Enhanced Forest Inventories: A QGIS Plugin to incorporate R processing tools in forest management

Larissa Maria Granja [1], Francesco Pirotti [1,2]

[1] Department of Land, Environment, Agriculture and Forestry (TESAF), University of Padova, Legnaro, Italy – larissamaria.cavalcantefalcaogranja@phd.unipd.it
[2] Interdepartmental Research Center of Geomatics (CIRGEO), University of Padova, Legnaro, Italy - francesco.pirotti@unipd.it

Keywords: Smart forestry, Airborne laser scanner, QGIS plugin, R, lidR package.

Abstract

The complexity of remote sensing systems makes it possible to collect huge amounts of data that public administrations often do not use at their full potential. The traditional forest inventories (samples and field campaigns) used to identify tree species and measure morphological and physiological parameters are financially burdensome and time-demanding. Thus, remote sensing can be an alternative used by public administration to reduce the efforts in the field and improve the quality of forest inventories and land cover mapping at a sustainable price. In this scenario, this work aims to bridge the gap between common inventory practices and enhanced forest inventories (EFI), that used lidar point clouds and GIS environments together. To support EFIs we developed and present here a QGIS plugin for accessing and processing 3D point clouds to enable decision-makers in the forestry sector to have easier and more intuitive data processing pipelines. Lidar data provides accurate and detailed information on the vertical structure of canopies and can provide estimated volume and biomass, as well as other parameters that are key in forest management. This plugin differs from other approaches in that it initiates a stream with an active R session and sends commands from QGIS with several solutions that re-use all intermediate steps avoiding recalculating them. This saves time and allows multiple processing threads to run in parallel, and thus test different combinations of input parameters to the workflow. Examples and results of the processing are given over a specific study area.

1. Introduction

Forests provide several benefits to society. They capture and store carbon, improve air and water quality, mitigate erosion, protect structures against landslides, house biodiversity, and are important for economic development (Trumbore et al., 2015). Thus, it is essential to perform forest inventories and provide information to forest managers, policy makers and practitioners about the current state of forests and the changes happening.

In the last century, remote sensing (RS) systems and their applications have advanced. The complexity of RS systems makes it possible to collect huge amounts of data that public administrations often do not use at their full potential. The traditional methods (field campaigns and chemical analysis) used to identify tree species and measure morphological and physiological parameters are financially and time-demanding. Therefore, remote sensing can be an alternative used by the public administration to reduce the efforts in the field and improve the quality of forest inventory databases and land cover mapping at a sustainable price. Several public authorities in almost all European countries have regular acquisition of remote sensing data using both passive and active sensors (Nex et al., 2015).

Nowadays, it is happening an effort to integrate Lidar point clouds to improve forest inventories because it offers a new perspective on evaluating forest structures. Light Detection Ranging (LIDAR) is an active sensor used to describe the vertical structure of forests. LIDAR describes the distance to the targets based on the return reflection. This technology emits laser pulses that have small footprint and measures the 3D distribution of vegetation in forest canopies. It is well stablished in the literature the robustness and repeatability of Lidar point clouds to improve the estimation of forest parameters in forest inventories (Bauwens et al., 2016: Fassnacht et al., 2024; Kershaw et al., 2016; White et al., 2016).

In this context, the goal of this work is to bring the results of higher-level procedures for processing point clouds in the R CRAN to the more commonly used GIS environment in a user-friendly context. To achieve this, we present a QGIS plugin for accessing and processing 3D point clouds to enable decision-makers in the forestry sector to have easier and more intuitive access to these data. Although there are other plugins available to process Lidar, their functioning is often dependent on old R libraries and old versions of software's like GRASS GIS that with time are being replaced. We try to fill the gap between researchers - who use coding and advanced analysis of point clouds - and practitioners/public administration in the forestry sector who are comfortable using GIS tools and raster/vector-based data (and are not used to working with 3D point clouds).

