Improving Aboveground Forest Biomass Maps: From High-Resolution to National Scale

Pilar Durante 1,2,*, Santiago Martín-Alcón 1, Assu Gil-Tena 1, Nur Algeet 1, José Luis Tomé 1, Laura Recuero 3, Alicia Palacios-Orueta 3 and Cecilio Oyonarte 2,4

1 Agresta Sociedad Cooperativa, 28012 Madrid, Spain; smalcon@agresta.org (S.M.-A.); agil@agresta.org (A.G.-T.); nalgeet@agresta.org (N.A.); jltome@agresta.org (J.L.T.)
2 Departamento de Agronomía, Universidad de Almería, 04120 Almería, Spain; coyonart@ual.es
3 Departamento de Sistemas y Recursos Naturales, ETSIMFMN, Universidad Politécnica de Madrid, 28040 Madrid, Spain; laura.recuero@upm.es (L.R.); alicia.palacios@upm.es (A.P.-O.)
4 Centro Andaluz para la Evaluación y Seguimiento del Cambio Global (CAESCG), 04120 Almería, Spain

* Correspondence: pdurante@agresta.org; Tel.: +34-969243064

Received: 26 January 2019; Accepted: 31 March 2019; Published: 3 April 2019

Abstract: Forest aboveground biomass (AGB) estimation over large extents and high temporal resolution is crucial in managing Mediterranean forest ecosystems, which have been predicted to be very sensitive to climate change effects. Although many modeling procedures have been tested to assess forest AGB, most of them cover small areas and attain high accuracy in evaluations that are difficult to update and extrapolate without large uncertainties. In this study, focusing on the Region of Murcia in Spain (11,313 km²), we integrated forest AGB estimations, obtained from high-precision airborne laser scanning (ALS) data calibrated with plot-level ground-based measures and bio-geophysical spectral variables (eight different indices derived from MODIS computed at different temporal resolutions), as well as topographic factors as predictors. We used a quantile regression forest (QRF) to spatially predict biomass and the associated uncertainty. The fitted model produced a satisfactory performance ($R^2$ 0.71 and RMSE 9.99 t·ha$^{-1}$) with the normalized difference vegetation index (NDVI) as the main vegetation index, in combination with topographic variables as environmental drivers. An independent validation carried out over the final predicted biomass map showed a satisfactory statistically-robust model ($R^2$ 0.70 and RMSE 10.25 t·ha$^{-1}$), confirming its applicability at coarser resolutions.

Keywords: mediterranean forest; climate change; ALS; MODIS; quantile regression forest; uncertainty

1. Introduction

The Mediterranean basin represents a hotspot of biological diversity, being a socioecological system very sensitive to climate change effects [1]. A keystone concern is how to develop adaptive planning for the assessment, monitoring, and management of organic carbon in ecosystems, which is essential to achieve climate change commitments at a national level. Particularly in terrestrial ecosystems, forests play a basic role as carbon sinks containing about 80% of global terrestrial aboveground biomass (AGB). However, such information is difficult to produce, and the uncertainties about magnitude and location are often large [2,3]. Much of this uncertainty is due to the lack of detailed information about the spatial distribution of carbon stored in biomass. Thus, forest biomass is an important measure for environmental management to provide more insights into the amount and spatial distribution of carbon storage for supporting future climate change mitigation actions [4].

The use of light detection and ranging (LiDAR) technology aimed at studying forested ecosystems started in the 1970s, especially airborne laser scanning (ALS) systems. Airborne laser-based acquisitions...
have been extensively used in forestry inventories following area-based approaches, such as in AGB mapping [5–7]. This is due to ALS’s high sensitivity to structural features [8–10] and high precision ability to measure stand parameters in three dimensions previously calibrated with plot-level ground-based measures [11,12]. Moreover, several previous studies have proven the use of existing data sets, particularly national forest inventory plots, to be a feasible option to provide training data for stand-level ALS inventories (e.g., [13,14]). ALS data combined with machine learning modeling techniques are able to find substantial accuracy in forest biomass estimates at local scale [15,16]. Applications linking local and national—even global—scales allow the design of dynamic assessment and management planning, therefore increasing the accuracy and predictive capacity of global models predicting forest AGB. Despite technological advancement, the high cost of acquisition of ALS data results in much higher revisit time compared to satellite data, which reduces their continuous temporal and spatial availability, therefore making their use at national scale inoperative. Thus, ongoing efforts are needed to solve the challenges involved in integrating environmental assessment and management at concordant scales in order to address climate change at a national level [17].

In the last few years, different studies have integrated optical satellite data into different spatial and temporal resolutions for land surface monitoring in a highly operative way [18]. Satellite remote sensing technology is considered as the most effective approach to estimate biomass, since it provides adequate methods to produce a high-temporal characterization of vegetation attributes over spatially continuous large areas [19]. The information provided by the moderate resolution imaging spectroradiometer (MODIS) has been extensively applied to monitor forests at regional and large scales since 2010, therefore enabling the derivation of biophysical parameters which are highly valuable for forest management such as forest biomass [20,21]. MODIS offers a short revisit time, and the information acquired in several spectral bands can be summarized in different vegetation indices (VIs) that can be used as proxies of photosynthetic activity of vegetation [22]. The theoretical basis for empirical-based VIs is derived from typical spectral reflectance signatures of contrast between blue-red and near-infrared wavelengths’ response. This contrast can be quantified through the use of ratios, differences, weighted differences or linear band combinations [23]. Hence, a MODIS-based hypertemporal series of VIs allow to extract valuable information about vegetation phenology [24] and to estimate vegetation biomass [25]. On the one hand, the normalized difference vegetation index (NDVI) [26] has been widely used in many forestry applications. The NDVI is the ratio of contrasting reflectance between maximum absorption of red wavelength due to chlorophyll pigments and maximum reflectance of infrared wavelength owing to leaf cellular structure [27,28]. These characteristics have been highly correlated with green biomass, green leaf area index (LAI), and percent green vegetation cover [3,29]. Since forest biomass is conditioned by temporal dynamics, the annual profile of the NDVI time series can be considered as an indicator of seasonal dynamics of forest biomass. Furthermore, NDVI annual mean (as net primary production estimator) and its variation coefficient (as indicator of vegetation seasonality) are commonly used in forest biomass assessment [30]. On the other hand, NDVI shows saturation in dense and multilayered forest canopy [31]. Therefore, substantial efforts have been made to improve and develop new vegetation indices, such as the enhanced vegetation index (EVI), which is considered a modified NDVI. EVI improves sensitivity across regions with high biomass and shows a high vegetation monitoring capability through a decoupling of the canopy background signal with a reduction in atmospheric influences [32]. The soil adjusted vegetation index (SAVI) [33] was generated to minimize the soil reflectance effect. Additionally, Reference [34] noted that NDVI is limited by its nonlinear response, making it insensitive to differences at very low and high densities. Hence, other vegetation indices are considered more linearly related to biophysical parameters of vegetation, such as the renormalized difference vegetation index (RDVI) [35], which was developed to take combined advantage of the difference vegetation index (DVI = NIR – Red) [36] and the NDVI; and the modified simple ratio (MSR) [37], proposed over the RDVI because it shows more sensitivity to various biophysical parameters.

