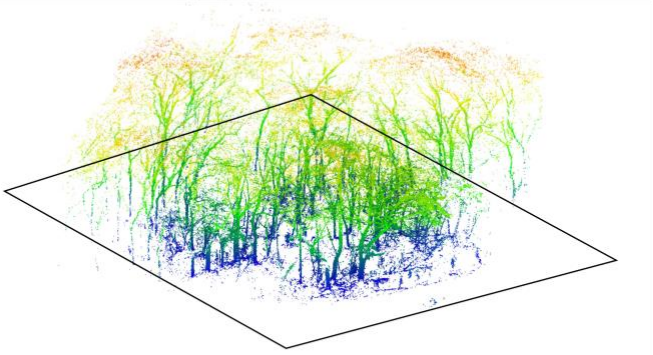

1-your_initial_data
my_original_point_cloud.las
my_original_point_cloud_10pct.las
automated import
file
1_to_2.ipynb
2-intensity_filtered_point_cloud
automated export
my_original_point_cloud_RDI.las
my_original_point_cloud_10pct_RDI.las
manual import
software
CloudCompare: Edit/Segment , Scalar_fields/Split_cloud , Plugins/CANUPO
3-trees_point_cloud
manual export
trees_cloud.las
terrain_cloud.las
3D Forest: Vegetation/Automatic_tree_segmentation
3_to_4.ipynb
4-separated_trees
3DForest_segmented_regions/ID1.txt , ID2.txt , ...
trees/tree_1.las , tree_2.las , ...
3D Forest: Trees , QSM
5-QSM_analysed_data
features.xlsx , on_field_data.xlsx
generated_reports/plot_1.png , report.docx , ...
results_plotting.ipynb



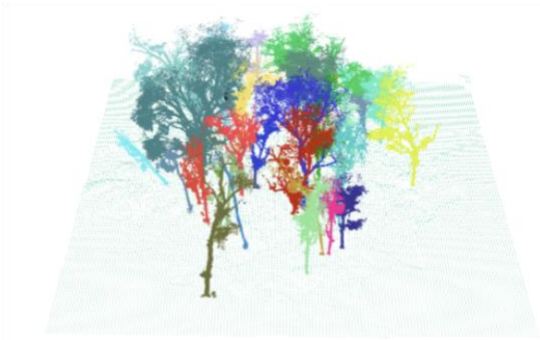
1-your_initial_data



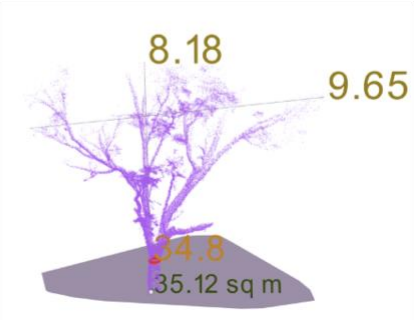
2-intensity_filtered_point_cloud



3-trees_point_cloud



4-separated_trees



5-QSM_analysed_data