Public administration and practitioners would greatly benefit from using directly or indirectly 3D point clouds. Lidar 3D data notably provides accurate and detailed information on the vertical structure of canopies and can validly estimate volume and biomass, as well as other parameters that are key in forest management. Point cloud data are not usually processed by forester managers and forest practitioners.

The R CRAN provides several packages that approach point cloud processing. The lidR package is, at the time of writing this note, one of the more commonly used R packages to process Lidar data. LidR is open source and implements several algorithms to advanced processing and visualization of airborne laser scanner data with emphasis on forestry applications. The users can calculate any metrics at individual, pixel, tree or stand/scale levels defined by them without conforming with predefined metrics like in other packages/software's (Rousell et al., 2020; Pirotti et al., 2017).

https://doi.org/10.5194/isprs-archives-XLVIII-3-2024-203-2024

2. Materials and Study area

The QGIS framework allows plugin developers to extend the functionality of QGIS to several applications. The rationale is that many tools are available for forestry applications and Lidar data processing in the R CRAN environment. Thus, it would be useful to bridge the QGIS interface to the R CRAN tools. We use Python to develop a plugin that initializes an R session through the "subprocess" functionality and allows QGIS users to call specific R routines using the QGIS interface.

We therefore present initial results of pipelines that are useful for forestry practitioners. Specifically, we show how to use the plugin for setting up a project environment, for extracting information from the Lidar data, and then for semi-automatically extracting tree positions and heights. This is the precursory required information for estimating the volume at the parcel level, which is the key information required by practitioners and public administrators to optimize the management of forests.

The study area is in the Autonomous Province of Trento (Figure 1). A mountainous region in the northern part of Italy. Only 13% of the territory is located below 600 m. Trento forests cover 63% of its territory and the species composition in terms of surface area are spruce (32%), beech (14%), larch (13%), and silver fir (11%) (Department of Civil Protection, forest, and fauna of Trento).

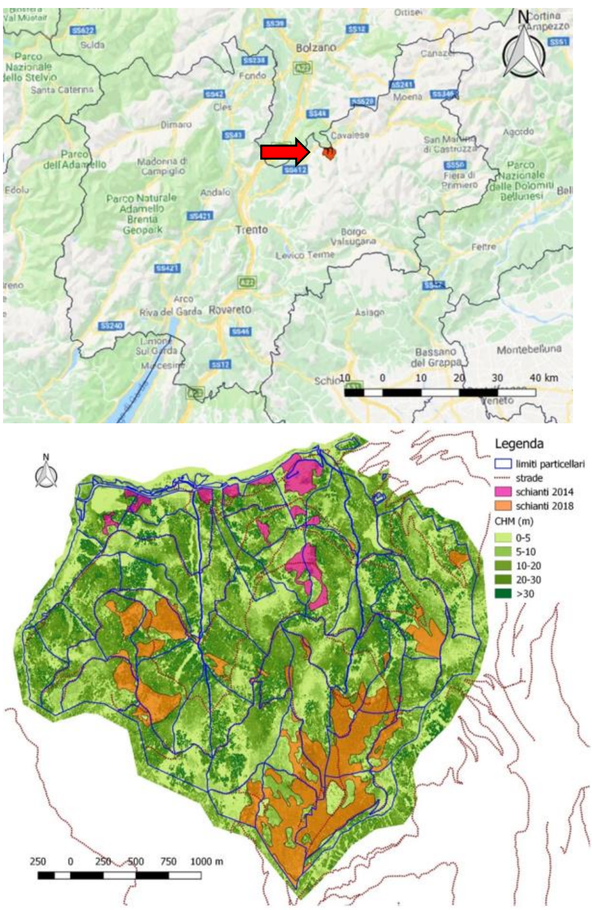

Figure 1. Study area in the Province of Trento (Paterno, 2020).