In combination with dynamic and easily updated remotely-sensed VIs, terrain variables have also been used in many studies predicting AGB as supplemental information aimed at increasing
predictive power and statistical discrimination [38]. Since topography controls the main landscape processes related to structural characteristics of vegetation, such as water distribution and potential solar radiation [39], the digital elevation model (DEM) and the derivation of DEM-based predictor variables partially explain spatial patterns in AGB [40].

On the basis of the above, the use of high-resolution ALS data is a widely proven tool to collect and subsequently characterize structural forest stand attributes, with some limitations due to cost, revisit time, logistics, and data volumes involved in monitoring and mapping large areas [41]. Optical remote sensing data provide abundant and highly frequent observations to monitor extensive forest areas. However, coarse-resolution satellite data are often limited by the ground-based validation and nature of the measures, which can be largely insensitive to vertically distributed attributes [42].

The main hypothesis of this study is that there is a robust correlation between ALS-based forest AGB and VIs derived from MODIS data. Therefore, forest AGB spatial distribution can be estimated in those places where ALS information is not available. This study aimed to develop a methodological approach to improve nationwide assessment and mapping of AGB through combining high-resolution ALS data with hypertemporal and moderate-resolution satellite images in Mediterranean forests. A two-stage upscaling approach was applied where ALS-based forest AGB assessments, calibrated with plot-level ground-based measures, were used to train a machine learning method (quantile regression forest). Predictor variables consisted of MODIS-derived indices and topographic factors. As an approach of the estimation accuracy in the upscaling processes, the estimated variance of the AGB estimator was assessed according to Reference [43]. This methodology was mainly developed with open source software and always considering free available data.

2. Materials and Methods

2.1. Study Area

Our study case was a Mediterranean region in Southeastern Iberian Peninsula (Region of Murcia), which is a particularly vulnerable area to global climate change threats due to its environmental characteristics [1]. The area is dominated by a dry, semiarid climate characterized by warm mean annual temperatures (ca. 18 °C), scarce annual rainfall (ca. 300–350 mm), and high mean annual evapotranspiration (ca. 900 mm) [44]. Climate variation is related to a diverse orography, which combines mountainous areas (up to 2000 m asl), high plateaus (500–1000 m), and badlands. The lithology is mainly represented by Calcisols, Leptosols, Regosols, and Fluvisols.

The combination of these environmental conditions shapes the land uses. Most of the Region of Murcia (11,313 km²) is highly influenced by human activity and, therefore, it is occupied by cultivated land. Fragmented forest areas represent approximately 45% of the study area. The wooded area is mostly composed of pine forest, Pinus halepensis Mill being the main species (ca. 90%) along with Pinus nigra Arn. salzmanii and Pinus pinaster Ait, and holm oak (Quercus ilex L) to a lesser extent. That is to say, all the main species in this forest community are perennial. The shrubland area covers about of 60% of the forested area in the Region of Murcia.

2.2. From Field Plots to ALS Benchmark Map

2.2.1. Field Data

ALS plot-level biomass models were calibrated from the field plot database of the Fourth Spanish National Forest Inventory (SNFI) measured in 2010 in the Region of Murcia [45]. The SNFI is an extensive systematic survey carried out every 10 years on Spanish-forested areas on a basis of a regular network of 25 m radius plots at a density of 1 plot per km. From the complete database of the SNFI in Murcia (i.e., 1284 SNFI plots), a training dataset (i.e., 242 SNFI plots) was selected according to a stratified random sampling done within six predefined strata that covered the entire range of growing stock volume measures.
The main difficulty when using plots from the SNFI in combination with ALS data is the lack of precision in the plot center coordinates. Global navigation satellite system (GNSS) devices used in SNFI data collection, as other studies showed [46,47], have a nominal accuracy of 5–15 m. Following the methodology and the information sources proposed by Reference [48], the selected plots were relocated manually to improve the accuracy of the plot center coordinates. In this case, plots were shifted an average of 12.5 m in the relocation procedure.

The biomass model was based on four different information sources: (i) 2 m resolution ALS canopy models, (ii) sketches of plot location made by the field teams, (iii) high resolution orthoimagery from the Spanish National Plan for Aerial Orthophotography (PNOA in Spanish), and (iv) total height, species, and location of each measured tree in the field. The plot center relocation was intended to minimize possible errors derived from this lack of precision in the positioning of the plots. The SNFI plot size varies depending on tree diameter at breast height (dbh), that is to say, circular plots with four concentric subplots of a 5, 10, 15, and 25 m radius where trees with a dbh larger than 75, 125, 225, and 425 mm were respectively measured. In the SNFI, percent cover and mean height (in dm) were only estimated for each shrub species in the first plot radius (5 m).

In ALS-based forest inventories, the individual cell size of a grid that covers the entire study area has to be equal to the field plot size [49]. In this study, the selected SNFI plots were converted to plots with a maximum radius of 14.1 m (625 m$^2$), in which we calculated forest AGB and extracted the ALS-derived metrics. Thus, only trees whose distance to the center of the SNFI plot was less than 14.1 m were taken into account to fit a maximum plot area equivalent to a 25 m $\times$ 25 m pixel. Therefore, the plots remained with only three concentric subplots of 5, 10, and 14.1 m. We calculated AGB for each selected SNFI plot as the sum of forest AGB related to each shrub species and single tree measured in these three subplots. AGB was scaled up to the hectare level according to the area of each subplot. Tree AGB was calculated through species-specific general models, reporting a goodness-of-fit (adj-R$^2$) that was always higher than 0.95, as a function of dbh and total height [50]. To estimate shrub biomass, we applied a general model for Mediterranean shrublands (adj-R$^2$: 0.63) that considers the shrub cover and its mean height as variables [51]. Grassland was excluded from the estimation due to ALS data inaccuracy to predict attributes related to this cover type.

