

ALEPPO PINE ALLOMETRIC MODELING THROUGH INTEGRATING UAV IMAGE-BASED POINT CLOUDS AND GROUND-BASED DATA

Fernando J. Aguilar^{1,*}, Abderrahim Nemmaoui¹, Manuel A. Aguilar¹, Rafael Jiménez-Lao¹

¹ Department of Engineering and Research Centre CIAIMBITAL, University of Almería, Almería, Spain
(faguilar, an932, maguilar, rjl020)@ual.es

Commission III, WG III/10

KEY WORDS: Mediterranean Forest, Aleppo Pine, UAV, Point Cloud, Machine Learning, Allometric Models, AGB Mapping.

ABSTRACT:

Effective monitoring of Mediterranean forest is essential to determine the role of forest management in mitigating climate change and ensuring the maintenance of its environmental services. Most of the allometric models to estimate dry above-ground biomass (AGB) at tree level are based on knowing the diameter of the trunk at breast height (DBH). However, it is difficult, if not impossible, to estimate DBH from airborne/spaceborne sensors within the context of a remote sensing-oriented approach, being common to draw upon regression models to relate DBH to remotely sensed dendrometric variables such as total tree height (H) and tree crown diameter (CD). This study uses UAV (unmanned aerial vehicle) image-based data to estimate the dendrometric variables H and CD of the repopulated Aleppo pine (*Pinus halepensis* Mill.) located in a semiarid continental Mediterranean forest of Almería (southeast of Spain). DBH data were gathered through field work. Both bivariate ($DBH = \Phi(H)$) and multivariate ($DBH = \Psi(H, CD)$) allometric models were developed by applying least-squares-based regression and machine learning regression methods. The results showed that multivariate allometric models performed better than bivariate at predicting DBH, both in terms of goodness-of-fit and stability against changes in training or testing samples. In addition, least-squares-based regression (linear and potential) provided statistically similar results to those obtained from complex machine learning ensemble algorithms. In this way, the easy-to-apply multivariate linear allometric model $DBH = -4.84 + 1.73149H + 3.08114CD$ ($R^2 = 89.23\%$) would be recommended to locally estimate DBH in Aleppo pine from remotely sensed H and CD data.

1. INTRODUCTION

According to the general definition of forests by FAO, in 2015 there were around 88 million ha of forest area in Mediterranean countries, 18 million of them located in Spain (FAO, 2015). This means that forests occupied 10.04% of the total area of Mediterranean countries in 2015, equivalent to the combined size of Spain and Morocco, also storing 5066 billion tons of carbon, equivalent to 1.7% of global forest carbon (FAO, 2015).

Focusing on the importance of Mediterranean forest, the Mediterranean region has more than 25 million ha of Mediterranean forests and about 50 million ha of other Mediterranean wooded lands (e.g., open oak woodlands of *Quercus* species such as Spanish *Dehesas*) (FAO and Plan Bleu, 2018) that can effectively contribute to carbon storage and climate change mitigation. Furthermore, they host a large variety of forest ecosystems, contain an impressive plant and animal diversity, and generate a large number of environmental services (ES) that make crucial contributions to rural development, poverty alleviation, food security, as well as the agricultural, water, tourism, and energy sectors (Vilà-Cabrera et al., 2018).

We know that forest area in Mediterranean countries has been increasing since 1990. For example, between 2000 and 2015 there has been an increase of 8 million ha in forest area (FAO and Plan Bleu, 2018). This increase in forest size is both the result of the European Common Agriculture Policy (as in the case of Spain) and forest regeneration in rural areas following abandonment (Nogueira and Rico, 2017). Unfortunately, an increasing forest area, despite being good news, tell us nothing about forest degradation and potential capacity to adapt to

climate change. It is needed to take a closer look to cope with monitoring forest structure at stand and tree level. Note that forest structure is formed through the action of very diverse factors such as silvicultural practices, fires, droughts, heat waves, storms, and pests or diseases, being the key that explains, in turn, relevant variables of ecosystems such as biodiversity, erosion control, water availability and landscape complexity.

Taking into account the aforementioned antecedents, it seems clear that effective monitoring of Mediterranean forest structure turns out to be a key role for adapting to climate change (Pascual et al., 2020), ensuring at the same time the maintenance of ES and the preservation of the functional characteristics of Mediterranean forests (Vilà-Cabrera et al., 2018). But how to deal with it? The Intergovernmental Panel on Climate Change (IPCC) recommends combining remote sensing and ground-based data to estimate forest area, carbon stocks and changes at large scales (Espejo et al., 2020). However, this integrated approach requires the development of new allometric tools based on individual tree measures (tree-centric approach) collected from spaceborne or airborne sensors (e.g., total tree height (H) and tree crown diameter (CD)) to make full use of remote sensing techniques in implementing enhanced forest inventories and mapping carbon stocks (Coomes et al., 2017; Jucker et al., 2017; White et al., 2016).

Within available remote sensing techniques, the recently emerged computer vision algorithm Structure from Motion (SfM) with Multi-View Stereo (MVS) (Furukawa and Ponce, 2010) has boosted efficiency in building very dense and accurate 3D point clouds of comparable quality to laser-based methods (Wallace et al., 2016). When coupled with Unmanned Aerial

* Corresponding author

Vehicle (UAV) very-high resolution images, it has demonstrated to be able to generate a Canopy Height Model (CHM) that can be successfully used to estimate total tree height in forest inventories (Goodbody et al., 2017; Lisein et al., 2013; Panagiotidis et al., 2017; White et al., 2013).

On the other hand, tree dry above-ground biomass (AGB) is traditionally estimated using forest inventory data from sample plots and allometric models that usually rely on stem diameter (diameter at breast height: DBH) and H as key inputs (Chave et al., 2014). However, it is impossible to measure DBH from airborne or spaceborne sensors, being necessary to carry out traditional field work or terrestrial laser scanning inventories to obtain this dendrometric variable (Peñalver et al., 2021). In this regards, it is crucial to develop locally calibrated allometric models to allow predicting DBH from other dendrometric variables such as H and CD that can be estimated from airborne or spaceborne sensors. The fitting of this site-specific allometric relationships can be faced by using both least-squares-based regression and machine learning regression methods (Aguilar et al., 2021).

