Above ground biomass estimation from UAV high resolution RGB images and LiDAR data in a pine forest in Southern Italy

Mauro Maesano (1), Giovanni Santopuoli (2), Federico Valerio Moresi (1), Giorgio Matteucci (3), Bruno Lasserre (4), Giuseppe Scarascia Mugnozza (1)

Knowledge of forest biomass is an essential parameter for managing the forest in a sustainable way, as forest biomass data availability and reliability are necessary for forestry and forest planning, but also for the carbon market as well as to support the local economy in the mountain and inner areas. However, the accurate quantification of the above-ground biomass (AGB) is still a challenge both at the local and global levels. The use of remote sensing techniques with Unmanned Aerial Vehicle (UAV) platforms can be an excellent trade-off between resolution, scale, and frequency data of AGB estimation. In this study, we evaluated the combined use of RGB images from UAV, LiDAR data and ground truth data to estimate AGB in a forested watershed in Southern Italy. A low-cost AGB estimation method was adopted using a commercial fixed-wing drone equipped with an RGB camera, combined with the canopy information derived by LiDAR and validated by field data. Two modelling methods (stepwise regression, SR and random forest, RF) were used to estimate forest AGB. The output was an accurate maps of AGB for each model. The RF model showed better accuracy than the Steplm model, and the R² increased from 0.81 to 0.86, and the RMSE and MAE values were decreased from 45.5 to 31.7 Mg ha⁻¹ and from 34.2 to 22.1 Mg ha⁻¹ respectively. We demonstrated that by increasing the computing efficiency through a machine learning algorithm, readily available images can be used to obtain satisfactory results, as proven by the accuracy of the Random forest above biomass estimation model.

Keywords: Above Ground Biomass, UAV, Random Forest, Forest Biomass, Machine Learning

Introduction

Forest biomass is one of the most important indicators of Sustainable Forest Management (SFM – Maesano et al. 2018) and its accurate estimation has environmental, economic, and social implications. Nevertheless, its quantification still presents challenges from local to global levels. Indeed, the conventional field-based methods are labour- and time-intensive as well as generally expensive. Above Ground Biomass (AGB) accounts for 70% to 90% of total forest biomass and can be analyzed and correlated with remote sensing data (Kumar & Onisimo 2017). Forest inventory data integrated with remote sensing data has become a powerful technique to estimate AGB (Dang et al. 2019). Based on information captured by remote sensors, AGB has been correlated with ground truth for defining biomass estimation models (Lu et al. 2016). Particularly useful is the use of Light Detection and Ranging (LiDAR) data for defining forest attributes like tree height, which is directly linked to forest biomass (Mura et al. 2015, Sačkov et al. 2016). In addition, the use of high-resolution RGB images has emerged as extremely functional (Watts et al. 2012) and highly complementary to LiDAR data (Lu et al. 2016). Many studies estimate AGB using remote sensing data from satellites or airborne equipped with different types of optical sensor data (Anaya et al. 2009, Lu et al. 2016). However, these technologies have constraints due to spatial and temporal resolution. Unmanned-Aerial-Vehicle (UAV) platforms are viable alternatives to airborne and satellite platforms, as they are characterized by low cost and an excellent trade-off between resolution, scale, and frequency data at the local level (Jayathunga et al. 2018). Several approaches using various predictor variables derived from remotely sensed images have been developed in order to improve the accuracy of estimation of AGB.
for different forest types and management goals (Silveira et al. 2019). The most common method used for AGB estimation from remote sensing data is the area-based approach (ABA – Hawyrolo et al. 2017). The aim of the ABA is to derive a predictive model that links LiDAR metrics \( X \) to a target attribute measured \( Y \) – e.g., AGB – at a specific ground location. This integrated approach tries to fill the four critical issues in the biomass estimation model from remote sensing data: (i) the sample plot number, (ii) the metrics selection, (iii) the suitable algorithms, and (iv) the model accuracy. Conventional regression techniques are commonly used to estimate AGB from remote sensing, though in the last few years the machine learning techniques are increasingly employed (Maxwell et al. 2018), as higher accuracy can be generally obtained compared with traditional methods and better ability in identifying the relationships between predictors and the AGB from field data (Chen et al. 2018).