Since SNFI data were collected over forested areas exclusively (i.e., forests and shrublands), we applied a forest mask to the study area based on the national forest map. The Spanish forest map (SFM) provides a detailed and updated version for the Region of Murcia at 1:25000 [45]. The SFM is the base cartography for the SNFI and describes the current distribution and seral stage of forest cover, based on structural and ecological descriptive items.

### 2.2.2. ALS Data Acquisition and Processing

ALS data were provided by PNOA. Data from two flight campaigns were used. A recent study demonstrated temporal transferability of pine forest attributes modeling by means of using low-density ALS data, even if data captured with different ALS instruments were considered [52]. In our case, data captured in the campaign of 2009 (one year before SNFI plot survey) were used to fit the aboveground biomass ALS model. ALS data captured in 2016–2017 were employed to implement the model and to obtain an updated prediction of forest biomass. Data from 2009 were captured from October to December 2009 using a Leica ALS50 system (Leica Geosystems AG, Heerbrugg, Switzerland) with a mean density of 0.5 pulse m$^{-2}$ and vertical RMSE $\leq$ 0.20. Data from 2016–2017 were captured from August 2016 to March 2017 using a Leica ALS60 system with a mean density of 0.7 pulse m$^{-2}$ and vertical RMSE $\leq$ 0.08. In both cases, ALS point cloud data were acquired in 2 $\times$ 2 km tiles covering the whole region, which were processed using the FUSION software [53].

Firstly, a 2 m resolution DEM was generated to subtract the orthometric elevation of the DEM from the z-coordinate of each ALS return aimed at normalizing the ALS point cloud. This procedure was applied separately to the ALS point cloud from each flight. Then, a normalized ALS point cloud of 2009 was used to compute a total of 15 metrics (Table 1) for each field plot using a threshold height of 2 m to separate trees
from understory vegetation [54]. Amongst these considered metrics, tree canopy cover (TCC) refers to the ratio between the number of first returns above 2 m and the total number of first returns, whereas the canopy relief ratio (CRR) describes the relative canopy shape from altimetric observation [55]. CRR reflects the degree to which canopy surfaces are in the upper (>0.5) or lower (<0.5) portions of the height range [56]. Two other metrics of normalized ALS point cloud for each field plot were computed through filtering from the point cloud those points with a height lower than 2 m and using a height threshold of 0.5 m to separate shrubs from bare soil or herbaceous vegetation. These metrics correspond to mean height of the low stratum (Hmean_LS) and canopy cover of the low stratum (TCC_LS; i.e., the ratio between the number of first returns above 0.5 m and the total number of first returns).

Table 1. Summary statistics of the response variable and the airborne laser scanning (ALS) metrics used for modeling in the set of Fourth Spanish National Forest Inventory (SNFI) plots (n = 242; 14.1 m from the plot center).

<table>
<thead>
<tr>
<th>Group</th>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>Aboveground forest biomass (t·ha⁻¹)</td>
<td>6.88</td>
<td>143.51</td>
<td>57.20</td>
<td>28.03</td>
</tr>
<tr>
<td></td>
<td>Hmean</td>
<td>2.75</td>
<td>13.34</td>
<td>6.38</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>Hsd</td>
<td>0.55</td>
<td>4.86</td>
<td>1.99</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Hvar</td>
<td>0.30</td>
<td>23.61</td>
<td>4.49</td>
<td>3.55</td>
</tr>
<tr>
<td></td>
<td>Hcv</td>
<td>0.16</td>
<td>0.50</td>
<td>0.31</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Hiq</td>
<td>0.76</td>
<td>8.13</td>
<td>2.90</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>Hkur</td>
<td>1.43</td>
<td>4.77</td>
<td>2.51</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Hp01</td>
<td>2.02</td>
<td>5.00</td>
<td>2.39</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Hp05</td>
<td>2.04</td>
<td>6.97</td>
<td>3.07</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Hp10</td>
<td>2.06</td>
<td>8.06</td>
<td>3.66</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>Hp25</td>
<td>2.27</td>
<td>10.94</td>
<td>4.93</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td>Hp50</td>
<td>2.48</td>
<td>14.00</td>
<td>6.44</td>
<td>2.17</td>
</tr>
<tr>
<td></td>
<td>Hp75</td>
<td>3.10</td>
<td>17.64</td>
<td>7.83</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td>Hp90</td>
<td>3.51</td>
<td>19.24</td>
<td>8.94</td>
<td>2.67</td>
</tr>
<tr>
<td></td>
<td>Hp95</td>
<td>3.77</td>
<td>19.59</td>
<td>9.50</td>
<td>2.78</td>
</tr>
<tr>
<td></td>
<td>Hp99</td>
<td>4.24</td>
<td>20.15</td>
<td>10.25</td>
<td>2.94</td>
</tr>
<tr>
<td></td>
<td>CRR</td>
<td>0.29</td>
<td>0.73</td>
<td>0.48</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>TCC</td>
<td>1.88</td>
<td>93.86</td>
<td>44.55</td>
<td>19.67</td>
</tr>
<tr>
<td>*ALS metrics</td>
<td>Hmean_LS (low stratum) = mean height of the low stratum (total returns above 0.5 m and below 2 m), and TCC_LS = canopy cover of the low stratum.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, the normalized ALS point cloud from 2016–2017 was used to compute the same set of metrics for the forest area in the entire region in 25 × 25 m pixels. We used this pixel resolution to reduce the impact on forest biomass scaling-up estimation and because this resolution is compatible with SNFI plot size.

2.2.3. ALS Benchmark Map: Model, Spatial Prediction, and Resampling

The nonparametric regression algorithm random forest [57] was used to model the relationship between ALS-derived metrics with the ground-based measure of forest AGB computed from the SNFI plots (14.1 m radius). Analyses were carried out using the packages randomforest [58], VSURF [59], and rfUtilities [60] implemented in the R statistical software [61]. The VSURF package was applied to identify the combination of ALS metrics that best predicted forest AGB. This package provides an automated method to select predictor variables based on their importance scores while minimizing the redundancy among them. The out-of-bag error statistic provided in RF reports the goodness of model fit but not necessarily the predictive performance. For this reason, an independent data withholding of 10% was performed for a more precise model validation. A bootstrapping validation technique was followed by randomizing the independent data and executing the RF model a thousand times. Withheld data were predicted at each replicate. This cross-validation procedure was used by applying the random forest regression model cross-validation tool implemented in the rfUtilities
package. Cross-validated estimates of root-mean-square error (RMSE) were calculated between the observed and predicted values using the training data. RMSE was normalized by the observations mean (RMSE%). The median percentage of explained variation (pseudo-R²) and RMSE% were reported for 1000 cross-validations.