3. Methods

There are plenty of routines in the lidR R package (Roussel et al., 2020) that provide state of the art processing for aerial lidar data in forest environments, the novelty of the present study is to bridge the QGIS python environment with the R environment with a smoother approach than existing methods. The pipeline for processing lidar data is usually to take the point cloud files, usually in LAS or LAZ format, and to store them in a single folder, that we will call from now on the working directory (WD). We assume that the files have been pre-processed to remove outliers and are classified for ground points vs non-ground points. These two pre-processing steps can also be integrated in the proposed pipeline, but this will not be done as the main scope of this work is to assess forestry applications. Another reason is that it is reasonable to expect that public administrations or practitioners will make use of professional lidar aerial surveying services, which will provide a high-quality product to the buyers.

3.1 Sub-process heuristics

The following steps bridging R LidR package with QGIS are addressed in this work:

1. create a file catalogue with all LAS/LAZ files in a user-specified folder
2. extract the digital terrain model (DTM)
3. normalize the LAS files using a tin model of ground points from the DTM
4. extract the canopy height model (CHM)
5. extract tree positions and heights with a user-defined function to determine the variable window size.
6. calculate tree canopy area and tree volumes with allometric models inserted by the user in the interface (Figure 2)
7. extract forestry statistics in parcels - total volume, tree height/diameter/volume distribution statistics (mean, median, standard deviation, quantiles), plots.

To avoid re-doing any procedure that was already carried out, intermediate steps are either saved in an R data file with the typical "rda" extension, which is read at the beginning of the procedure, or in spatial files (geopackage for vector data and GeoTIFF for raster data). All procedures output objects have one of two states: either they have been processed/created, or they have to be. To account for different parameters (e.g. the resolution of the output DTM and CHM raster), for each parameter combination a folder will be created with all intermediate products. The folder name will encode the parameters (e.g. outputResRaster1_0MaxHeigh50_0). This allows the user to access the output data intuitively directly in the folder. A log file is also created in the folder in html format. Logging will include the creation of a pdf with plots and maps that derive from the processing and that practitioners can use in reports or other deliverables.

3.2 File Catalogues and parallelization

LAS or LAZ files in the working directory can be stored as catalogues in R using the lidR pipeline. This allows to easily parallelize the workflow, as chunks can be processed in different threads, thus decreasing the processing time. LidR takes care of the necessary buffering between tiles to avoid border effects and duplicated trees at the border.

https://doi.org/10.5194/isprs-archives-XLVIII-3-2024-203-2024

3.3 Digital height models

The CHM is basically a normalized DSM with ground points equal to zero and the rest of the vegetation with height above the ground. If the point cloud is classified according to the LAS specifications (The American Society for Photogrammetry & Remote Sensing, 2019), then only vegetation classes are used (class 3 to 5, respectively low to high vegetation). This is an important advantage as it remove buildings and human artifacts that might then be incorrectly mixed with trees. After normalization of the LAS files, the “pitfree” algorithm from Khosravipour et al. (2016) is used. This method considers all the returns from Lidar to generate a high-resolution DSM without great irregularities (spikes) that can compromise accuracy by increasing the number of undetected trees or wrongly detection of trees. The DSM generated with the contribution of all returns gives a better representation of the canopy structure. The advantage of using this approach is an improved treetop detection including small trees.

3.4 Tree position and heights

To detect the trees and extract the heights with the R package lidR we first used the functions locate_trees and the local maximum filter (lmf). The locate_trees function processes each square region of the acquisition with a buffer that ensures a correct identification of the trees including the ones in the edges of the files. The local maximum filter is point cloud based and locates the tree tops without using the raster. For a chosen point the algorithm will analyse the other points near to determine if the point being processed is the highest. We used the lmf with a variable window size because it adapts to forests that are not homogeneous in terms of tree size. The window size is a function of height. The higher trees have larger crowns thus to detect their treetops it is necessary to use bigger windows. The next step was to perform tree crown delineation using the function segment_trees. When applying this function, it is possible to choose which algorithm, raster-based or CHM based, is more suitable for the dataset being processed. This step is necessary if the user wants to calculate single trees related metrics using the function crown_metrics.