This work uses UAV image-based data to estimate the dendrometric variables H and CD in a Mediterranean forest mainly composed of repopulated Aleppo pine (*Pinus halepensis* Mill.). DBH data were collected almost simultaneously through field work, allowing the development of bivariate and multivariate allometric models to predict DBH from H and CD. These models were fitted by using both least-squares-based regression (linear and potential models) and supervised machine-learning regression methods. These locally obtained allometric models could be used to improve forest AGB and carbon estimation, especially in large-scale inventories where only H and CD can be estimated from airborne or spaceborne sensors.

2. MATERIALS AND METHODS

2.1 Study Site

The study site is located at “Sierra de María-Los Vélez” Natural Park (Fig. 1). This Natural Park is a Spanish protected natural area located northwest of the province of Almería (southeast of Spain), Andalusia.



Figure 1. Location of the study area. a) Province of Almería (Spain) in yellow. b) Region of “Los Vélez” (in red). c) Limits of the “Sierra de María-Los Vélez” Natural Park (in green). d) Reference field plots represented as white dots.

“Sierra de María-Los Vélez” was declared a natural park in 1987, occupying an area of 22,562 ha and a maximum elevation of 2,045 m AMSL. It presents an average annual rainfall of about 400 mm, with an average annual temperature of 11° C and semiarid continental Mediterranean climate.

Although in the most humid areas of the natural park there can be small forests of Laricio pine and deciduous trees, the most representative forest ecosystem in this area presents a two layers structure. The emergent upper layer (dominant trees) is composed of repopulated Aleppo pine (*Pinus halepensis* Mill.), while the lower canopy layer (understory vegetation) is mainly formed of little holm oak trees (*Quercus ilex* L.) and different species of shrubs (Aguilar et al., 2019a). This forest structure is considered very representative of the Mediterranean forest.

2.2 Field Data and Processing Methods

The test field sites of this work consisted of 18 square plots of 100 m side (1 hectare) that were selected trying to represent the variability of the stand density and tree height of the Aleppo pine population in the Natural Park (white dots in Fig. 1).

High resolution UAV images for each reference plot were obtained in March 2021 by using a DJI Phantom 4 Advanced®. Its 8.8 mm focal length integrated RGB camera is equipped with a 20 megapixel 1” CMOS sensor (2.52 $\mu\text{m}/\text{pixel}$) fitted to a 3-axis stabilized gimbal to maintain nadir image capture. The flying height of the mission was set to approximately 75 m above ground, which allowed yielding an average ground sample distance of 2.1 cm. High forward and side overlaps of 90% and 85%, respectively, were set in the UAV flight plan of each reference plot to avoid potential forest occlusions.

Between seven and eight ground control points (GCP) constituted of rectangular 60x40 cm wood panels (black and white chess-board style painted) targets were evenly distributed over each reference plot, trying to choose open terrain sites to ensure their visibility in the UAV images. Those GCP were surveyed with a couple of GNSS RTK multiband receivers Emlid Reach RS2 (rover and base). The geographical coordinates of the base in each reference plot were obtained by applying the differential corrections provided through NTRIP caster from the continuous reference GPS station of Huércal-Overa (Andalusian network of GPS positioning). The rover receiver was used to measure the ETRS89 UTM 30N projected coordinates and EGM08 REDNAP (Spanish High Precision Levelling Network) orthometric heights of each GCP in RTK mode using the base receiver as a reference.

All images for each reference plot were photogrammetrically processed using the SfM-MVS algorithm implemented in the software Agisoft Metashape Professional 1.7.2 (Agisoft LLC, St. Petersburg, Russia), a well-known SfM-MVS software capable of producing image-based high quality point clouds that has been widely employed to conduct UAV-based forest inventories (e.g., Chen et al., 2021; Jensen and Mathews, 2016; Panagiotidis et al., 2017; Wallace et al., 2016). Metashape was also used to produce a high-resolution RGB 3 cm/pixel orthoimage.

The seven-eight surveyed GCP were manually marked on the digital images to carry out an iterative bundle adjustment to estimate the 3D coordinates of the automatically matched features (sparse point cloud) (Aguilar et al., 2019b). Next, a camera self-calibration process was performed in each reference plot to optimize the camera model, always maintaining fixed the focal length. After estimating internal and external camera

orientation parameters, the depth information of each image was combined through a multi-view reconstruction of the scene geometry into a single and very dense 3D point cloud, thus generating very high density UAV point clouds ranging from 800 up to 1500 points/m². Note that the selected Agisoft Metashape parameters were those recommended by Tinkham and Swayze (2021) to carry out individual tree detection from CHM. In this way, the photo alignment process was performed on the original images (i.e., without changing their original spatial resolution or "High Accuracy" setting), while the dense cloud was computed by selecting "High quality" (i.e., original image size downscaling by factor of four) and "Mild" depth filtering settings.

The point cloud corresponding to each reference plot was automatically classified into ground and non-ground points through applying the filtering algorithm triangular irregular network (TIN) iterative approach proposed by Axelsson (2000) and implemented into Agisoft Metashape. After a trial and error procedure, the set of chosen parameters was cell size = 10 m, distance = 0.3 m and angle = 30°. Those potential outliers in the automatically filtered ground points were automatically removed by adapting the parametric statistical method for error detection in digital elevation models published by Felicísimo (1994) to be used on scattered points, then setting the neighbourhood size radius to 15 cm. In addition, a local maxima filtering with a neighbourhood radius of 10 cm was applied to each UAV point cloud to obtain the corresponding Canopy Surface Model (CSM).

