Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data

Gaia Vaglio Laurin
Johannes Balling
Piermaria Corona
Walter Mattioli
Dario Papale
Nicola Puletti
Maria Rizzo
John Truckenbrodt
Marcel Urban
Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data

Gaia Vaglio Laurin,a,* Johannes Balling,b Piermaria Corona,c Walter Mattioli,c Dario Papale,a Nicola Puletti,c Maria Rizzo,c John Truckenbrodt,b and Marcel Urbanb

aUniversity of Tuscia, Department for Innovation in Biological, Agro-Food and Forest Systems, Viterbo, Italy
bFriedrich-Schiller-University, Department for Earth Observation, Jena, Germany
cCouncil for Agricultural Research and Economics, Research Centre for Forestry and Wood (CREA-FL), Arezzo, Trento, Italy

Abstract. The objective of this research is to test Sentinel-1 SAR multitemporal data, supported by multispectral and SAR data at other wavelengths, for fine-scale mapping of above-ground biomass (AGB) at the provincial level in a Mediterranean forested landscape. The regression results indicate good accuracy of prediction ($R^2 = 0.7$) using integrated sensors when an upper bound of 400 Mg ha$^{-1}$ is used in modeling. Multitemporal SAR information was relevant, allowing the selection of optimal Sentinel-1 data, as broadleaf forests showed a different response in backscatter throughout the year. Similar accuracy in predictions was obtained when using SAR multifrequency data or joint SAR and optical data. Predictions based on SAR data were more conservative, and in line with those from an independent sample from the National Forest Inventory, than those based on joint data types. The potential of S1 data in predicting AGB can possibly be improved if models are developed per specific groups (deciduous or evergreen species) or forest types and using a larger range of ground data. Overall, this research shows the usefulness of Sentinel-1 data to map biomass at very high resolution for local study and at considerable carbon density. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JRS.12.016008]

Keywords: above-ground biomass; forest; Sentinel-1; ALOS2.

Paper 170685 received Aug. 3, 2017; accepted for publication Dec. 13, 2017; published online Jan. 5, 2018.

1 Introduction

The retrieval of forest above-ground biomass (AGB, or biomass hereafter) using remote sensing data has received increasing attention in the last decades for several reasons: primarily, for the ability of remote sensing to spatially extrapolate point field information on forest parameters, enabling AGB mapping over large areas; for the increased availability of different Earth Observation data types; and, in a global scenario of climate change and biodiversity loss, for the relevance of forest biomass with respect to global carbon accounting and reporting, conservation and management of natural resources, and environmental modeling.1–7

Different methods have been developed to estimate AGB with remote sensing and ground data, based on passive and/or active instruments.8–10 Active sensors, such as light detection and ranging (LIDAR) and synthetic aperture radar (SAR), have the advantage to penetrate the canopy; for this reason, they are considered the most useful tools for providing vertical structure or volumetric forest measures.1,12

*Address all correspondence to: Gaia Vaglio Laurin, E-mail: gaia.vl@unitus.it

In the last decades, different new SAR missions have been developed with improved features, such as polarimetric capabilities, enhanced spatial resolution, and increased revisiting time. New missions are also planned for the near future, with a consistent increase in SAR data types available to scientific and user communities. This data availability calls for a better understanding of the utility of new SAR information in ecological research.

SAR sensors provide information on the backscattered energy from the illuminated target. Depending on the structural and dielectric characteristics of forest canopies, the relative contribution of volume, surface–volume interactions, and surface scattering from forest in the total backscatter signal of an imaging radar can vary. SAR sensors can have different wavelengths (bands) and thus different abilities to penetrate the forest. Moreover, they have the important advantage of being insensitive to cloud coverage and, when mounted on satellites, to cover large areas with repeated measures. As evidenced by different models, P- and L-bands provide stronger backscatter from branches and trunks with respect to X- and C-bands, respectively, which collect most of the energy from leaves and needles. In forest parameters characterization, notable results have been obtained with P- and L-bands or combining different wavelengths. Accurate results in forest attributes assessment have been obtained using SAR interferometry and polarimetric interferometry techniques, even if the scarce availability of interferometric couples collected in short time intervals, together with the high expertise required for the application of these methods, limits their use in operational forest monitoring.