A huge variety of machine-learning algorithms has been investigated to estimate forest biomass, such as K-nearest neighbours (KNN – Chirici et al. 2012), Support Vector Regression (SVR – Sivasankar et al. 2019), Artificial Neural Networks (ANN – Ingram et al. 2005), but the most widely used algorithm is the Random Forest (RF – Dang et al. 2019), as it does not need assumptions about the initial data and the relationship between the response variable and predictive variables.

Some previous studies have estimated forest biomass using the multi-sensor data and machine learning approach for classification and forest biomass estimation (Dhanda et al. 2017). In this study, a unique combination of LiDAR with high-resolution optical data by UAV was adopted, aiming to evaluate the accuracy of two different predictive models to estimate forest biomass. In particular, we compared the accuracy of AGB estimates obtained with a machine learning algorithm, the Random Forest (RF), with a typical stepwise linear regression model (Steplm). Using remote sensing data derived by the UAV platform we produced spatial maps of AGB for each model. The study was carried out in a forested watershed belonging to the Calabrian Apennine Mountains (southern Italy).

## Materials and methods

### Study area

The experimental area is a mountain basin called “Bonis”, located in the Calabria Region, Southern Italy (39° 25′ 15″ N, 16° 12′ 38″ E – WGS 84). The Bonis basin covers 140 hectares with altitude ranging from 975 to 1301 m a.s.l. (average = 1129 m) and a mean slope of 40%. The mean annual precipitation is 1179 mm distributed over 97 rainy days, on average, and the mean annual air temperature is 9 °C. The forest stands in the study area are mostly pure stands of Calabrian pine (Pinus nigra ssp. laricio (Poir.) Maire) planted between 60 and 50 years ago. These stands cover about 95 hectares, while the remaining 45 hectares are composed of riparian vegetation, agricultural fields, roads, bare soil and a small portion of mixed forest, where Calabrian pine is mainly mixed with chestnut (Castanea sativa Mill) and alder (Alnus glutinosa [L.] Gaertn.). The study presented in this paper is focused only on the pure pine forests (Fig. 1).

### Field data

Field data were collected within 16 circular inventory plots located in the Calabrian pine pure forest. The distribution of the inventory plot was chosen at random. Each plot centre was georeferenced through a global navigation satellite system (GNSS) receiver. Real-Time Kinematic GPS (RTK-GPS) Leica CS10 GNSS® (Leica Geosystems AG, Switzerland) with sub-centimetre precision was employed. Each plot was 20 m in radius (0.13 ha). Within each plot, the following parameters were measured on each individual tree: diameter at breast height (DBH), tree height (H), crown diameter (CD), and the vitality of trees with DBH>5 cm. Tree height and the width of horizontal tree crown projection were recorded using a Vertex III hypsometer and Transponder T3® (Haglöf, Sweden), while callipers were employed to measure diameters. Each
Biomass estimation from UAV by RGB images and LiDAR data

Remote sensing data
The SenseFly eBee Ag fixed-wing UAV (Switzerland) was used for image acquisition. UAV images were collected in July 2017 using a Canon PowerShot 110 HS RGB camera (16.1 MP) with an average Ground Sampling Distance (GSD) of 6.59 cm (Fig. 2). Seven Ground Control Points (GCPs) were measured with mean RMSE error = 0.065 m accuracy. The photogrammetric point clouds provide information on the tree-top height of the forest canopy, while the Digital Terrain Model (DTM) was derived by the LiDAR campaign (summer 2017).

The LiDAR data were acquired by a system mounted on a helicopter; laser points have been captured at angles perpendicular to the terrain (LiDAR full-waveform Riegl LMS-Q560 with frequency 400,000 Hz, FOV 60°, inclination 20°). The system was set to acquire 20 points per square meter flying at 700 meters above ground level (AGL) and geo-referenced to the projection system of the UTM zone 33N, WGS 84. The data were acquired within the ALForLab project (https://www.alforlab.it).