3.5 Parcel Tree Volumes and Statistics

At the end of the process, we will have trees with heights, volume and diameter. The latter two from allometric models that are input from the user. At this stage only volume is deployed, but future developments will see the user adding a lookup table (LUT) where for each tree height volumes and diameters are provided for different scenarios. Scenarios are related to ground fertility that affects the expected tree volume at a certain tree height. As a rule of thumb, taken a specific diameter, fertile soils will have taller slender trees than less fertile soils. This results in a tree, with a specific measured height from the lidar data, having very different volume depending on the scenario.

4. Results

4.1 User interface

One of the objectives is to have the user input all the necessary information in a single panel. At time of writing the panel is defined as seen in Figure 2.

4.1.1 User inputs. As the objective is to have a simple tool for practitioners, the inputs must be limited to only the necessary information. First input is the folder path with all the LAS files. This path will be the root path of the working directory (WD) where a new folder will be created with all intermediate products which are necessary for the processing.

The second input is the resolution of the CHM that will be created (see methods section). This depends on the point cloud density and influences the detection of tree locations.

The third input is a polygon vector layer that was loaded in the QGIS project (or a file) that represents the parcels for which to extract the data. This input is trivially quite important as it will aggregate results such as the total tree volume in the parcel and the volume per hectare and other forest parameters.

The fourth and last input is the necessary equation to estimate the volume having the tree height. Lidar data will be processed to extract the location of each tree and the height of the tree; thus, we must have the allometric models to estimate the volume for each tree first, and then for each parcel.

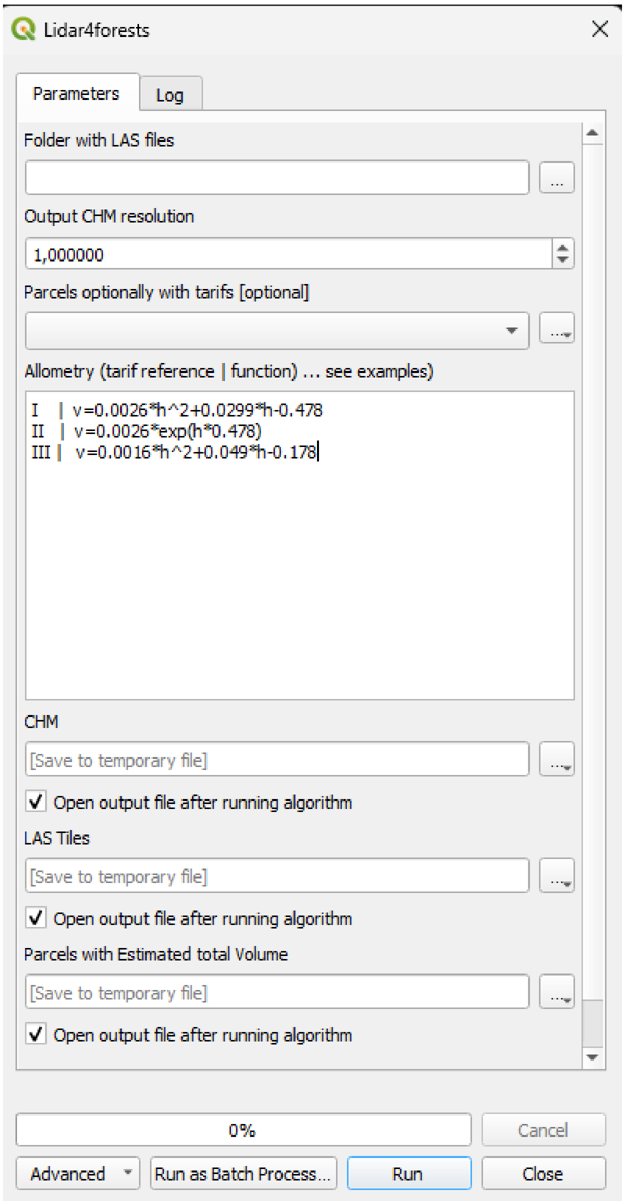

Figure 2. Panel interface for inserting parameters.