The selected model was applied to the ALS metrics computed over the 2016–2017 point cloud covering forest area in 25 × 25 m pixel resolution. Therefore, a continuous 25 m ALS map of forest AGB in tons per hectare (t·ha⁻¹) was obtained. The resulting spatial cover was resampled to 250 m moderate resolution (MODIS variables resolution, see next subsection) using the bilinear method implemented in the resample package in R [62]. This ALS-based AGB map at 250 m pixel resolution was used as the forest AGB benchmark of the study area, henceforth ALS benchmark map.

2.3. From ALS Benchmark Map to MODIS Map

2.3.1. Optical Remote Sensing and Topographic Variables

A stack based on 1060 environmental predictors was generated according to dynamic and static variables. The dynamic variables refer to 8 VIs calculated using the newly release collection 6 of MOD09Q1 [63] and MOD13Q1 [23] MODIS products. These products respectively consist of 8-day and 16-day composites, both at 250 meters spatial resolution and corresponding to the closest dates to the 2016–2017 ALS data acquisition. Therefore, we obtained data for the years 2015, 2016, and 2017 from the Land Processes Distributed Active Archive Center (LP DAAC) [64]. Images were re-projected to the UTM zone 30, datum WGS-84 using the “MODIS Reprojection Tool” (MRT), and quality flags were decoded using the “Land Data Operational Products Evaluation” tool (LDOPE). Both tools were provided by the LP DAAC.

Several vegetation spectral indices were proposed as potential AGB estimators. We used red and near-infrared reflectance bands provided by the MOD09Q1 product to calculate the NDVI, MSR, DVI, RDVI, SAVI, and MSAVI, whereas near-infrared and short infrared reflectance values were used to calculate NDI7 (see equations in Table 2). Time series were filtered in order to avoid abnormal values caused by clouds and certain atmospheric conditions. Low-quality pixel values were removed based on MODIS quality data bands and replaced by the average of the previous and subsequent dates values with good quality in the time series. A smooth filter was also applied to remove outliers that fall outside the mean plus/minus two times the standard deviation within a five-date period window. The anomalous values were also replaced by the average of the previous and subsequent date values in the time series. In addition, time series from the MOD09Q1 product were smoothed with a Savitzky–Golay filter using a window width of 11 observations.

Table 2. Equations of vegetation indices derived from moderate resolution imaging spectroradiometer (MODIS) data.

<table>
<thead>
<tr>
<th>Vegetation Index</th>
<th>Equation ¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized difference vegetation index (NDVI)</td>
<td>( NDVI = \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}}} ) [26]</td>
</tr>
<tr>
<td>Enhanced Vegetation Index (EVI)</td>
<td>( EVI = G \cdot \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}} - \rho_{\text{BLUE}}} - L ) [32]</td>
</tr>
<tr>
<td>Normalized Difference Index 7 (NDI7)</td>
<td>( NDI7 = \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}}} + 1 + L ) [65]</td>
</tr>
<tr>
<td>Modified Simple Ratio (MSR)</td>
<td>( MSR = \sqrt{\frac{\rho_{\text{NIR}}}{\rho_{\text{RED}}} - 1} ) [37]</td>
</tr>
<tr>
<td>Difference vegetation index (DVI)</td>
<td>( DVI = \rho_{\text{NIR}} - \rho_{\text{RED}} ) [36]</td>
</tr>
<tr>
<td>Renormalized difference vegetation index (RDVI)</td>
<td>( RDVI = \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}} + \rho_{\text{BLUE}}} ) [35]</td>
</tr>
<tr>
<td>Soil adjusted vegetation index (SAVI)</td>
<td>( SAVI = \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}} + L} ) ( L = 0.5 ) [33]</td>
</tr>
<tr>
<td>Modified soil adjusted vegetation index (MSAVI)</td>
<td>( MSAVI = 0.5 \left( 2\rho_{\text{NIR}} + 2 - \sqrt{(2\rho_{\text{NIR}} + 1)^2 - 8(\rho_{\text{NIR}} - \rho_{\text{RED}})} \right) ) [66]</td>
</tr>
</tbody>
</table>

¹ \( \rho_{\text{RED}}, \rho_{\text{NIR}} \) and \( \rho_{\text{SWIR3}} \) respectively correspond to the reflectance in the red (620–670 nm), near-infrared (841–876 nm), and shortwave infrared (2105–2155 nm) wavelengths. In SAVI, the accommodation coefficient of the soil background is \( L = 0.5 \), related to moderate canopy cover areas. In EVI, the canopy background adjustment for correcting the nonlinear is \( L = 1 \); the coefficients of the aerosol resistance term—which use the blue band to correct for aerosol influences in the red band—are \( C1 = 6 \) and \( C2 = 7.5 \), and the gain or scaling factor is \( G = 2.5 \).
Reflectance bands were compiled, and indices were computed using the Environment for Visualizing Images (ENVI 4.5) software.

The properties of the VIs time-series can be summarized as ecologically relevant measures in a variety of indices [67]. Different variables derived from MODIS VIs time series were estimated to characterize seasonal dynamics of forest biomass in different time frames. To obtain the annual mean profile of each VIs, the 8-day and 16-day composites were layer-stacked to respectively create a 46- and a 23-band image for each year. We calculated the annual profile as the mean corresponding values of the three years considered (2015–2017).

In order to assess the maximum VIs values related to the growing season, for the 8-day composite of the 2015–2017 time period, we generated aggregates for the computed VIs (Table 2) time series using the maximum-value composite (MVC) approach [68]. The assumption behind the MVC technique is that the maximum NDVI value of a set of images will correspond to the ideal atmospheric and vegetation conditions. For the remaining VIs, the value corresponding to the date of the maximum NDVI value was selected.

As long-term ecosystem functioning descriptors, we calculated two variables from the 2001–2016 NDVI time series of the MOD13Q1 product: NDVI mean, as annual primary production prediction factor, and its standard deviation (NDVI sd) as the coefficient of seasonal variation, the latter being an indicator of vegetation seasonality [31]. Both variables were calculated according to the method described in [30].

According to [39,69], a local digital elevation model (DEM) was used to derive the static variables describing the topography to inform models with locally relevant information on landscape position. These static variables were 17 topographic parameters (Table 3) derived from a 25 m DEM, which is available at the Centre of Spanish National Geographic Institute (IGN). The current DEM version was generated by the IGN from ALS data interpolation. We used the SAGA GIS software to derive the parameters, which were included in the terrain analyst functions [70]. Previously to the terrain parameters computation, the 25 m DEM was aggregated to 250 m through the same method used for the ALS benchmark map.