Forest characterization has also been successfully carried out using C- or X-bands, especially when dense temporal series are used. In fact, these wavelengths are more sensitive to surface moisture. In this case, a multitemporal approach can reduce the signal variability; in addition, the availability of SAR data from different dates allows the selection of those periods and scenes having stronger sensitivity of the backscatter to the target.

SAR and other data types (e.g., LIDAR) can also be used in conjunction for forest research. The fusion of SAR and LIDAR has been attempted to improve SAR-based prediction, and it can be promising for large area efforts; a review showed that the main advantage is in the use of an accurate LIDAR digital elevation model for SAR-based AGB prediction or in the upscale of local LIDAR metrics to large areas with SAR data. Combining SAR and optical data has brought accurate AGB predictions in a number of cases because the structural SAR information can be complemented with canopy density, forest type, and foliage-related information collected by optical instruments.

Even with the increase in new data, fusion techniques, and innovative researches, AGB prediction based on SAR remains a challenging task, as saturation of the signal is common in forests, where the backscatter is influenced by several factors, such as the soil moisture, especially when the vegetation coverage is low; the forest type, the leaf presence with needleleaf forests possibly having the highest saturation level; the seasonal and weather conditions;! and the canopy roughness; and the SAR signal polarization as well as the incidence angle. Therefore, studies clarifying SAR response in different conditions and forests types are important, as they can support the effective exploitation of new SAR data streams into ecological research and monitoring.

The launch of the first Sentinel satellite (S1A: C-band SAR) by the European Space Agency (ESA) in 2014, followed by the S1B having the same characteristics as the first, now allows very frequent SAR data acquisitions. The subsequent launch of the Sentinel-2 satellites in 2015 and 2017 (S2A and S2B: enhanced multispectral sensors), together with a policy of free data, has strongly broadened the opportunities for forest monitoring. These new ESA data add to the growing archives of multispectral Landsat and ALOS SAR L-band SAR data.

The development of new perspectives and methods is illustrated by recent research, exploiting the potential of Sentinel data for forest characterization, such as new methods for forest deforestation, degradation, and fire monitoring; new detailing of canopy in terms of biophysical properties or for functional types classification; and innovative data fusion for mapping plantation extent. Very few studies addressed the applicability of S1, S2, or joint data, for biomass monitoring. For S2, a preliminary test was conducted by Chrysafis et al. in a heterogeneous Mediterranean forest to predict growing stock (GS) volume based on S2 and Landsat 8 data, while Majasalmi and Rautianen used simulated S2 data for characterizing biophysical variables in a boreal forest. For S1, Chang and Shoshany explored the potential of this data.
joined with S2 for biomass mapping in Mediterranean shrublands. In central Italy, where forests are fragmented and included in a highly heterogeneous landscape, fine-scale information on forest attributes is important to orient management and conservation activities. At the country level, AGB and related parameters (GS, basal area, etc.) were mapped in a number of remote sensing-based efforts. Some research addressed the testing of new methods and data at site level,\textsuperscript{50,51} while others covered the entire country.\textsuperscript{52,53} These large-scale efforts were based on optical medium-resolution imagery and produced important data sources. However, some limitations in these maps were recognized,\textsuperscript{53} in part due to the intrinsic features of the remote sensing and validation data used and in part due to the spatial resolution (1 km), too coarse to capture the variability of fragmented landscapes.