4.2 Outputs

Intermediate and final outputs are saved in a subfolder of the working directory. Intermediate files of interest for the practitioner are the following: (i) the tiles covered in each LAS lidar tile - figure 3, (ii) digital height models, both the DTM and CHM - figure 4 (iii) tree positions with tree heights - figure 5 and (iv) final per-parcel statistics - figure 6.

4.2.1 Tiles. The catalogue file is used in LidR to organize multiple LAS files in a way as to process them separately. This helps with memory allocation and can potentially lead to parallel processing in a single or multiple machines. The proposed procedure opens the catalogue produced in R from the LidR processing in QGIS (Figure 3).

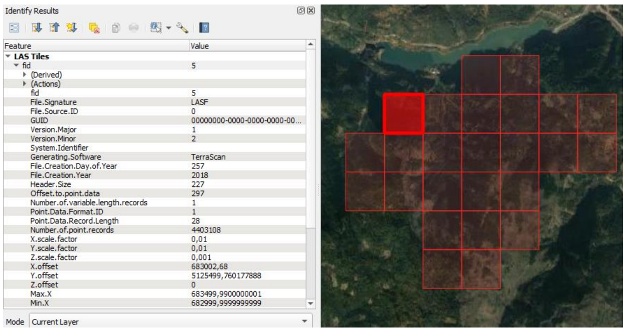

Figure 3. Visual representation of the LAS cataloque from LidR in QGIS.

4.2.2 Canopy and terrain height models. Canopy and terrain height models are intermediate products used to determine tree positions and tree heights, as well as other vertical metrics, as well as canopy size, which is known to be a valid covariate for predicting volume and biomass. Figure 4 shows the automated product at 1 m ground sampling distance.

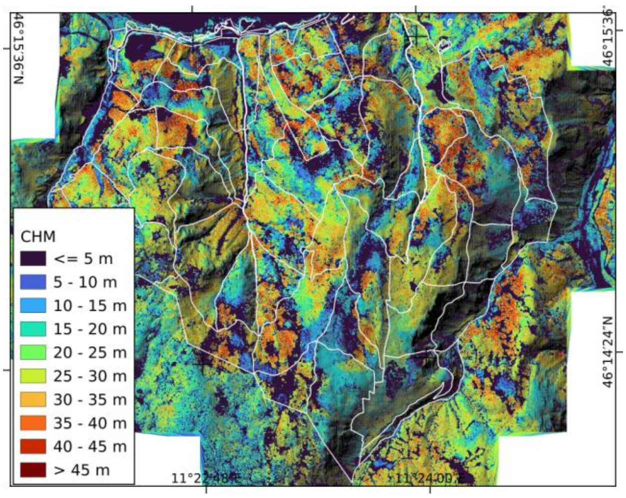

Figure 4. Canopy height model created automatically for all tiles.

4.3 Tree position and heights

Tree-based forest volume estimation is based on the detection of trees and their heights. This is done from Lidar data by local maxima calculation. Allometric models are then used to estimate volume from the tree heights. Results are then aggregated at parcel level to get statistics that can be used to support decision making and forest management. Figure 5 shows a detail of tree height frequency distribution on a specific area, and the overall tree height map. This can be used also for cross-validating results using field campaigns to check if the extracted tree height corresponds to the values measured from the lidar point cloud.