<table>
<thead>
<tr>
<th>Table 3. Initial covariates used in the aboveground biomass predictive model from the airborne laser scanning (ALS) benchmark map to the MODIS map.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Covariate</strong></td>
</tr>
<tr>
<td>Dynamic (^\text{1})</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Static</td>
</tr>
</tbody>
</table>

\(^{1}\) Normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), normalized difference index 7 (NDI\(^{7}\)), modified simple ratio (MSR), difference vegetation index (DVI), renormalized difference vegetation index (RDVI), soil adjusted vegetation index (SAVI), modified soil adjusted vegetation index (MSAVI).
All the dynamic and static variables were stacked to build the covariate space which matches the same projection, extent, and pixel size. All the statistical and geo-information analyses in this section were performed using the R software.

2.3.2. NDVI Annual Profile per Vegetation Type

Before correlating the ALS-based forest AGB assessment and MODIS-derived indices, we carried out a comparison among vegetation type profiles aimed at characterizing the influence of seasonal dynamics of the understory captured by satellite images on multilayer covers. This was performed mainly because grassland was excluded from ALS biomass estimation (see Section 2.2.1). Due to the relationship of the NDVI annual profile with green biomass [27,30], we generated the annual mean profile of an NDVI 8-day composite per vegetation type in the study area. Pixels of forest areas were aggregated into the main homogeneous vegetation formations at different percentages of canopy cover to describe their seasonal patterns. We classified the main vegetation in three cover types based on the SFM information: dense tree cover, mixed tree and shrub/herbs cover, and dense shrub/herbs cover. Each cover type was divided according to the vegetation species and canopy cover, which resulted in a set of 17 classes. To obtain the annual mean profile, the 8-day MODIS composite was layer-stacked to create a 46-band image for each year. We calculated the annual profile as the mean of the corresponding values of the three years considered (2015–2017).

2.3.3. MODIS Map: Spatial Predictive Model and Associated Uncertainty

The pixel values of the ALS benchmark map (250 m resolution) were considered as forest AGB synthetic data representative of the surface covered by the corresponding MODIS pixel. A systematic sample of pixels (39,647 pixels) was selected from the ALS benchmark map, increasing the distribution and sample size of the target variable (forest AGB). Covariate values were then extracted at each location from the covariate stack (Figure 1). The most explicative and low correlated predictors were selected through the VSURF package implemented in the R software [59] in combination with a low variance inflation factor (VIF, car package in R). Prior to the calculation of the VIF criterion, a GLM model was generated through the car package algorithm in R [71].

![Figure 1](image-url) **Figure 1.** Methodological scheme of the target and predictor variables extraction for modeling the final moderate-resolution aboveground biomass.
A machine learning algorithm was applied to model the relationship between the target variable (AGB obtained from the ALS benchmark map) and the covariates (selected MODIS derived indices and topographic factors). Due to the performance problems of RF with high dimensional data [72,73] and the need to conduct pixel-based uncertainty estimation, the extracted pixels of the ALS benchmark map were used to train a quantile regression forest (QRF). QRF is a generalization of RF used to estimate an accurate approximation of the full conditional distribution of the response variable. The QRF infers conditional quantiles to build predicted intervals interpreted as a surrogate of uncertainty associated with the response variable for each pixel value. The quantregForest package [74] was applied in the R software to perform predictive and uncertainty maps. Out-of-bag predictions were used to evaluate the quality of the conditional quantile approximations. We computed 2 maps at 250 m resolution: the predicted values of forest AGB based on MODIS-derived index and topographic factors, henceforth MODIS map, and their associated uncertainty. An independent validation of the predicted MODIS map was performed by testing a random sample of 3000 pixels from the ALS benchmark map. The linear correlation between tested and observed data showed the information values of the final MODIS map validation.

In order to analyze the relationship between estimated values of the final MODIS map and their associated uncertainty, a regression line was plotted to visually depict the forest AGB data and their corresponding relative uncertainty at 250 m pixel resolution.

According to [43], ignoring field-to-LiDAR uncertainty can lead to underestimation of estimator variance in a hierarchical modeling process. To address the analysis of the estimation accuracy in the up-scaling process, a generalized hierarchical model-based (ghmb) estimation was therefore performed based on the previous publication (ghmb, HMB package in R). To make the ghmb model computationally affordable, the dimension of the covariate space has been reduced based on the covariate’s importance measures depicted in the random forest analysis (randomForest package in R). Moreover, to reduce the number of input data (n = 39,647), a representative zone of the study area was selected (Figure 2). The results of the ghmb model are the estimated population mean of AGB and its variance as proxies of the estimation accuracy in the upscaling processes. For the same area, the AGB mean based on ALS map and MODIS map was also assessed to compare with the ghmb model results.

![Figure 2. 25 m airborne laser scanning (ALS) map of aboveground biomass estimation (t ha\(^{-1}\)) in the Region of Murcia (Spain), which was calibrated with plot-level ground-based measures (14.1 m radius plots were represented as black dots and were based on the SNFI). The inner rectangle depicts the selected testing area to estimate the global estimator variance through ghmb.](image-url)
3. Results

3.1. ALS-Based Forest AGB Assessment

The VSURF procedure for variable selection has identified as the best combination of ALS-derived variables to predict forest AGB: mean height of total returns above 2 m (H\text{mean}), 25th percentile of total returns above 2 m (H\text{p25}), 50th percentile of total returns above 2 m (H\text{p50}), tree canopy cover (TCC), and canopy cover of the low stratum (TCC_LS). The RF predictive model was calibrated and subsequently used to predict forest AGB at the field plots which were withheld from the model training subset at each cross-validation bootstrap. The RF model fitted with these variables showed a satisfactory performance in terms of variance explained and RMSE\% (explained percentage variance: 69.09; median cross-validation RMSE\%: 27.18).

The 25 m ALS map (Figure 2) showed remarkable spatial heterogeneity with a substantial increase in high elevation areas mainly distributed along the northwestern to southeastern axis of the study area. According to the 25 m ALS map prediction, the maximum values corresponded to the most densely forested areas mainly located in the middle and the northwest parts of the study area. By contrast, the minimum values were associated to shrubland areas. In the 25 m ALS map, the mean AGB value attained in forested areas was 25.44 t·ha\(^{-1}\) and 21.7 t·ha\(^{-1}\) standard deviation (sd). In tree cover areas (i.e., excluding pure shrubland area), the mean value of AGB was 37.64 t·ha\(^{-1}\) (15.64 t·ha\(^{-1}\) sd). Since the bilinear interpolation method used to compute the resampling from 25 to 250 m pixel resolution assigns the weighted average of nearest neighboring cells to the output cell value, the biomass mean value decreased at this moderate spatial resolution. Therefore, forest AGB at the 250 m pixel resolution showed a lower mean value and more homogeneous deviation at this coarser spatial resolution: 20.56 t·ha\(^{-1}\) (18.57 t·ha\(^{-1}\) sd). This resampled ALS biomass map was referred to as the forest ALS benchmark map.