The objective of this research is to test S1 C-band SAR multitemporal data, also used in conjunction with ALOS2 L-Band SAR and S2 multispectral data, for detailed mapping of AGB at the provincial level in a Mediterranean forested landscape. The research is based on the hypotheses that S1 time series can provide biomass predictions at very high resolution, also due to the possibility of exploiting different seasonal information according to leaf presence (evergreen or deciduous species). Additionally, it has been hypothesized that biomass prediction could be improved by the integration of SAR at different frequency or with optical data. Specifically, the aims are to provide fine-scale maps of biomass for the Viterbo province in Central Italy and to improve the understanding of (i) the impact of leaf presence (deciduous or evergreen species) in S1 SAR response, (ii) the usefulness of S1 time series for accurate AGB prediction, and (iii) the advantages offered by the integration of S1 data with L-band SAR (ALOS2) or S2 multispectral data.

2 Materials and Methods

2.1 Study Area

Several mature forest stands were surveyed in different locations of the Viterbo province study area (Fig. 1). These forests have a high landscape conservation value and are included within the Natura 2000 sites network.

Main forest categories according to the European Forest Type classification system\textsuperscript{54} are Mediterranean pine forest, thermophilous deciduous forest, and mountainous beech forest. High forests (beech—\textit{Fagus sylvatica} and Turkey oak—\textit{Quercus cerris}) and coppice stands...
with standards (sweet chestnut—Castanea sativa and Turkey oak) are included. The soils are fertile, deep, and loose with acid pH, mainly classified as andisols and identified as “black soil.” The bioclimate is meso-mediterranean subhumid with a mean annual temperature of 12.8°C and annual rainfalls ranges between 1250 and 1550 mm. July is the hottest month with an average temperature of 22°C, and January is the coldest one with 4.2°C; summer drought does not usually occur due to the humid air coming from the Tyrrhenian Sea and the lakes of the area, although a subarid period between July and August may occur. Beech stands are also considered a habitat of European priority interest 9210*—Apennine beech forests with Taxus and Ilex, in accordance with the EU Habitats Directive (92/43/EEC). The only exceptions among the studied forest stands are represented by the Italian stone pine (Pinus pinea) forests located along the coastline, established from the 30s to the 50s of the last century by the State Forest Service and by the mixed forest stands located within “Villa Giustiniani-Odescalchi” in Bassano Romano, mainly composed by holly oak (Quercus ilex), sweet chestnut, and Norway spruce (Picea abies).

2.2 Ground Data

Surveyed plots (Table 1) are located in areas where major disturbances, such as forest fires, floods, or clear-cutting, have not been recorded for many years. The sampled stands differ for main forest type, leaf presence (evergreen or deciduous species), silvicultural system, and type of field plots (size, shape, and number). All ground data were collected between 2014 and 2016.

In each plot, the following attributes were measured: diameter at breast height (DBH) of living and dead stems (diametric threshold >5 cm), species, 30 height samples per main species per plot, and tree age (by coring) on stems with mean DBH. The following variables were derived: number of stems (living and dead), basal area, mean DBH, mean and top height, G5s; above-ground woody biomass, current annual increment of wood volume, and dominance of deciduous or evergreen trees. The mean stand height was calculated by the height–diameter curve with respect to the mean stand DBH. GS of wood volume for each plot was computed using equations, with DBH and height as explanatory variables, developed with data from the Italian National Forest Inventory (INFI). AGB for each plot was computed from GS by the models proposed by Federici et al.