Both CSM and DTM derived point clouds were finally interpolated to 5 cm grid spacing products by using the Gaussian Markov Random Field (GMRF) algorithm developed by Aguilar et al. (2016) (<https://github.com/3DLAB-UAL/dem-gmrf>). The grid DTM was finally subtracted to the grid CSM to obtain a 5 cm grid spacing Canopy Height Model (CHM).

Stratified random sampling was carried out in the 18 reference plots to manually digitize onto the RGB 3 cm/pixel orthoimage the 2D crown boundary (shp file) of up to 785 Aleppo pines, also estimating their corresponding total tree height and planimetric coordinates from the previously obtained CHM (given by the point with maximum height within the digitized crown boundary of each sample tree). The crown diameter of each tree was computed from the area of its crown boundary. This office work was done in the ArcGIS environment.

Fieldwork was carried out between the end of May and the beginning of June 2021 to locate and measure the DBH and total tree height of the 785 sample trees. A pair of Emlid Reach RS2 GNSS RTK multiband receivers allowed the precise location of each tree in the field to measure its DBH (diameter at 1.3 m above ground) using a tape measuring. Furthermore, a Nikon® Forestry Pro II rangefinder/hypsometer was used to measure tree height, although only 451 Aleppo pines out of 785 sampled could be measured acceptably in this case due to occlusions caused by the surrounding trees in high density stands.

2.3 Allometric Models

Seven allometric models were tested to predict DBH from the explanatory variables H (bivariate model) and H + CD (multivariate model). One of them was based on traditional linear regression, while another used a potential form. The remaining five focused on supervised machine learning algorithms. An individual-based modelling approach was used by considering each individual tree measurement as an instance of the relationships modelled.

The allometric linear model tested took the following form:

$$DBH = \alpha + \beta H + \varepsilon \quad \text{Bivariate model,} \quad (1)$$

$$DBH = \alpha + \beta H + \gamma CD + \varepsilon \quad \text{Multivariate model,} \quad (2)$$

where DBH = diameter at breast height (cm)
 H = total tree height (m)
 CD = crown diameter (m)
 ε = fitting error
 α = intercept (to assure that the model will be unbiased)
 β, γ = regression coefficients for H and CD

The potential model used in this study was based on the allometric model proposed by Jucker et al. (2017). After taking logarithms to linearize the potential expression, we obtain the following equations:

$$DBH = e^{(\alpha + \beta \ln(H))} e^\varepsilon = e^{(\alpha + \beta \ln(H))} e^{\frac{\sigma^2}{2}}, \quad (3)$$

$$DBH = e^{(\alpha + \beta \ln(H \cdot CD))} e^\varepsilon = e^{(\alpha + \beta \ln(H \cdot CD))} e^{\frac{\sigma^2}{2}}, \quad (4)$$

where DBH is given in cm, and H and CD in m. α and β are model coefficients, and ε is an error term. If it is considered that the error term is normally distributed with zero mean and standard deviation σ , the mean of e^ε could be approximated by $e^{\frac{\sigma^2}{2}}$ (Jucker et al., 2017). This additional term would work as a correction factor applied to back transform the predicted values and remove bias from the logarithmically transformed data.

Regarding supervised machine-learning methods, this work has tested five tree-based regression learners. The first was Decision Tree Regression (DTR), which is only based on one tree model. The remaining four were some derive ensemble algorithms (based on several tree models built on datasets randomly extracted from the original sample) that can be grouped in bagging (Random Forest Regression: RFR) and boosting algorithms (AdaBoost Regression: AdaBoost, Gradient Boosting Regression: GBoost, and Categorical Boosting Regression: CatBoost). The optimal combination of hyperparameters for each machine-learning model was computed by applying grid search with cross-validation method (Zhang et al., 2020).

The accuracy assessment of the allometric models tested was based on the fact that the data used to train the model could never be used for validation. In this way, the testing set for validation consisted of 20% of the 785 available trees, leaving the remaining 80% as a set for training and computing the regression model. This procedure was repeated 100 times, splitting the original data between the training and testing sets by using random sampling. It allowed to study the stability of the goodness-of-fit (R^2) for the tested regression models against changes in the training and test samples.

Some error indicators of the DBH values predicted by the regression models were calculated according to the following expressions:

$$Bias (\%) = \frac{100}{N} \sum_{i=1}^N \left(\frac{DBH_{p_i} - DBH_{o_i}}{DBH_{o_i}} \right), \quad (5)$$

$$RMSE (cm) = \sqrt{\frac{\sum_{i=1}^N (DBH_{p_i} - DBH_{o_i})^2}{N}}, \quad (6)$$

$$relative RMSE (\%) = 100 \frac{RMSE}{DBH_o}, \quad (7)$$

where DBH_p = DBH predicted
 DBH_o = DBH observed
 N = number of pine trees in the testing dataset
 $\overline{DBH_o}$ = mean value of DBH observed values

Note that the Bias indicator constitute a measure of the systematic error of the model, while RMSE (root-mean-square error) is a quantitative indicator of its random error. The entire procedure mentioned above was coded in Python 3.8 with the support of the scikit-learn and catboost libraries.

3. RESULTS

3.1 Field Measured Tree Height vs Tree Height Estimated from Image-Based Point Clouds

Figure 2 graphically shows the good agreement between the total tree height measured with a hypsometer (H_{hyp}) during the field work and that estimated from the photogrammetrically derived CHM (H_{CHM}) for the sample of 451 Aleppo pines. In fact, the coefficient of correlation (Pearson's r) between H_{CHM} and H_{hyp} took a high value of 0.9929. Moreover, the RMSE value of the residuals (prediction errors) was only 40.65 cm, which meant a very low relative RMSE of 4.57% (< 5%) with respect to the mean value of the tree height measured with a hypsometer.