3.2. From ALS Benchmark Map to MODIS Map

3.2.1. NDVI Annual Mean Profile per Vegetation Type

The signal of forest cover types was extracted from MODIS hypertemporal series of VIs which allowed to analyze valuable information about the vegetation phenology. The differences in seasonal patterns characterized various types of vegetation based on their NDVI annual profile. The comparison of the mean annual profile (8-day composite of the 2015–2017 time series) among the main vegetation formations allowed a discernment of the influence of vegetation different from tree and shrubland depending on the cover density level. The different forest cover types showed a dominant signal signature comparable to the tree canopy cover reflectance pattern (Figure 3a) even in combination with other shrub or grass vegetation types (Figure 3b,c). Most of the mean seasonal NDVI profiles of the vegetation types were characterized by a bimodal pattern with two maximum values in April–May (0.43 max in pine cover) and December–January (0.47 max in pine cover). The growing seasons started in March/April and September according to spring and autumn in the European Mediterranean basin. The tree-shrubland mixed cover (Figure 3b) presented the same profile pattern (i.e., shape) with slightly lower values than the tree cover (Figure 3a). The mean annual NDVI profile of pure shrubland of different vegetation formations revealed greater differences regarding the others forest cover types (Figure 3c). As stated before, the pure grassland cover was excluded for biomass prediction.
3.2. From ALS Benchmark Map to MODIS Map

3.2.1. NDVI Annual Mean Profile per Vegetation Type

The signal of forest cover types was extracted from MODIS hypertemporal series of VIs which allowed to analyze valuable information about the vegetation phenology. The differences in seasonal patterns characterized various types of vegetation based on their NDVI annual profile. The comparison of the mean annual profile (8-day composite of the 2015–2017 time series) among the main vegetation formations allowed a discernment of the influence of vegetation different from tree and shrubland depending on the cover density level. The different forest cover types showed a dominant signal signature comparable to the tree canopy cover reflectance pattern (Figure 3a) even in combination with other shrub or grass vegetation types (Figure 3b,c). Most of the mean seasonal NDVI profiles of the vegetation types were characterized by a bimodal pattern with two maximum values in April–May (0.43 max in pine cover) and December–January (0.47 max in pine cover). The growing seasons started in March/April and September according to spring and autumn in the European Mediterranean basin. The tree-shrubland mixed cover (Figure 3b) presented the same profile pattern (i.e., shape) with slightly lower values than the tree cover (Figure 3a). The mean annual NDVI profile of pure shrubland of different vegetation formations revealed greater differences regarding the others forest cover types (Figure 3c). As stated before, the pure grassland cover was excluded for biomass prediction.

Figure 3. (a–c): NDVI annual mean profile of the 8-day MODIS composite of the 2015–2017 time series per main vegetation types and percentage of canopy cover (CC) based on Spanish forest map (SFM) information. DOY: day of the year.

3.2.2. Model Fitting, Spatial Prediction, and Uncertainty Assessment

The previous VSURF selection of the initial set of 1060 covariates aimed to fit the predictive model from ALS benchmark map values decreased the dimension of the covariate space to 35 (Table 4), after taking into account a low VIF criterion in order to minimize the statistical redundancy of the predictors. The dynamic predictive factors were associated to dates in April, early May, late August, September, and early October. NDVI was the main vegetation index at different time frames.

Table 4. Selected covariates in the fitted aboveground biomass predictive modeling of the airborne laser scanning (ALS) benchmark map.

<table>
<thead>
<tr>
<th>Description</th>
<th>Selected</th>
<th>Covariates ¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-day composite mean annual profile (2015–2017)</td>
<td>9/05, 21/08</td>
<td>NDVI, MSR</td>
</tr>
<tr>
<td>16-day composite mean annual profile (2015–2017)</td>
<td>9/05, 13/08</td>
<td>EVI, NDW7</td>
</tr>
<tr>
<td>Mean value and standard deviation (2001–2016)</td>
<td>2 indices</td>
<td>NDVImean, NDVIsn</td>
</tr>
<tr>
<td>Maximum value (2015–2017)</td>
<td>6 indices</td>
<td>NDVI, MSR, DVI, RDVI, SAVI, MSAVI</td>
</tr>
<tr>
<td>Topographic variables</td>
<td>17 parameters</td>
<td>DEM, and 16 derived parameters included in Terrain Analyst functions (SAGA GIS software)</td>
</tr>
</tbody>
</table>

¹ Described in Table 3.
Then, two models were obtained through QFR which corresponded to conditional median for forest AGB (MODIS map) and conditional standard deviation, the latter as a surrogate of uncertainty. Based on the determination coefficient ($R^2$) and the root-mean-square error (RMSE) as model information criteria, a substantial accuracy increase of the biomass estimates was observed in QRF models compared to GLM techniques previously generated to calculate the VIF. The latter showed an $R^2 = 0.52$ and RMSE = 12.9 t ha$^{-1}$ versus $R^2 = 0.71$ and RMSE = 9.97 t ha$^{-1}$ in the QRF model. The spatial interpolation of the two QRF models resulted into a 250 m spatial resolution map with two bands related to forest AGB estimate (MODIS map) and its associated uncertainty in each pixel (Figure 4).

3.2.3. Final Predicted MODIS Map and Validation

Figure 4 depicts the spatial variability of the forest AGB MODIS map in the study area and its associated uncertainty. The biomass values ranged from 0.1 to 88.34 t ha$^{-1}$ and presented a very similar spatial distribution to the 25 m ALS map (see Figure 2). The highest AGB values also corresponded to tree cover in the central and northwest part of the study area, and the minimum ones were associated to shrubland covers in the southeast. The uncertainty values ranged from 0.1 to 44.25 t ha$^{-1}$, 11.34 t ha$^{-1}$, and 4.44 t ha$^{-1}$ as mean and sd values, respectively. The relationship between the forest AGB of MODIS map predicted values and their relative uncertainty associated the highest values of uncertainty with the lowest values (dots within the red box in Figure 5) and vice versa.