Remote Data Processing
Photogrammetric processing of the images was carried out through an automated system based on Structure from Motion (SfM) algorithms. The SfM workflow was implemented in the commercial software package Agisoft PhotoScan Professional (Agisoft LLC, Russia) to obtain 3D models from 2D images. For better results, the images were pre-processed improving colour correction, tone scale, edge enhancement and white balance. Based on this, a high-quality orthomosaic and a Digital Surface Model (DSM) were produced. The free and open-source desktop software QGIS version 3.2 (QGIS Development Team 2018) was used to define the Canopy Height Model (CHM) with a grid cell size of 6.59 x 6.59 cm using the DSM and Digital Terrain Model (DTM). To derive CHM metrics and reduce time-consuming processes, the orthomosaic and CHM layer were clipped over the sixteen sample areas in QGIS. Procedures for images segmentation and classification for extracting forest canopy in the sample areas were carried out by the software Trimble eCognition Professional developer v. 9.1 (Trimble GeoSpatial, Munich, Germany). Multiresolution segmentation based on Object-Based Image Analysis (OBIA) techniques was used, while a CHM mask and a Green Leaf Index (GLI) mask were employed for the forest vegetation classification. The CHM mask included the segmentation polygons taller than 5 m height, according to the Italian forest definition (INFC 2005), while for the second mask, polygons with GLI>0 were considered. The GLI index was calculated using the following formula (Louhaichi et al. 2001 – eqn. 1):

\[
GLI = \frac{R - (G + R)}{R + G + B}
\]

where R, G, and B are the digital number (0 to 255) of red, green and blue channels, respectively. GLI index has values between -1 and +1 and the positive values are related to the green leaves. Afterwards, the classification results by eCognition® were exported in QGIS and the LiDAR-derived metrics from the CHM were extracted. For each plot, we calculated twelve CHM metrics (Tab. 1).

Above ground biomass modelling and accuracy assessment
Metric derived by CHM are candidate predictor variables (independent variables) which can be correlated with AGB calculated by field data at plot level (dependent variable in Mg ha⁻¹) using statistical methods such as stepwise regression model (Steplm) and Random Forest (RF). StepLm was the traditional statistical method used into the study. The stepwise method provides the inclusion of each predictor one-by-one in the model, testing whether the included variable improves the model performance based on AIC method (Akaike Information Criteria). If the addition of the variable worsens the model accuracy, the predictor is then taken out. The multicollinearity between the independent variables was estimated by the Variance Inflation Factor (VIF) using the package "car" in R environment (Fox & Weisberg 2019), and variables with VIF > 10 were discarded (O'Brien 2007). Pearson's correlation coefficient (r) was calculated between predictor variables and AGB.

Global validation of linear model assumptions was performed using the package "gvlma" in R environment. To assess the prediction accuracy of the AGB estimation model obtained, the Leave-One-Out Cross-Validation method (LOOVC) was performed. RF is a powerful non-parametric machine learning algorithm that can be used for both regression and classification. As for regression, RF generates an arbitrary number of simple trees (a subset of independent variables – CHM-derived metrics), which are used to obtain an estimate of the dependent variable (AGB). In this context, RF algorithm was applied to predict the AGB from 12 CHM-derived metrics obtained from high-resolution imagery. The 16 inventory plots were randomly split into a training dataset and validation dataset in the ratio of 70:30. The best setting of the RF model had 6 variables, 500 decision trees and 3 as the number of variables per level. In order to improve the accuracy of predicted AGB, the OOB error (out of bag)

Tab. 1 - CHM-derived metrics.

<table>
<thead>
<tr>
<th>CHM Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H_count</td>
<td>Total of the pixels value</td>
</tr>
<tr>
<td>H_sum</td>
<td>Sum of the pixels value</td>
</tr>
<tr>
<td>H_mean</td>
<td>Average of the pixels value</td>
</tr>
<tr>
<td>H_median</td>
<td>Median of the pixels value</td>
</tr>
<tr>
<td>H_stdev</td>
<td>Standard deviation of the pixels value</td>
</tr>
<tr>
<td>H_min</td>
<td>Minimum of the pixels value</td>
</tr>
<tr>
<td>H_max</td>
<td>Maximum of the pixels value</td>
</tr>
<tr>
<td>H_range</td>
<td>Range of the pixels value</td>
</tr>
<tr>
<td>H_minority</td>
<td>Minority of the pixels value</td>
</tr>
<tr>
<td>H_majority</td>
<td>Majority of the pixels value</td>
</tr>
<tr>
<td>H Variety</td>
<td>Variety of the pixels value</td>
</tr>
<tr>
<td>H_variance</td>
<td>Variance of pixel values</td>
</tr>
</tbody>
</table>