These results evidenced the high geometrical accuracy of the CHMs generated from UAV image-based point clouds, which allowed us to expand the sample of Aleppo pines to 785 by adding those that could not be measured during field work with the hypsometer due to occlusions caused by surrounding trees.

Similar results have been reported by other authors, who generally have found high agreement between field and remote-sensed data for total tree heights (Goodbody et al., 2017; Guerra-Hernández et al., 2016; Lisein et al., 2013; Panagiotidis et al., 2017; Wallace et al., 2016). It has been also demonstrated that tree heights computed from photogrammetrically derived CHM can even improve the results provided by terrestrial laser scanning, which usually yields underestimated tree heights (Shimizu et al., 2022).

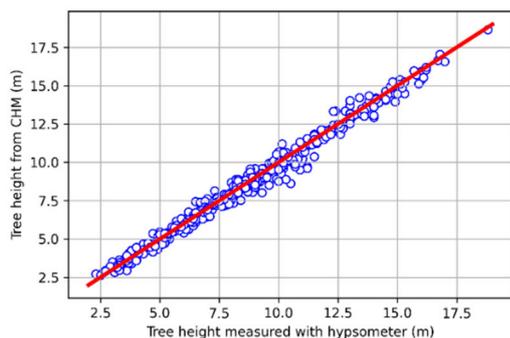


Figure 2. Comparison between tree heights measured with a hypsometer (field work) and extracted from the image-based CHM. Red line refers to 1:1 line.

3.2 Bivariate Regression Models

Table 1 shows the goodness-of-fit (R^2) statistics of the seven bivariate allometric models tested in which the explanatory variable was only H. Regarding R^2 mean value, potential and

linear allometric models were at the top of the list, providing fit figures close to 80%, although they did not return significantly different values compared to those based on boosting machine learning regression methods. In contrast, RFR (bagging algorithm) and DTR (just one decision tree) performed significantly worse ($p < 0.05$) than the other competitors. DTR showed noticeably poorer performance, exhibiting a low R^2 mean value of 62% along with high variability (standard deviation of R^2) across the 100 randomly varying repetitions of the training and test data sets. It is worth noting that small changes in the learning sample can cause dramatic changes in the constructed tree when it is only based on a decision tree, leading to unstable and not robust results. In this sense, most recent studies have adopted multi-tree bagging and boosting ensemble algorithms (Luo et al., 2021; Zhang et al., 2020).

Traditional least-squares-based regression methods (linear and potential) proved to be very competitive, providing results statistically similar to those yielded by complex ensemble boosting algorithms. RFR worked significantly worse ($p < 0.05$) than boosting or least-squares-based regression methods, in addition to a greater variability in prediction when varying training samples. It is necessary to consider that ensemble learning is a branch of machine learning that builds and combines multiple learners (decision trees) to improve the outcomes of the learning process. RFR, as an ensemble bagging method, uses bootstrap samples randomly generated from the original dataset to train several tree models, then aggregating the ensembles to obtain final predictions by majority voting. Because RFR generally improves predictions by decreasing variance and avoiding overfitting, it is most advisable when developing models that include multiple explanatory variables. (Aguilar et al., 2021).

Regression method	R^2 mean value (%)	R^2 standard deviation (%)	R^2 range (min-max %)
Potential	79.61 ^a	2.40	72.56-85.86
Linear	79.41 ^a	2.44	73.80-84.53
CatBoost	77.88 ^a	2.71	68.12-84.57
GBoost	77.73 ^a	2.88	70.17-83.87
AdaBoost	77.30 ^a	2.46	69.61-83.17
RFR	70.58 ^b	3.95	61.11-80.00
DTR	62.01 ^b	5.41	49.17-74.12

Table 1. Statistics of R^2 for bivariate allometric models $DBH = \Psi(H)$. Mean values with different superscript letters are significantly different ($p < 0.05$) (two-sample t-test).

Regression method	Mean Bias (%)	Mean RMSE (cm)	Relative RMSE (%)
Potential	4.38 (1.87)	4.24 (0.17)	19.00 (0.81)
Linear	4.15 (1.81)	4.29 (0.21)	19.20 (0.97)
CatBoost	4.24 (1.90)	4.41 (0.21)	19.80 (1.03)
GBoost	4.33 (2.10)	4.42 (0.24)	19.79 (1.12)
AdaBoost	5.40 (2.28)	4.45 (0.19)	19.91 (0.89)
RFR	3.38 (2.23)	5.09 (0.27)	22.78 (1.21)
DTR	3.00 (2.26)	5.72 (0.30)	25.66 (1.40)

Table 2. Statistics of systematic and random error for bivariate allometric models $DBH = \Phi(H)$. Figures in parentheses correspond to the standard deviation.

Regarding systematic error, the bias of the bivariate allometric models tested in this work is depicted in Table 2, providing mean bias values of less than 5% in all cases, except AdaBoost (5.4%), with a reasonably low standard deviation (values in parentheses). Note that the positive sign of bias in all cases pointed to a slight overestimation of the observed values of DBH.

In the case of random error, the group of best performance allometric models (linear, potential and boosting) yielded mean RMSE values lower than 5 cm in the prediction of DBH, which meant a reasonably low relative RMSE ranging between 19% and 20% approximately (Table 2).

3.3 Multivariate Regression Models

Table 3 presents the goodness-of-fit statistics of the seven multivariate allometric models tested in this work ($DBH = \Phi(H, CD)$). Note that the prediction results were clearly better than those provided by the bivariate allometric models shown in Table 1, especially in the case of machine-learning regression methods, explaining between about 87% and 89% of the variance of the observed data. Again, there were no significant differences ($p < 0.05$) between least squares regression methods and machine learning, in this case also including RFR, which clearly performed better than when applied to the bivariate allometric model. DTR was again ranked the worst, although notably improved its performance when dealing with multivariate regression. All the tested models showed lower variability in R^2 when varying training samples, which meant that multivariate regression provided greater stability than bivariate. These findings point to the need to have CD, obviously together with H, as a key variable to build accurate allometric models to predict DBH. Similar results were recently reported by Aguilar et al. (2021) working in teak plantations in Ecuador.