The resulting MODIS map also showed a predicted mean value very similar to the 25 m ALS map and slightly higher than the ALS benchmark map at 250 m: 24.91 t ha$^{-1}$ (17.16 t ha$^{-1}$ sd). Considering the associated uncertainty of MODIS-based predicted values, the results revealed a mean value of 18.77 t ha$^{-1}$ (13.31 t ha$^{-1}$ sd). The external, independent validation test attained a satisfactory fitting of the predictive model, since it presented a reasonable $R^2$ and similar RMSE (0.703 and 10.25 t ha$^{-1}$, respectively).

According to the AGB estimator for the global two-stage upscaling approach, the 35 selected covariates for AGB MODIS map modeling were reduced to 4 covariates based on the variable
importance measures of the QRF model. These covariates were NDVI sd, 16-day NDVI composite of the 04/2016, 16-day EVI composite of the 08/2015 and DEM. The selected area for this analysis (Figure 2) covers an extent of 2773 km$^2$, and includes 72 field plots and a sample of 6474 pixels from the ALS benchmark map.

The ghmb model attained values of 41.63 t ha$^{-1}$ for the estimated AGB mean and 13.22 for the estimated variance with a 95% confidence interval of 34.51–48.76. Likewise, the AGB mean values were 34.72 t ha$^{-1}$ and 31.01 t ha$^{-1}$ for ALS and MODIS maps, respectively.

4. Discussion

4.1. Forest AGB Estimate from Local to National Scale

The assessment of forest aboveground biomass at the national extent and, subsequently, carbon stock is a future key challenge still unsolved in Spain with large implications related to climate change. The novelty of this study is to apply a two-stage upscaling process for AGB assessment over large extents using open data. Even though there is a time frequency inconsistence in ALS data in Spain, the provided methodology combines a sample of ALS data to bridge the scale gap between a sparse sample of field plots and satellite images.

The methodological framework presented herein provides a robust approach to map detailed, spatially continuous, and easy to update vegetation biomass from local to large scale. We demonstrated that forest attributes such as biomass can be modeled with relevant accuracy in large areas through optical remote sensing data as long as there is free availability of ALS data and NFI plots, such as in the case of Spain. This methodology has increasingly been employed [15] to estimate AGB across heterogeneous forested landscapes around the world [75]. Nevertheless, some authors such as that of [76] advised that LiDAR-based AGB assessment should be used with great care for further upscaling to satellite imagery, since they found lower uncertainties in NFI-calibrated biomass maps. The studies that considered the technique used in this study for feasible prediction of AGB refer to a reasonable density of training data to calibrate LiDAR data, short time lag data acquisition between NFI plots and LiDAR data, and a representative plot sample area related to LiDAR pixel resolution [38]. Once we addressed these considerations, the discrepancy between ALS biomass estimate and plot-level data
was significantly reduced (Table 1) and, therefore, the error propagation, too. We obtained a satisfactory performance of the fitted model in terms of prediction of independent test data (explained percentage variance: 69.09; median cross-validation RMSE%: 27.18). As a general comparison, we contrasted our results with the review conducted by the authors of [77] over other forestry studies, which focused on estimating growing stock volume/biomass/carbon. All the considered studies in the review used field data, mostly from local inventories of NFIs, and contrasting sources of remotely sensed data. To facilitate the comparison of the results accuracy assessment, the RMSE% (defined as RMSE normalized by the mean observed values) was the measure most commonly used and attained a median RMSE% of 37%, with a standard deviation of ±31.6% across these types of attributes. The studies that used only LiDAR-based feature variables (81 studies) revealed a median RMSE% of 31.3%, with a median reference set size of 124 plots (our study accounted for 242 plots). These studies confirmed that to obtain low RMSE%, large reference sets are necessary depending on the complexity and diversity of the forest. Therefore, our results of the AGB estimated based on the ALS model can be deemed as a promising result relative to these and national benchmarks [78].

Since the pixel values of the ALS benchmark map can be considered as forest AGB synthetic data, we integrated ALS pixel plots in order to increase the distribution and sample size of local data to model AGB at moderate resolution (250 m). As the authors of [41] claimed, this integration mitigates the cost of ground plot installation and offers spatially extensive and representative sampling of calibration and validation data to support the modeling of forest attributes, such as biomass canopy cover (in our study, more than 30,000 ALS plots and 3000 test data, respectively). However, the costs, limited revisited time and spatial continuity, logistics and data volumes involved in PNOA LiDAR data have fostered the combination of field NFI plot data and synthetic ALS dataset with MODIS imagery. As preliminary results, the extrapolation of AGB assessment from ALS to 250 m pixel resolution based on VIs and topographic variables (MODIS map) offered reasonably robust, accurate, and scalable results (R² 0.71 and RMSE 9.97 t ha⁻¹), with a satisfactory model fitting for the external, independent validation (similar R² and RMSE of 0.703 and 10.25 t ha⁻¹, respectively). Our MODIS model relative RMSE (48.47%) was consistent with other studies such as that of Reference [79], which revealed an RMSE% of 69% for aboveground tree biomass estimated by NFI ground plots and 250 m MODIS data. However, other studies, which a large amount of Landsat data, provided lower RMSE% at regional scale [80].

The main limitation of measure in our model was due to a discrepancy between the ALS estimation capacity and the surface reflectance signal captured by MODIS in grasslands. ALS presents measurement constraints related to its operational mode with discrete returns. It only returns a few (up to 4) points per laser shot and, therefore, represents a potential limitation for the measurement of short understory structure [81]. In this sense, canopy attribute measurement was correlated with the MODIS VIs reflectance signal, which also captures variability in understory structure related to composition and configuration of herbaceous species [82]. Our analysis of the NDVI annual profile characterizing the main types of vegetation in the Region of Murcia (Figure 3) showed slight differences in the shapes of seasonal patterns between stands of pure tree and mixed with grass or shrubland cover (Figure 3a–c), in which the pattern of the tree canopy cover was mainly predominant. This may be partially due to an understory scattered cover or an intermediate to dense pine canopy cover, where understory reflectance can be neglected.

In addition, since SNFI plots mainly encompass forest tree areas, biomass in pure shrub areas was estimated without plot-level validation. Therefore, and according to studies carried out by the authors of [83,84], high biomass uncertainties may be found in modeling herbs/shrub areas in the MODIS map generated in this study due to low ground cover or low vegetation height. Even though shrubland biomass estimation was less accurate, the model performance analysis revealed the convenience of including it, thus showing higher final model accuracy. This can be supported by the “edge effect” of the input data aggregation on upscaling [85] created in fragmented forest landscapes such as in the Region of Murcia, where tree canopy cover presented narrower and discontinuous spatial patches in comparison with the enlarged surface of shrubland. This is particularly relevant in forest biomass
modeling in Mediterranean areas where fragmentation of forests is highly frequent due to strong human influence and management [86]. Moreover, in Mediterranean fire-prone landscapes, there is a current knowledge gap related to shrubland fuel load at the landscape scale and, therefore, lots of assumptions need to be taken in order to assess fire behavior at large scale [87]. In this sense, despite the acknowledged limitations, the shrubland AGB estimation carried out in this study represents a novel tool with relevant implications regarding fire regime modeling (and subsequently fire management, e.g., prescribed burning).