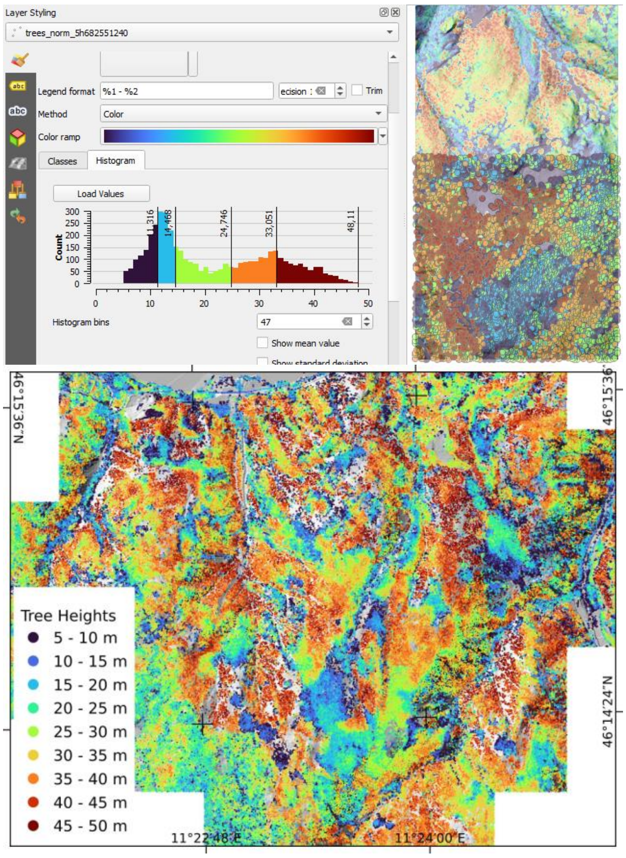

Figure 5. Top - tree height distribution of frequencies; bottom - total tree height map.

4.4 Parcel volumes

Automated processing allows also to apply allometric models (Figure 2) to tree heights and calculate total volume inside parcels applying a simple geometric predictor of intersection.

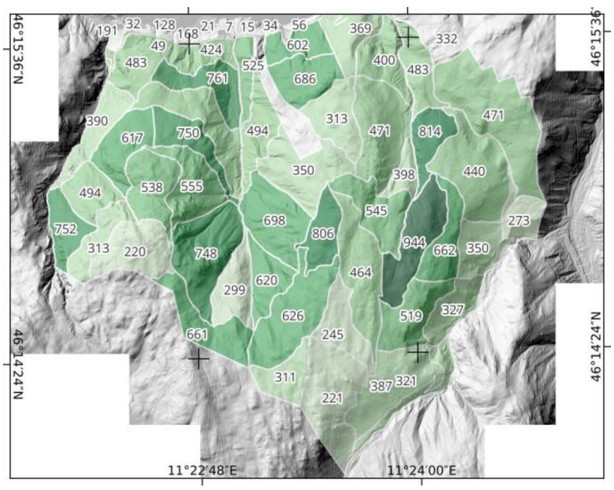

Figure 6. Final volume values (m^3/ha) for each parcel.

This aggregation is useful to practitioners that aim at knowing the volume or biomass in a specific area, for management purposes. Allometric models are species-specific and also location-specific. In this study area we applied the models for conifers reported in the figure 7 below.

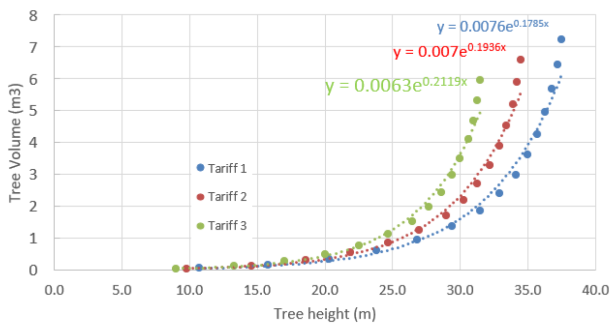

Figure 7. Example of allometric models for three different tariffs for conifers in the region.