Regression method	R^2 mean value (%)	R^2 standard deviation (%)	R^2 range (min-max %)
CatBoost	89.37 ^a	1.42	82.67-93.45
GBoost	89.17 ^a	1.42	84.78-91.69
Linear	89.08 ^a	1.23	86.16-91.77
RFR	87.98 ^a	1.57	84.86-92.12
Potential	87.75 ^a	1.61	82.66-91.76
AdaBoost	87.41 ^a	1.58	82.53-91.03
DTR	79.63 ^b	2.97	71.68-85.71

Table 3. Statistics of R^2 for multivariate allometric models $DBH = \Phi(H, CD)$. Mean values with different superscript letters are significantly different ($p < 0.05$) (two-sample t-test).

Regression method	Mean Bias (%)	Mean RMSE (cm)	Relative RMSE (%)
CatBoost	2.18 (1.24)	3.04 (0.18)	13.64 (0.76)
GBoost	2.48 (1.41)	3.04 (0.17)	13.61 (0.74)
Linear	2.90 (1.52)	3.07 (0.15)	13.74 (0.73)
RFR	2.01 (1.24)	3.21 (0.19)	14.41 (0.84)
Potential	2.81 (1.45)	3.24 (0.19)	14.64 (0.85)
AdaBoost	3.98 (1.55)	3.27 (0.18)	14.73 (0.78)
DTR	1.77 (1.76)	4.15 (0.26)	18.63 (1.16)

Table 4. Statistics of systematic and random error for multivariate allometric models $DBH = \Phi(H, CD)$. Figures in parentheses correspond to the standard deviation.

The bias of the multivariate allometric models can be considered low, as can be seen qualitatively in Figure 3 for the case of Linear and DTR models. The quantitative results represented in Table 4 also confirmed this finding, providing low mean bias values of around 3% in all cases, with the exception of DTR, always yielding a slight overestimation of the observed values of DBH. The variability of bias (standard deviation) was also less than that recorded when bivariate models were tested.

Focusing on the random error of the multivariate allometric models, both the regression methods based on least squares and those of machine learning (except DTR) presented mean RMSE values close to 3 cm, thus clearly improving the results provided

by the bivariate allometric models (Table 4). These results, when converted to relative RMSE values, led to low relative random error between 13.61% (GBoost) and 14.73% (AdaBoost), relegating DTR to last position with a relative RMSE value of 18.63%.

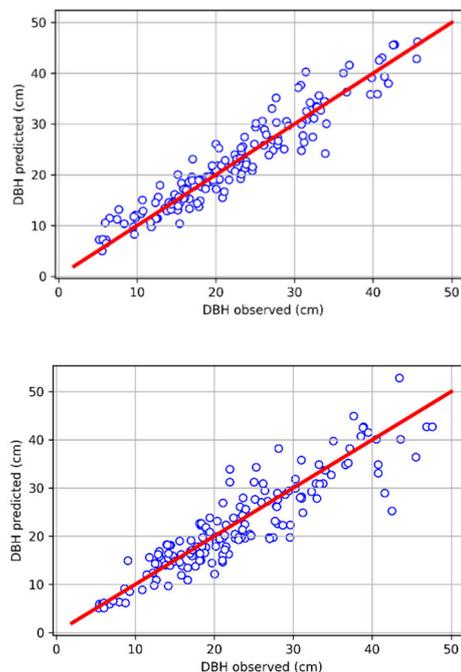


Figure 3. Plots of the observed/predicted values of DBH for a test dataset case (157 Aleppo pine trees not used for training) given by two multivariate allometric models ($DBH = \Phi(H, CD)$). Top: Linear Regression ($R^2 = 89.02\%$). Bottom: Decision Tree Regression ($R^2 = 79.86\%$). Red line refers to 1:1 line.

3.4 Recommended Allometric Model

In view of the results described in sections 3.2 and 3.3, the multivariate allometric model based on Linear Regression (Equation 2) would be the most appropriate locally adjusted model to predict the DBH of Aleppo pine as a function of H and CD in “Sierra de María-Los Vélez” Natural Park. In fact, its goodness of fit was statistically similar to that obtained when applying more complex and sophisticated machine learning regression methods, such as Categorical Boosting or Gradient Boosting, showing even slightly higher stability (i.e., lower standard deviation) against changes in the training data set (Table 3). It should be noted that machine learning or deep learning regression methods are capable of modelling complex non-linear allometric relationships, thus surpassing the linear regression method when applied to tree species that present this type of morphology (Aguilar et al., 2021; Bayat et al., 2020; Chen et al., 2020; Ercanlı, 2020; Vieira et al., 2018). However, this potential improvement is not applicable if the allometric relationships are mostly linear, which may be common in some pine species (Filho et al., 2019). Furthermore, linear models have the advantage that they are easy to fit and apply (for example, simply by using an Excel tabulator).

Figure 4 shows the fit of observed and predicted values of DBH for the entire dataset (785 Aleppo pines) according to the multivariate linear allometric model. It can be seen that the model is unbiased, also explaining up to 89.23% of the variance of the observed DBH values (R^2 adjusted = 89.20%). In addition, the

DBH residuals followed a normal distribution (Figure 5), while both the regression coefficients and the intercept were statistically significant ($p < 0.001$), thus giving rise to the formulation of the following allometric model:

$$DBH = -4.84 + 1.73149H + 3.08114CD, \quad (8)$$

where DBH is given in cm, and H and CD in m.

Considering the allometric equation to estimate AGB proposed by López-Serrano et al., (2005) for Aleppo pine in south-eastern Spain ($AGB = 0.128 \cdot DBH^{2.29}$), where AGB is expressed in kg and DBH in cm, AGB estimates at tree level could be obtained from airborne and spaceborne sensors by only collecting two key variables such as total tree height and crown diameter (Equation 8).