### Table 1 Main characteristics of surveyed sample plots.

<table>
<thead>
<tr>
<th>Area</th>
<th>Location</th>
<th>Main forest type</th>
<th>Plot type</th>
<th>Plot size (m²)</th>
<th>Plot number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bassano Romano</td>
<td>Mixed high forest (including six deciduous and six evergreen)</td>
<td>Circular</td>
<td>314</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Montalto di Castro</td>
<td>Stone pine high forest (evergreen)</td>
<td>Circular</td>
<td>531</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>Vetralla – Arcionara</td>
<td>Beech high forest (evergreen)</td>
<td>Irregular</td>
<td>48,000</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Vetralla – Convento</td>
<td>Beech and Turkey oak high forest (evergreen)</td>
<td>Irregular</td>
<td>60,000</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Mt. Cimino</td>
<td>Sweet chestnut coppice with standards (deciduous)</td>
<td>Circular</td>
<td>452</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Caprarola – Mt. Venere</td>
<td>Beech high forest (evergreen)</td>
<td>Polygon</td>
<td>11,150</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>Grotte Santo Stefano</td>
<td>Turkey oak coppice with standards (deciduous)</td>
<td>Circular</td>
<td>1257</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>Caprarola – Mt. Venere</td>
<td>Turkey oak high forest (deciduous)</td>
<td>Circular</td>
<td>1257</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>Caprarola – Le Coste</td>
<td>Sweet chestnut coppice with standards (deciduous)</td>
<td>Circular</td>
<td>531</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>Rocca dei Papi</td>
<td>Turkey oak coppice (deciduous)</td>
<td>Circular</td>
<td>531</td>
<td>4</td>
</tr>
</tbody>
</table>
Overall, the measured variables cover a broad range of values: number of stems ranges from 100 (mature beech high forest) to about 5000 tree per ha (20-year-old sweet chestnut coppice), whereas basal area ranges from about $20 \text{ m}^2 \text{ ha}^{-1}$ (22-year-old Turkey oak coppice and coastal Stone pine high forest) to $50.9 \text{ m}^2 \text{ ha}^{-1}$ in the artificial mixed high forest stand within the historical park of Bassano Romano. AGB goes from values lower than $100 \text{ Mg ha}^{-1}$ in the coastal Stone pine high forest and the young sweet chestnut coppice to values higher than $600 \text{ Mg ha}^{-1}$ in the unmanaged holly oak stand within the historical park in Bassano Romano and in the old-growth high forest in Caprarola. Mean DBH ranges from 8 cm in the young coppice to 67 cm in the mature high stand; similarly, mean height ranges from 10 to 37 m in the same stands. Table 2 reports the mean values observed for the selected parameters in each sampled forest stand.

For comparison purposes, AGB data provided by the INFC, collected in plots located in the study area, were used as an independent dataset. Details on the data collection for these plots are given in Ref. 57.

### Table 2 Mean values observed of the selected parameters per area.

<table>
<thead>
<tr>
<th>Area</th>
<th>Location</th>
<th>Number of stems (n ha$^{-1}$)</th>
<th>Basal area (m$^2$ ha$^{-1}$)</th>
<th>GS (m$^3$ ha$^{-1}$)</th>
<th>AGB (Mg ha$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Bassano Romano</td>
<td></td>
<td>1738</td>
<td>50.9</td>
<td>709.1</td>
<td>521.5</td>
</tr>
<tr>
<td>2 Montalto di Castro</td>
<td></td>
<td>271</td>
<td>19.6</td>
<td>184.7</td>
<td>113.2</td>
</tr>
<tr>
<td>3 Vetralla – Arcionara</td>
<td></td>
<td>481</td>
<td>31.9</td>
<td>452.2</td>
<td>356.6</td>
</tr>
<tr>
<td>4 Vetralla – Convento</td>
<td></td>
<td>231</td>
<td>23.0</td>
<td>278.6</td>
<td>221.6</td>
</tr>
<tr>
<td>5 Mt. Cimino</td>
<td></td>
<td>1150</td>
<td>28.2</td>
<td>254.6</td>
<td>148.1</td>
</tr>
<tr>
<td>6 Caprarola – Mt. Venere</td>
<td></td>
<td>182</td>
<td>34.0</td>
<td>542.1</td>
<td>428.1</td>
</tr>
<tr>
<td>7 Grotte Santo Stefano</td>
<td></td>
<td>966</td>
<td>20.6</td>
<td>154.8</td>
<td>126.4</td>
</tr>
<tr>
<td>8 Caprarola – Mt. Venere</td>
<td></td>
<td>711</td>
<td>46.2</td>
<td>607.8</td>
<td>490.4</td>
</tr>
<tr>
<td>9 Caprarola – Le Coste</td>
<td></td>
<td>4982</td>
<td>27.5</td>
<td>157.3</td>
<td>113.7</td>
</tr>
<tr>
<td>10 Rocca di Papa</td>
<td></td>
<td>740</td>
<td>24.7</td>
<td>250.8</td>
<td>177.6</td>
</tr>
</tbody>
</table>