Fig. 3 - Importance of variables, percentage of the increase of mean-squared error (%IncMSE, left panel) and the increase in node purity (IncNodePurity, right panel) for AGB estimation using the RF model.
was used as the minimization criterion. For removing redundant information all CHM-derived metrics were explored and the optimal variables for forest AGB mapping were determined. To this purpose, the percentage increase in the mean-squared error (%IncMSE) and the increase in node purity (IncNodePurity) were calculated to evaluate the variable importance. The larger the %IncMSE and the IncNodePurity of a variable, the more important is the variable (Fig. 3).

Modelling and accuracy assessment of AGB estimations were conducted by applying the Steplm and RF algorithms in the R open source environment (R Core Team 2019). Coefficient of determination (R²), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were calculated to compare the performance of the two algorithms.

Results

In the study site of the Bonis Basin, 1065 trees were measured in the 16 circular plots (sample areas). At the plot level, high-resolution RGB images by UAV platform were collected. At the plot level, the DBH ranged from 6 to 80 cm, with an average value of 32 cm, while the tree height ranged from 4.5 to 35 m, with an average value of 24.4 m. The AGB calculated using the species-specific allometric relationships ranged from 189.82 to 527.13 Mg ha⁻¹, with an average value of 383.09 ± 104.16 Mg ha⁻¹ (Tab. 2).

Pearson’s correlation analysis was performed to analyse the relationship between CHM-derived metrics and computed AGB from field data for each sample area. According to the goodness of fit, a strong positive relationship between AGB and H_sum (r=0.83), H_mean (r=0.72) and H_Median (r=0.70) were found (Fig. 4).

In the first approach, the Steplm algorithm was used for variable selection (CHM-metrics). The Steplm algorithm allowed for the optimal combination among predicted variables that produced the smallest AIC value and the root mean squared error of the estimate (RMSE). CHM metrics, including H_sum, H_majority, H_variance and H_median from CHM were selected to predict total biomass, with an R² of 0.92 and an RMSE of 27.3 Mg ha⁻¹. The best model selected for AGB estimation was as follows (eqn. 2):

\[
AGB = 1.42 + (4.76 \times 10^{-5} \times H_{sum}) \\
+ (2.75 \times H_{majority}) \\
+ (4 \times H_{variance}) \\
+ (1.57 \times H_{median})
\]  

(2)

The model validated with LOOCV had an accuracy of R² of 0.81 and an RMSE of 45.5 Mg ha⁻¹.

In the second approach, the RF model was applied to estimate the AGB based on CHM metrics. The RF model showed better accuracy than the Steplm model (Tab. 3, Fig. 5). The R² increased from 0.81 to 0.86, and the RMSE and MAE values decreased from 45.5 to 31.7 Mg ha⁻¹ and from 34.2 to 22.1 Mg ha⁻¹, respectively (Tab. 3).

Two maps of the Calabrian pine forests in the Bonis Basin have been generated. The Steplm model map showed an AGB mean of 303.18 ± 89.4 Mg ha⁻¹, while the random forest model map resulted in a
mean AGB of $314.38 \pm 30.2$ Mg ha$^{-1}$.

**Discussion**

The presented method combines texture and canopy information derived by RGB imagery and LiDAR, and integrates a forest allometric model to estimate the above-ground biomass. In the literature, several methods have been used to estimate forest biomass and AGB from remote sensing data through LiDAR sensors (Shao et al. 2018), satellite imagery (Li et al. 2019) or both (Shang et al. 2019), but few studies used RGB images from Unmanned Aerial Vehicles (UAV – Lu et al. 2019). Both AGB maps of the Bonis Basin provide similar AGB mean values but a different distribution of the forest biomass, with lower variability in the random forest model (Fig. 6). The AGB values derived by RF can be considered more reliable than those calculated with StepIm, because all the Calabrian pine pure forest stands are even aged. This is statistically confirmed by the very high accuracy of the RF model and the lower “salt-and-pepper” map obtained.