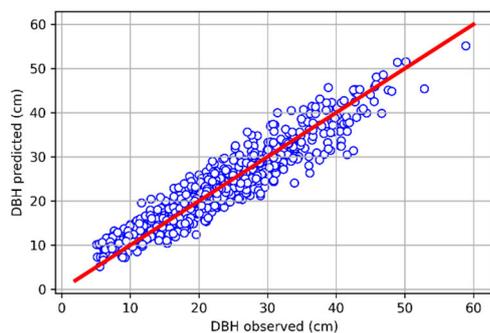


Figure 4. Plot of the observed/predicted values of DBH for the entire dataset (785 Aleppo pine trees) corresponding to the multivariate allometric model based on Linear Regression ($DBH = \Phi(H, CD)$). Red line refers to 1:1 line.

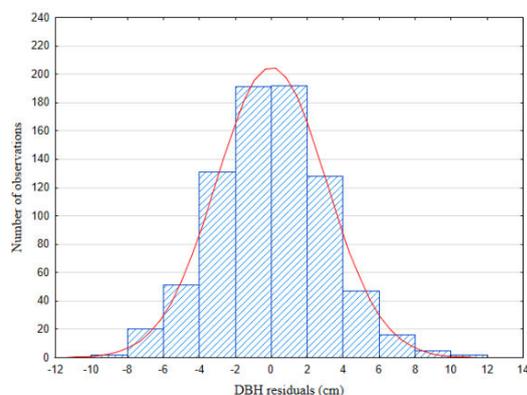


Figure 5. Histogram of the DBH residuals for the multivariate linear allometric model ($DBH = \Phi(H, CD)$). The corresponding normal distribution is overlaid in red.

Note that one of the main drawbacks of using image-based reconstruction in forested areas is the fact that the ground must be visible on aerial imagery, allowing ground points to be generated to build an accurate DTM below the canopy. Indeed, an accurate ground reference is needed to obtain a high-quality CHM in order to correctly calculate tree height. In this sense, it is worth noting that the vegetation cover of the reference plots that participated in this study ranged between 29% and 85% (average value of 49.4%), also having relatively low height pines with Lorey’s mean height (plot-level) ranging from 3.4 m to 16.2 m.

Under these conditions, and although these results are not presented in this paper due to lack of space, the vertical accuracy assessment of the UAV image-based extracted DTMs (UAV-DTM) was performed using airborne LiDAR data available for the study area, yielding systematic error (bias) values of 8.1 cm and 7.6 cm for the mean and median DTM vertical error, respectively. It meant that the UAV-DTM slightly overestimated the z-terrain reference values provided by the LiDAR data. In any case, this reasonably low bias in the construction of the ground reference (DTM) should be interpreted as adequate to support the generation of CHMs aimed at estimating AGB maps over the natural park studied.

Finally, the LiDAR data used as ground truth were provided by the PNOA (National Plan of Aerial Orthophotography of Spain). Data were taken on December 7, 2014, by means of a Leica ALS60 discrete return sensor with up to four returns measured per pulse and an average flight height of 2700 m. The point density of the test area, considering overlapping, turned out to be 0.97 points/m² (all returns). The nominal (at nadir) horizontal accuracy (RMSE_{xy}) and nominal vertical accuracy (RMSE_z) after processing had values lower than 0.3 m and 0.2 m, respectively

4. CONCLUSIONS

In this study, several allometric models were tested to relate DBH (dependent variable) and H and CD (explanatory variables) in the case of the population of Aleppo pine trees in “Sierra de María-Los Vélez” Natural Park. Since UAV carrying visible sensors are gaining ground in local and small-scale data acquisition, UAV image-based point clouds to build CHMs and high-resolution orthoimages were successfully tested to estimate total tree height (Pearson’s $r = 0.9929$; relative RMSE = 4.57%) and tree crown diameter, increasing the efficiency in obtaining these parameters compared to traditional field inventory. Furthermore, this UAV-based technique has proved to be very useful to estimate tree heights that could not be measured during field work (hypsoneter) due to occlusions caused by surrounding trees.

Multivariate allometric models (including H and CD as explanatory variables) performed better at predicting DBH, both in terms of goodness-of-fit and stability of regression models against changes in training samples, than those based solely on H. It highlights the need to count on CD, in addition to H, to build accurate predictions of DBH. Within the tested multivariate allometric models, the linear model showed statistically similar goodness-of-fit than that provided by ensemble machine learning regression methods. This finding is attributed to the linear relationships between DBH and H/CD in the case of Aleppo pine. The machine learning regression method based on just one decision tree (i.e., DTR) performed significantly worse than linear, potential and ensemble machine learning regression methods.

According to the results obtained in this work, the easy-to-apply multivariate linear allometric model would be the recommended model to estimate DBH in Aleppo pine, and consequently AGB at tree level, from total tree height and crown diameter collected from airborne or spaceborne sensors. This remote sensing-oriented approach is gaining importance in recent years because it enables large-scale mapping of AGB for forest management and monitoring in the context of mitigating climate change (Reducing Emissions from Deforestation and forest Degradation (REDD) monitoring programmes).

ACKNOWLEDGEMENTS

This study was funded by the following projects: 1) “*Enabling interdisciplinary collaboration to foster Mediterranean forest sustainable management and socio-economic valuation (ECO2-FOREST)*” (Proyecto Retos Junta de Andalucía, Spain. Grant number P18-RT-2327). 2) “*Intervention strategies for an integrated and sustainable management of the Mediterranean forest based on an interdisciplinary analysis and its economic assessment*” (Programa Operativo FEDER Andalucía 2014-2020, Spain. Grant number UAL2020-SEJ-D1931). The authors wish to thank the support of the Territorial Delegation in Almería of the Ministry of Agriculture, Fisheries and Sustainable Development of Andalusia. Special thanks are due to Jaime de Lara, Director-Conservator of the Natural Park of Sierra de María-Los Vélez. Finally, this work takes part of the general research lines promoted by the Agrifood Campus of International Excellence ceiA3, Spain (<http://www.ceia3.es/>).