5. Discussion

Users knowledgeable in this topic might ask why the *R provider* QGIS plugin was not used instead of this approach. The answer is that the *R processing provider* plugin is great for scripts, but calls a new R process every time it is used, loading a set of R libraries every time. Another drawback is that, at the time of writing, scripts in the R processing provider have some limitations such as not being able to save the users' input, and not to load an output layer with a specific style. The idea behind this work is to bridge the gap between "super-users", that can easily script their own processing pipeline, and "end-users" that are comfortable using a GIS interface only.

The proposed plugins launch an R subprocess and communicates directly with it through *stdin* and *stdout* pipes. This makes the procedure smoother and allows to calculate several runs with different parameters to compare which are more likely to be correct in a specific scenario. Lidar data are nowadays becoming a key source of information in forestry, and rigorous methods for extracting parameters of interest for forestry.

The CHM resolution is an input from the user which must be also discussed. It is a parameter which depends on the point cloud density available in the original LAS files. Future implementations will provide a preferred value of CHM resolution by checking LAS point density. It is trivial to say that if the LAS data have 1 point per square meter, a 2 m CHM is not ideal. If we take into consideration the Nyquist–Shannon sampling theorem, the CHM resolution could be at least half the point density (in the case of 1 point per square meter the CHM could sample up to 0.5 m resolution). It should be noted that this "oversampling" should use at least bilinear interpolation, not nearest neighbour, otherwise gaps in the canopy would result in "pits" in the CHM. This aspect can be partly ignored because the "pitfree" algorithm from Khosravipour et al. (2016).

The estimation of forest parameters using only Lidar data without field samples for validation is well described in the literature (Ferraz et al. 2016; Puliti et al., 2020). However, there are still some challenges. Regarding volume estimation with Lidar, traditionally, it is done combining the measurements of tree height and DBH in allometric equations. In this study, only the tree heights were available. Airborne Laser Scanners (ALS) are capable of measuring accurate heights but do not perform well measuring DBH in high density forests. An alternative to overcome these limitations is combining ALS with terrestrial laser scanners (TLS) (Ali et al., 2020; Cabo et al., 2018, Holopainen et al., 2013).

Another challenge is the sensitivity of the model at high values of tree heights. As can be seen in figure 7, tall trees can have very different volumes. Due to the nature of the equation, this results in a large error if tree height is overestimated. Taking figure 7 as an example, it is trivial to see that a tree that is 50 m tall will have a volume that is grossly overestimated. If this overestimation is done for many trees, this will propagate to the final aggregated statistics per parcel, thus giving wrong values and leading to wrong decisions. For this reason, integration with surveys that measure diameter can be of high importance.

The use of a dataset acquired with a handheld laser scanner like the X120Go SLAM from Stonex that uses the Simultaneous Localization and Mapping (SLAM) algorithm can be an option to acquire precise DBH measurements and consequently improve the volume estimations using the QGIS plugin proposed in this study. The SLAM algorithm estimates the laser scanner position while building a high precision 2D or 3D point cloud map of the surroundings without a GPS (Taheri and Xia, 2021). The three integrated cameras are capable of acquire texture information, producing point clouds with colours and panoramic images.

6. Conclusion

In this work we reported on a new tool for QGIS to process remotely sensed data from aerial lidar sensors. The rationale is that aerial lidar in forestry is becoming a very common data source and is now available often to practitioners and public administrations. These stakeholders are becoming increasingly familiar with QGIS software, due to its open nature and free access. Less familiarity is usually the case regarding Lidar data. Thus, a practical and easy-to-use plugin for processing forestry-related lidar data is likely an important added value. In particular, a user-friendly interface that provides the full pipeline from LAS files to parcels with estimated tree volume, tree locations and various forest metrics and statistics is the final aim of the work. Future developments will see an improved integration with the QGIS interface and a better and more intuitive way to provide allometric models to the workflow.