For forest biomass models derived from remote sensing data, the RF non-parametric model represents one of the best estimation models in terms of accuracy, with larger $R^2$ and smaller RMSE as compared to other parametric and non-parametric methods (Sun et al. 2019). The linear regression achieved less accuracy of the estimate, probably due to the nonlinear relationship between remote sensing predictors and the observed forest AGB. Therefore, the results obtained in this study can be considered satisfactory with the coefficient of determination of 0.81 and RMSE of 45.5 Mg ha$^{-1}$ for stepwise regression and 31.7 Mg ha$^{-1}$ for random forest algorithms, respectively. These values are consistent with those reported by He et al. (2013) in a coniferous forest ($R^2 = 0.73$ in the regression model) and by Liu et al. (2017) in a mixed forest ($R^2 = 0.95$ with the random forest), while Lefsky et al. (1999) reported a coefficient of determination of 0.81 and an RMSE of 45.8 Mg ha$^{-1}$ in a deciduous forest.

Moreover, the study area is subject to anthropogenic and natural disturbances such as logging and forest fire that can affect ecosystem dynamics and forest structure (Moresi et al. 2019). The produced above-ground biomass maps quantify the biomass inside the forested basin and can be useful to estimate ecosystem services such as carbon stock, canopy status and health of the forest.

**Conclusions**

A comparative study was conducted for evaluating the forest AGB estimation accuracy using a ready-to-fly UAV to produce very high-resolution images. Recently, the use of UAVs is increasingly used for research purposes. Thanks to the combined use of multi-sensor remote sensing data and ground-truth, the accuracy of AGB estimation of forests has been improved (Lu et al. 2019), though there is still room for further improvements and to increase the estimation accuracy (Krisanski et al. 2018).

The UAV platforms equipped with various sensors potentially combine the advantage of collecting and delivering data useful for...
tree parameters estimation and forest monitoring in real-time with high spatial resolutions at reasonable costs. Moreover, remote sensing technology is a non-destructive method to estimate the AGB without negative impacts on the forest ecosystem (Ku & Popescu 2019) and is applicable to large areas and/or areas that are difficult to access, but it can also be viable for small forest owners (Di Lallo et al. 2016).

Nowadays, the high-resolution acquisition of RGB images is simple, making the use of these technologies less expensive and easy-to-use compared to other remote sensing data. This study demonstrated the viability of using machine learning algorithms to predict forest AGB and showed that the Random forest machine learning algorithm led to more accurate AGB estimates than the traditional stepwise regression model.

The presented approach can be considered a relatively low-cost method of forest AGB estimation if DTM and DSM layers or Lidar data are already available. In this context, a commercial fixed-wing drone equipped with an RGB camera was used. Indeed, once the biomass model for AGB estimation is established, new UAV flights can be planned through time to monitor the development of AGB and forest status within the basin. Remote sensing by UAV can be an active monitoring system for forest assessment and management at the local scale to meet specific environmental, economic, social, and cultural objectives (Pastorella et al. 2016).

This study demonstrated that by applying a machine learning algorithm, readily available images can be used to obtain reliable results, as demonstrated by the accuracy of the above-ground biomass estimation by the Random forest model.

Author Contributions

Acknowledgements
We would like to thank Dr. Gaetano Pelligcone, Dr. Antonella Velteri, Dr. Raffaele Froio for the technical assistance in field data collection.

Funding
This research was funded by: National financial resources (Revolving Fund Cohesion Action Plan (CAP), MIUR; Research Project PRIN-MIUR 2015 “The forest-wood value chain: biomass supply, traceability, C-footprint. Innovation for bioarchitecture and energy efficiency”; and the research project “ALForLab-Chain Value Forest- Wood-Products” (PON003PE_00024_1) cofunded by the Italian National Operational Programme for Research and Competitiveness (PON R&C) 2007-2013, through the European Regional Development Fund (ERDF).

References
INFC (2005). Linee generali del progetto per il secondo inventario forestale nazionale italiano [General lines of the project for the second Italian national forest inventory]. MIAPF – Ispettorato Generale del Corpo Forestale dello Stato, CRA-ISAFA, Trento. [in Italian]

456
iForest 15: 451-457

iforest • Biogeosciences and Forestry
Biomass estimation from UAV by RGB images and LiDAR data