Overall, the measured variables cover a broad range of values: number of stems ranges from 100 (mature beech high forest) to about 5000 tree per ha (20-year-old sweet chestnut coppice), whereas basal area ranges from about $20 \text{ m}^2 \text{ ha}^{-1}$ (22-year-old Turkey oak coppice and coastal Stone pine high forest) to $50.9 \text{ m}^2 \text{ ha}^{-1}$ in the artificial mixed high forest stand within the historical park of Bassano Romano. AGB goes from values lower than $100 \text{ Mg ha}^{-1}$ in the coastal Stone pine high forest and the young sweet chestnut coppice to values higher than $600 \text{ Mg ha}^{-1}$ in the unmanaged holly oak stand within the historical park in Bassano Romano and in the old-growth high forest in Caprarola. Mean DBH ranges from 8 cm in the young coppice to 67 cm in the mature high stand; similarly, mean height ranges from 10 to 37 m in the same stands. Table 2 reports the mean values observed for the selected parameters in each sampled forest stand.

For comparison purposes, AGB data provided by the INFC, collected in plots located in the study area, were used as an independent dataset. Details on the data collection for these plots are given in Ref. 57.

### 2.3 Remote Sensing and Ancillary Data

In the current study, Sentinel-1 C-band SAR data interferometric wide swath mode was used, with a 250-km swath width at $5 \times 20 \text{ m}$ spatial resolution, an incidence angle between 29.1 deg and 46.0 deg, and VH and VV dual polarization. The study area was fully included in the extent of a scene. A total of 25 scenes from all of 2015 were acquired; 10 stacks of 3-month scenes were composed using a one-month moving temporal window (first January to March to last October to December); and the number of scenes in each stack ranged from 6 to 8, except for the last two stacks of the year (August to October and September to December) composed of four scenes and for October 2015 no S1 data were available. Scenes were multilooked (one look in range and four in azimuth), geocoded based on Shuttle Radar Topography Mission (SRTM) data, and radiometrically calibrated with a final pixel spacing of $20 \times 20 \text{ m}$. Areas of data distortion, e.g., shadowing due to layover and foreshortening, were masked out. Pixels values were converted to backscattering (or normalized radar cross section), measured in decibel (dB), according to the following equation:

$$\sigma_0 \text{ dB} = 10 \log 10 \sigma_0,$$

in which $\sigma_0 \text{ dB}$ is the normalized radar cross section and $\sigma_0$ is the backscatter for a specific polarization. A multitemporal speckle filter was applied to reduce noise, according to Ref. 58.

Finally, data were averaged according to the 3-month stacks, as well as yearly.
The Japan Aerospace Exploration Agency ALOS2 L-band SAR data were acquired in fine beam double-polarization mode (HH and HV) characterized by 40- to 70-km swath width at 9.1 × 5.3 m spatial resolution, incidence angle between 28.6 deg and 32.9 deg and HH and HV dual polarization. The study area is covered by four scenes: two from March 9, 2015, and two from April 20, 2015. Scenes were multilooked (one look in range and two in azimuth), geocoded based on SRTM data, and radiometrically calibrated with a final pixel spacing of 20 × 20. A 5 × 5 Lee speckle filter was used for noise reduction. For conversion to dB, the same procedure indicated for S1 data was used.

For both S1 and ALOS2 data, the preprocessing was conducted in Sentinel Application Platform (SNAP) v. 4.0.0. A visual inspection of the misalignment between S1 and ALOS2 due to geometrical distortions in areas where plots occur resulted in an offset in a 2- to 10-m range.