Acknowledgements

This work was supported by Agritech National Research Center and received funding from the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Componente 2, Investimento 1.4 – D.D. 1032 17/06/2022, CN00000022; Missione 4 Componente 1 Investimento 3.4 and Missione 4 Componente 1 Investimento 4.1 - D.M 351 09/04/2022). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

Ali, I., Durmush, A., Suominen, O., Yli-Hietanen, J., Peltonen, S., Collin, J., & Gotchev, A., 2020. FinnForest dataset: A forest landscape for visual SLAM. *Robotics and Autonomous Systems*, *132*. https://doi.org/10.1016/j.robot.2020.103610.

Bauwens, S., Bartholomeus, H., Calders, K., & Lejeune, P., 2016. Forest inventory with terrestrial LiDAR: A comparison of static and hand-held mobile laser scanning. *Forests*, *7*(6), 127.

Cabo, C., Ordóñez, C., López-Sánchez, C. A., & Armesto, J., 2018. Automatic dendrometry: Tree detection, tree height and diameter estimation using terrestrial laser scanning. *International journal of applied earth observation and geoinformation*, *69*, 164-174. https://doi.org/10.1016/j.jag.2018.01.011.

Department of Civil Protection, forest, and fauna of Trento, 2024.https://forestefauna.provincia.tn.it/Foreste/Foreste-in-Trentino/Le-foreste-in-Trentino.

Fassnacht, F. E., White, J. C., Wulder, M. A., & Næsset, E., 2024. Remote sensing in forestry: current challenges, considerations and directions. *Forestry: An International Journal of Forest Research*, *97*(1), 11-37. https://doi.org/10.1093/forestry/cpad024

GRASS Development Team, 2017. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation. grass.osgeo.org (20 June 2024).

Kershaw Jr, J. A., Ducey, M. J., Beers, T. W., & Husch, B., 2016. *Forest mensuration*. John Wiley & Sons.

Khosravipour, A., Skidmore, A.K., Isenburg, M., 2016. Generating spike-free digital surface models using LiDAR raw point clouds: A new approach for forestry applications. *International Journal of Applied Earth Observation and Geoinformation* 52, 104–114. https://doi.org/10.1016/j.jag.2016.06.005

Nex, F., Delucchi, L., Gianelle, D., Neteler, M., Remondino, F., & Dalponte, M., 2015. Land cover classification and monitoring: the STEM open source solution. *European Journal of Remote Sensing*, 48(1), 811-831. https://doi.org/10.5721/EuJRS20154845.

Pirotti, F., Kobal, M., and Roussel, J. R.: A comparison of tree segmentation methods using very high density Airborne Laser Scanner data, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W7, 285–290. https://doi.org/10.5194/isprs-archives-XLII-2-W7-285-2017, 2017.

Pirotti, F., Paterno, C., and Pividori, M.: Application of tree detection methods over LiDAR data for forest volume estimation, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B3-2020, 1055–1060. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-1055-2020, 2020.

Roussel, J.-R., Auty, D., Coops, N.C., Tompalski, P., Goodbody, T.R.H., Meador, A.S., Bourdon, J.-F., de Boissieu, F., Achim, A., 2020. lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sensing of Environment* 251, 112061. https://doi.org/10.1016/j.rse.2020.112061.

Paterno C. 2020. Utilizzo di dati lidar per la redazione di un piano di gestione forestale aziendale in Provincia di Trento. (Master Thesis - University of Padova).

Taheri, H., & Xia, Z. C., 2021. SLAM; definition and evolution. *Engineering Applications of Artificial Intelligence*, *97*, 104032. https://doi.org/10.1016/j.engappai.2020.104032.

The American Society for Photogrammetry & Remote Sensing, 2019. LAS Specification 1.4 - R14.

Trumbore, S., Brando, P., & Hartmann, H., 2015. Forest health and global change. *Science*, 349(6250), 814-818. https://doi.org/10.1126/science.aac6759.

White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., & Tompalski, P., 2016. Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing*, 42(5), 619-641. https://doi.org/10.1080/07038992.2016.1207484.