4.2. Environmental Drivers of Forest AGB Spatial Variability

Apart from the influence of dynamic type variables derived by optical remote sensing on AGB modeling, we also found a strong terrain variable influence. These factors derived from the DEM improved our forest AGB spatial estimate, though biomass changes cannot be detected from them, since topography is a static feature. The combination of both types of ecological drivers (static and dynamic) has overcome this limitation and allows us to monitor and easily update the biomass forest environmental changes.

Regarding the dynamic type variables, the previous covariates selection before modeling forest AGB at moderate spatial resolution showed that NDVI, combined with MSR and EVI, exerted a strong influence as environmental drivers among the estimated MODIS VIs. Because of the widely acknowledged close relationship between vegetation characteristics described by satellite data and biomass, remote sensing data are often used to calculate biomass estimates, showing a high correlation mainly with NDVI and describing similar insights as ours [88]. Moreover, our results revealed the importance of other covariates regarding different time dynamics which span from short to large time-scale, including seasonal mean dynamics.

The implementation of QRF to a large data set allows capturing complex hidden patterns between covariates and the response variable which are difficult to detect, or even difficult to interpret ecologically. In this sense, the results of our analysis showed quite consistent ecological interpretations. In the VIs time series, we found that the strongest correlation between forest AGB and VIs variables were associated with dates in April, September, and October 2016. This short-time response was related to the data acquisition during the 2016 ALS flight (from August to November), hence highlighting the importance of a short time lag in data acquisition between NFI plots and LiDAR data to minimize the outcome uncertainty [38]. In annual mean VIs profiles, we detected strong correlations with May and August dates coinciding with the end of the spring growing season and the start of the autumn growing season, respectively. While some covariates of short-time and annual mean scales were linked to different moments of the high photosynthetic activity in spring (beginning in March–April and reaching a maximum in May), the remaining covariates were related to the summer drought recovery, which is typical in the European Mediterranean region. Finally, long-term ecosystem functioning factors were represented by the mean annual and standard deviation of the NDVI time series as annual primary production and vegetation seasonality indicators [89].

In spite of the satisfactory model accuracy reported in this research, the fitted AGB model should be tested including other ground cover specific variables, such as forest canopy cover, and accounting for more species-specific equations to calculate AGB for Mediterranean shrubland. The latter may improve the assessment of forest AGB and its associated uncertainty, relevant in the two-stage upscaling methods due to the error propagation at all scales. In this sense, this can be checked when comparing the AGB mean assessed in each stage with the ghmb model results. The AGB mean of the ALS map for the specified testing area was within the 95% confidence interval of the ghmb model, whilst the MODIS map AGB mean was slightly out of the lower limit. Therefore, as [43] claimed, ignoring the uncertainty in the AGB benchmark map may lead to under-estimation of estimator variance in a hierarchical process. However, both methodologies—ghmb and random forest—are not totally comparable since they are based on different modeling approaches (linear regression vs. decision trees, respectively) and the number of covariates also differs. Despite this, an independent,
external ground-based validation of the final MODIS map would be desirable, as it would allow the analysis of the accuracy and reduce the high uncertainty estimates which limit the understanding of the spatial variability of AGB.

5. Conclusions

The methodology proposed here has been shown to provide a cost-effective and easy to update model aimed at generating spatially extensive maps at management scales concordant with AGB complexity. This confirms our hypothesis regarding the correlation between ALS-based forest AGB and MODIS–derived VIs. In this study, we specifically leveraged the benefits of two complementary remote sensing technologies to map forest AGB over large areas and integrated ALS pixel plots as a means to increase the distribution and sample size of target data for modeling AGB as a function of VIs moderate resolution composites (MODIS).

The information necessary to perform the methodology tested in this study is available for the entire Spanish national territory and free to access. Even though there is a date inconsistency in ALS data in Spain with several years of time lag between two different areas owing to the time frequency of ALS flights each 6–7 years, enough networks of airborne ALS transect can be used to bridge the scale gap between field samples and satellite images. Due to the high temporal availability and moderate spatial resolution, MODIS imagery was shown to be feasible in forest AGB estimation for regional, even national, land use planning and management. The two-stage upscaling method for AGB assessment tested in this study can be applied over large areas, such as the whole country, to gain insights into national forest aboveground biomass, and subsequently into carbon stock, which remains a key challenge still unsolved in Spain.


Funding: Pilar Durante’s participation was supported by a predoctoral grant [DI-15-08093] and Nur Algeet [PTQ-14-07206], Santiago Martín-Alcón [PTQ-15-07975] and Assu Gil-Tena [PTQ-16-08741] by postdoctoral grants awarded by the ‘National Programme for the Promotion of Talent and Its Employability’ of the Ministry of Economy, Industry, and Competitiveness (Torres-Quevedo program), which are partially funded by the European Social Fund (ESF) from the European Commission. Laura Recuero was also supported by a predoctoral grant [FPU014/05633] from the Spanish Ministry of Science, Innovation and Universities.

Acknowledgments: Authors are grateful to the Andalusian Scientific Computing Centre (CICA) for the access to computational facilities and the support service as well as the Spanish National Geographic Institute (IGN) as the provider of LiDAR data.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References


23. Didan, K. MODI13Q1 MODIS/Terra Vegetation Indexes 16-Day L3 Global 250 m SIN Grid V006 [Data Set]; NASA EOSDIS LP DAAC: Sioux Falls, SD, USA, 2015. [CrossRef]


61. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Development Core Team: Vienna, Austria, 2014.


63. Vermote, E. MOD09Q1 MODIS/Terra Surface Reflectance 8-Day L3 Global 250 m SIN Grid V006 [Data Set]; NASA EOSDIS LP DAAC: Sioux Falls, SD, USA, 2015.


73. Xu, B.; Huang, J.Z.; Williams, G.; Wang, Q.; Ye, Y. Classifying very high-dimensional data with random forests built from small subsamples. *Int. J. Data Wareh. Min.* 2012, 8, 44–63. [CrossRef]


82. Ozdemir, I. Linear transformation to minimize the effects of variability in understory to estimate percent tree canopy cover using RapidEye data. *Glaci. Remote Sens.* 2014, 51, 288–300. [CrossRef]