REFERENCES

- Aguilar, F.J., Aguilar, M.A., Blanco, J.L., Nemmaoui, A., García-Lorca, A.M., 2016. Analysis and validation of grid DEM generation based on Gaussian markov random field. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* 41, 277–284. <https://doi.org/10.5194/isprsarchives-XLI-B2-277-2016>.
- Aguilar, F.J., Nemmaoui, A., Aguilar, M.A., Peñalver, A., 2021. Building Tree Allometry Relationships Based on TLS Point Clouds and Machine Learning Regression. *Appl. Sci.* 11, 10139. <https://doi.org/10.3390/APP112110139>.
- Aguilar, F.J., Nemmaoui, A., Aguilar, M.A., Peñalver, A., 2019a. Fusion of terrestrial laser scanning and RPAS image-based point clouds in Mediterranean forest inventories. *Dyna* 94, 131–136. <https://doi.org/10.6036/8892>.
- Aguilar, F. J., Rivas, J.R., Nemmaoui, A., Peñalver, A., Aguilar, M.A., 2019b. UAV-Based Digital Terrain Model Generation under Leaf-Off Conditions to Support Teak Plantations Inventories in Tropical Dry Forests. A Case of the Coastal Region of Ecuador. *Sensors* 19, 1934. <https://doi.org/10.3390/s19081934>.
- Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. *Int. Arch. Photogramm. Remote Sens.* 33 part B, 110–117. <https://doi.org/10.1016/j.isprsjprs.2005.10.005>.
- Bayat, M., Bettinger, P., Heidari, S., Khalyani, A.H., Jourgholami, M., Hamidi, S.K., 2020. Estimation of Tree Heights in an Uneven-Aged, Mixed Forest in Northern Iran Using Artificial Intelligence and Empirical Models. *Forests* 11(3), 324. <https://doi.org/10.3390/F11030324>.
- Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrizar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Péliissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., Vieilledent, G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Chang. Biol.* 20, 3177–3190. <https://doi.org/10.1111/gcb.12629>.
- Chen, J., Yang, H., Man, R., Wang, W., Sharma, M., Peng, C., Parton, J., Zhu, H., Deng, Z., 2020. Using machine learning to synthesize spatiotemporal data for modelling DBH-height and DBH-height-age relationships in boreal forests. *For. Ecol. Manage.* 466, 118104. <https://doi.org/10.1016/J.FORECO.2020.118104>.
- Chen, S., Liang, D., Ying, B., Zhu, W., Zhou, G., Wang, Y., 2021. Assessment of an improved individual tree detection method based on local-maximum algorithm from unmanned aerial vehicle RGB imagery in overlapping canopy mountain forests. *Int. J. Remote Sens.* 42, 106–125. <https://doi.org/10.1080/01431161.2020.1809024>.
- Coomes, D.A., Dalponte, M., Jucker, T., Asner, G.P., Banin, L.F., Burslem, D.F.R.P., Lewis, S.L., Nilus, R., Phillips, O.L., Phua, M.H., Qie, L., 2017. Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. *Remote Sens. Environ.* 194, 77–88. <https://doi.org/10.1016/j.rse.2017.03.017>.
- Ercanlı, İ., 2020. Innovative deep learning artificial intelligence applications for predicting relationships between individual tree height and diameter at breast height. *For. Ecosyst.* 7(12), 1–18. <https://doi.org/10.1186/S40663-020-00226-3>.
- Espejo, A., Federici, S., Green, C., Amuchastegui, N., D’Annunzio, R., Balzter, H., Bholanath, P., Brack, C., Brewer, C., Birigazzi, L., Cabrera, E., Carter, S., Chand, N., Donoghue, D., Eggleston, S., Fitzgerald, N., Foody, G., Galindo, G., Goeking, S., Grassi, G., Held, A., Herold, M., Kleinn, C., Kurz, W., Lindquist, E., McRoberts, R., Mitchell, A., Næsset, E., Notman, E., Quegan, S., Rosenqvist, A., Roxburgh, S., Sannier, C., Scott, C., Stahl, G., Stehman, S., Tupua, V., Watt, P., Wilson, S., Woodcock, C., Wulder, M., 2020. Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and guidance from the Global Forest Observations Initiative, Edition 3.0, U.N. Food and Agriculture Organization. Rome, Italy.
- FAO, 2015. Global forest resources assessment 2015: Desk reference. Rome.
- FAO and Plan Bleu, 2018. State of Mediterranean Forests 2018. Rome and Plan Bleu, Marseille.
- Feliciísimo, A.M., 1994. Parametric statistical method for error detection in digital elevation models. *ISPRS J. Photogramm. Remote Sens.* 49, 29–33. [https://doi.org/10.1016/0924-2716\(94\)90044-2](https://doi.org/10.1016/0924-2716(94)90044-2).
- Filho, S.V.S. da C., Arce, J.E., Montaña, R.A.N.R., Pelissari, A.L., 2019. Configuração de algoritmos de aprendizado de máquina na modelagem florestal: um estudo de caso na modelagem da relação hipsométrica. *Ciência Florest.* 29, 1501–1515. <https://doi.org/10.5902/1980509828392>.
- Furukawa, Y., Ponce, J., 2010. Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 1362–1376. <https://doi.org/10.1109/TPAMI.2009.161>.
- Goodbody, T.R.H., Coops, N.C., Marshall, P.L., Tompalski, P., Crawford, P., 2017. Unmanned aerial systems for precision forest inventory purposes: A review and case study. *For. Chron.* 93, 71–81. <https://doi.org/10.5558/ffc2017-012>.
- Guerra-Hernández, J., González-Ferreiro, E., Sarmiento, A., Silva, J., Nunes, A., Correia, A.C., Fontes, L., Tomé, M., Díaz