In addition, we utilized a Sentinel-2 image from June 25, 2016. S2 data are characterized by 13 spectral bands with 10-, 20-, and 60-m spatial resolution and a radiometric resolution of 12 bit. Bands at 60 m were excluded from the analysis. The image was first downloaded as a level-1C S2 top-of-atmosphere reflectance product, and atmospheric correction, using the “sen2cor” SNAP’s plug-in, and orthorectification were performed in the SNAP, which also allowed the generation of a cloud cover mask. The 10-m bands (three in the visible and three in the near-infrared (NIR) portion of the electromagnetic spectrum) were resampled to match the 20-m spatial resolution of the other S2 bands (four red edges and two SWIR bands) and of S1 and ALOS2 data. Normalized difference vegetation index (NDVI) and Red edge normalized difference vegetation index (RENDVI) were computed; RENDVI is calculated using the NDVI equation but replacing the red band with the red edge band at 740 nm.

The forest type map for the Viterbo province was obtained from the Lazio region; it was used as ancillary information to exclude nonforested areas in AGB maps production, as well as to obtain a deciduous/evergreen mask then used in the AGB prediction models exploiting this categorical information.

2.4 Data Analysis

The AGB values derived from field data collected in plots ranged from 70.4 to 884 Mg ha⁻¹, with a mean equal to 292.8 Mg ha⁻¹ and standard deviation to 217.9 Mg ha⁻¹. Two plots with values >800 Mg ha⁻¹, including each with a very large stem, were outliers and thus excluded; one plot of 119 Mg ha⁻¹ characterized by unusual backscattering values in S1 images (due to shadowing effects) was also excluded.

Hence, 29 plots were actually used for the analyses, with AGB ranging from 70.4 to 717 Mg ha⁻¹, with mean equal to 255 and standard deviation equal to 167.9 Mg ha⁻¹. A buffer was applied around the 20 circular plots having radius <20 m to obtain an equal radius of 20 m for all the circular plots. This procedure was conducted after evaluating with very high-resolution Google Earth imagery that the canopy in the buffer was as dense as that inside the plots. Four polygonal large plots and five circular plots of 20-m radius were not buffered. The buffer purpose was to reduce the possible influence caused by differences in plots size; having a larger number of pixels also helps to reduce the SAR signal variability among very small areas. Eventually, 25 plots of 20-m radius plus 4 larger polygons were obtained. A layer with plot surface was overlapped to imagery to allow extraction of the corresponding pixels.

Pixels having more than 70% of the surface included in the plot area were extracted and averaged. To exclude the possibility that larger polygonal plots had stronger influence in the prediction models, it was verified that the variance of the pixels values from larger plots was in the range of the variance of the circular plots.

For multitemporal S1 data, scenes from 10 three-month periods were stacked, and a yearly stack with all the 2015 scenes was also produced. Data for each of the stacks were averaged, generating the S1 response in 10 different time periods and for the whole year. In addition to single-polarization values, polarizations sum and difference were computed for S1 and ALOS2 data. Being SAR backscattering values expressed in dB scale, their computed difference in dB (VH – VV and HV – HH) corresponds to a quotient, while their computed sum (VH + VV and HV + HH) to a product. For S2, the NDVI and RENDVI vegetation indices were computed. In synthesis, the data extracted and averaged for each plot were...
Laurin et al.: Above-ground biomass prediction by Sentinel-1 multitemporal data...

- S1: 10 three-monthly and 1-yearly value per VV, VH, VH – VV and VH + VV polarizations (total 44 values),
- ALOS2: 1 value for HV, HH, HV – HH, and HV + HH polarizations (total four values), and
- S2: 10 bands and 2 vegetation indices (total 12 values).