- Varela, R., 2016. Using high resolution UAV imagery to estimate tree variables in Pinus pinea plantation in Portugal. *For. Syst.* 25, 5. <https://doi.org/10.5424/fs/2016252-08895>.
- Jensen, J., Mathews, A., 2016. Assessment of Image-Based Point Cloud Products to Generate a Bare Earth Surface and Estimate Canopy Heights in a Woodland Ecosystem. *Remote Sens.* 8, 50. <https://doi.org/10.3390/rs8010050>.
- Jucker, T., Caspersen, J., Chave, J., Antin, C., Barbier, N., Bongers, F., Dalponte, M., van Ewijk, K.Y., Forrester, D.I., Haeni, M., Higgins, S.L., Holdaway, R.J., Iida, Y., Lorimer, C., Marshall, P.L., Momo, S., Moncrieff, G.R., Ploton, P., Poorter, L., Rahman, K.A., Schlund, M., Sonké, B., Sterck, F.J., Trugman, A.T., Usoltsev, V.A., Vanderwel, M.C., Waldner, P., Wedeux, B.M.M., Wirth, C., Wöll, H., Woods, M., Xiang, W., Zimmermann, N.E., Coomes, D.A., 2017. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. *Glob. Chang. Biol.* 23, 177–190. <https://doi.org/10.1111/gcb.13388>.
- Lisein, J., Pierrot-Deseilligny, M., Bonnet, S., Lejeune, P., 2013. A Photogrammetric Workflow for the Creation of a Forest Canopy Height Model from Small Unmanned Aerial System Imagery. *Forests* 4, 922–944. <https://doi.org/10.3390/f4040922>.
- López-Serrano, F.R., García-Morote, A., Andrés-Abellán, M., Tendero, A., Del Cerro, A., 2005. Site and weather effects in allometries: A simple approach to climate change effect on pines. *For. Ecol. Manage.* 215, 251–270. <https://doi.org/10.1016/J.FORECO.2005.05.014>.
- Luo, M., Wang, Y., Xie, Y., Zhou, L., Qiao, J., Qiu, S., Sun, Y., 2021. Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests* 12, 216. <https://doi.org/10.3390/f12020216>.
- Nogueira, D.F., Rico, E.C., 2017. Cambios en los usos de suelo en la Península Ibérica: Un meta-análisis para el período, 1985-2015. *Biblio3W Rev. Bibliográfica Geogr. y Ciencias Soc.* 22, 1215. <https://doi.org/10.1344/B3W.0.2017.26442>.
- Panagiotidis, D., Abdollahnejad, A., Surový, P., Chiteculo, V., 2017. Determining tree height and crown diameter from high-resolution UAV imagery. *Int. J. Remote Sens.* 38, 2392–2410. <https://doi.org/10.1080/01431161.2016.1264028>.
- Pascual, A., Guerra-Hernández, J., Cosenza, D.N., Sandoval, V., 2020. The Role of Improved Ground Positioning and Forest Structural Complexity When Performing Forest Inventory Using Airborne Laser Scanning. *Remote Sens.* 2020, Vol. 12, Page 413 12, 413. <https://doi.org/10.3390/RS12030413>.
- Peñalver, A., Aguilar, F.J., Nemmaoui, A., Rivas, J.R., Triana, Á.A., Aguilar, M.A., Llanderal, A., 2021. Precisión y eficiencia del inventario de plantaciones de teca en Ecuador mediante escáner láser terrestre. *Madera y Bosques* 27(1), e2712097. <https://doi.org/10.21829/MYB.2021.2712097>.
- Shimizu, K., Nishizono, T., Kitahara, F., Fukumoto, K., Saito, H., 2022. Integrating terrestrial laser scanning and unmanned aerial vehicle photogrammetry to estimate individual tree attributes in managed coniferous forests in Japan. *Int. J. Appl. Earth Obs. Geoinf.* 106, 102658. <https://doi.org/10.1016/J.JAG.2021.102658>.
- Tinkham, W.T., Swayze, N.C., 2021. Influence of Agisoft Metashape Parameters on UAS Structure from Motion Individual Tree Detection from Canopy Height Models. *Forests* 12, 250. <https://doi.org/10.3390/F12020250>.
- Vieira, G., de Mendonça, A., da Silva, G., Zanetti, S., da Silva, M., Dos Santos, A., 2018. Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. *Sci. Total Environ.* 619–620, 1473–1481. <https://doi.org/10.1016/J.SCITOTENV.2017.11.138>.
- Vilà-Cabrera, A., Coll, L., Martínez-Vilalta, J., Retana, J., 2018. Forest management for adaptation to climate change in the Mediterranean basin: A synthesis of evidence. *For. Ecol. Manage.* 407, 16–22. <https://doi.org/10.1016/J.FORECO.2017.10.021>.
- Wallace, L., Lucieer, A., Malenovsky, Z., Turner, D., Vopěnka, P., 2016. Assessment of forest structure using two UAV techniques: A comparison of airborne laser scanning and structure from motion (SfM) point clouds. *Forests* 7, 62. <https://doi.org/10.3390/f7030062>.
- White, J., Wulder, M., Vastaranta, M., Coops, N., Pitt, D., Woods, M., 2013. The Utility of Image-Based Point Clouds for Forest Inventory: A Comparison with Airborne Laser Scanning. *Forests* 4, 518–536. <https://doi.org/10.3390/f4030518>.
- 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. *Can. J. Remote Sens.* 42, 619–641. <https://doi.org/10.1080/07038992.2016.1207484>.
- Zhang, Y., Ma, J., Liang, S., Li, X., Li, M., 2020. An Evaluation of Eight Machine Learning Regression Algorithms for Forest Aboveground Biomass Estimation from Multiple Satellite Data Products. *Remote Sens.* 12, 4015. <https://doi.org/10.3390/rs12244015>.