In the light of the phytoclimate seasonality of the Viterbo province,62 the correlation between the 10 S1 time series and field AGB values was first explored for deciduous and evergreen dominated plots (17 and 12 plots, respectively) using Pearson’s correlation coefficient. Stepwise regressions using different combinations of S1 data (44 inputs) and S1 + ALOS2 data (48 inputs) were performed. These regressions were also repeated using the leaf presence information as an added categorical variable to test any difference in signal behavior in these two groups. The addition of S2 data (12 inputs) to S1 and ALOS2 datasets was eventually attempted. All stepwise regressions were performed using a 0.15-alpha value for inputs to enter and to leave. The best regression results were used to produce Viterbo province AGB maps, masking out nonforest areas on the basis of the ancillary information and back-transforming logarithmic AGB values using a correction factor.63 Considerations were made with respect to the differences between the two maps in terms of basic statistics. Similarly, considerations were made with respect to differences between the maps and the INFC 2005 independent dataset (34 circular plots, 13-m radius); in this case, AGB values were extracted from the two maps using a 3 × 3 pixel window centered in the INFC plot center.

The analyses were performed in ENVI (Exelis Visual Information Solutions, Boulder, Colorado), R (R Foundation for Statistical Computing, Vienna, Austria), MATLAB® (The MathWorks Inc., Natick, Massachusetts), and Minitab (Minitab, Inc. Statistical Software, State College, Pennsylvania) software.

3 Results

The correlation between S1 backscatter and AGB field values, in different time periods, is shown in Fig. 2 for deciduous and in Fig. 3 for evergreen plots. For evergreen plots, the correlation was always very high (>0.75) with minimal variations among the different time series and polarizations. Lower correlation values (<0.7) were found for the deciduous group, with large variations in different time periods.

The stepwise regression tests were conducted using AGB values transformed in natural logarithm, with results summarized in Table 1. The selected inputs were always significant (p-value < 0.05). When the 44 S1 inputs were tested, only the “S1 VH-VV Feb-Apr” time series was selected, and the regression obtained with leave-one-out (LOO) cross validation had a coefficient of determination ($R^2$) equal to 0.55 and an RMSE equal to 126 Mg ha$^{-1}$. The four ALOS2 inputs were added to this S1 input, repeating the stepwise procedure with 48 inputs: the results were the selection of “S1 VH-VV Feb-Apr,” “S1 VH Apr-June,” and “ALOS2 HV-HH” inputs, and an LOO-validated $R^2$ equal to 0.67 with an RMSE of 100 Mg ha$^{-1}$. The inclusion of

![AGB-S1 correlation - deciduous](image)

**Fig. 2** Pearson’s correlation coefficient calculated between S1 backscattering and AGB for the 17 deciduous plots and for different time series.
quadratic terms determined an improvement of both LOO-validated $R^2$ (equal to 0.71) and RMSE (74.5 Mg ha$^{-1}$). Categorical information on leaf presence was then included but not selected as significant input at this stage.

Prediction for a smaller AGB range, arbitrarily set up to 400 Mg ha$^{-1}$, was also attempted, and this allowed keeping a reasonable number of plots (24), ranging from 70.4 to 393 Mg ha$^{-1}$, with an average AGB of 201 Mg ha$^{-1}$ and a standard deviation equal to 103.2 Mg ha$^{-1}$. When S1 inputs were tested using this reduced AGB range, the stepwise procedure resulted again in the selection of “VH-VV Feb-Apr” (as for the full set of AGB) as input with an LOO-validated $R^2$ of 0.52 and RMSE of 71 Mg ha$^{-1}$; adding ALOS2 data resulted also in a selection similar to the one obtained with the full AGB range, with “VH-VV Feb-Apr” and “ALOS2 HV-HH” inputs and an increase of LOO-validated $R^2$ to 0.56 and a decrease of RMSE to 68 Mg ha$^{-1}$. The use of quadratic terms did not improve this result, but, when the categorical information on deciduous or evergreen dominance was added, the LOO-validated $R^2$ increased to 0.7 and RMSE decreased to 46.6 Mg ha$^{-1}$. Figure 4 shows the predicted versus observed AGB values for this LOO regression.

When the S2 inputs were added to those used in the previous regression test (with reduced AGB range), the stepwise procedure discarded ALOS2 and selected the “S1VH-VVFeb-Apr” from S1 and B6, B8, and B8A from S2, with the deciduous or evergreen information being not significant; similarly to the previous test, the LOO $R^2$ was equal to 0.7 and RMSE to 47 Mg ha$^{-1}$. Figure 5 shows the scatter plot of predicted versus observed AGB LOO values.

**Fig. 4** Scatterplot of predicted versus observed values [slope for blue evergreen (0) line = 0.9 and slope for red deciduous (1) line = 0.76]. The predicted AGB values are obtained using back-transformed AGB values from the following equations based on S1 and ALOS2 inputs and categorical information ($R^2 = 0.7, \text{RMSE} = 46.6 \text{ Mg ha}^{-1}$): evergreen (0): $\text{ln}(\text{AGB}) = 2.595 - 0.3986 \text{ S1VH-VVFeb-Apr} - 0.1848 \text{ ALOS2 HV-HH}$; deciduous (1): $\text{ln}(\text{AGB}) = 3.957 - 0.3986 \text{ S1VH-VVFeb-Apr} + 0.1848 \text{ ALOS2 HV-HH}$. 

---

**Fig. 3** Pearson’s correlation coefficient calculated between S1 backscattering and AGB for the 12 evergreen plots and for different time series.
The (f) and (g) inputs in Table 3 were considered better suited for AGB prediction, even in view of the upper limit of 400 Mg ha\(^{-1}\) of model validity, due to lower RMSE values and the absence of quadratic forms in the regressions. These inputs were used to produce two different AGB maps for the forested areas in the Viterbo province: one being based on SAR only and the other on joint SAR and optical data (Fig. 6).

A statistical comparison of the maps values is reported, but it should be considered just as an indication because values external to the range used for AGB regressions are not fully reliable. According to the S1-S2 map, the mean AGB in the Viterbo forests considered here (total surface equal to 678.1 km\(^2\)) is 365.4 Mg ha\(^{-1}\), with standard deviation of 153.2 Mg ha\(^{-1}\), and a total AGB of 24699.6 Gg. The S1-ALOS2 map showed a mean AGB equal to 219.2 Mg ha\(^{-1}\), with standard deviation of 113.9 Mg ha\(^{-1}\) and total AGB of 14862.2 Gg. The frequency distribution of the AGB found in the two maps is also compared with the measurements extracted from the INFC that shows in general lower values (Fig. 7); the per-pixel Pearson’s correlation coefficient between the two maps equaled 0.36.

### Table 3

<table>
<thead>
<tr>
<th>Inputs Selection</th>
<th>(R^2)</th>
<th>RMSE (Mg/ha(^{-1}))</th>
<th>RMSE %</th>
</tr>
</thead>
<tbody>
<tr>
<td>29 plots (up to 717 Mg ha(^{-1}))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) 44 S1 S1 VH-VV Feb-Apr</td>
<td>0.55</td>
<td>126</td>
<td>19.5</td>
</tr>
<tr>
<td>(b) 44 S1 + 4 ALOS2 S1 VH-VV Feb-Apr, S1 VH Apr-June, ALOS2 HV-HH</td>
<td>0.67</td>
<td>100</td>
<td>15.5</td>
</tr>
<tr>
<td>(c) 45 S1 + 4 ALOS2 S1 VH-VV Feb-Apr, S1 VH Apr-June, ALOS2 HV-HH, + S1 quadratic terms</td>
<td>0.71</td>
<td>74.5</td>
<td>11.5</td>
</tr>
<tr>
<td>24 plots (up to 400 Mg ha(^{-1}))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d) 44 S1 S1 VH-VV Feb-Apr</td>
<td>0.52</td>
<td>71</td>
<td>21.5</td>
</tr>
<tr>
<td>(e) 44 S1 + 4 ALOS2 S1 VH-VV Feb-Apr, ALOS2 HV-HH</td>
<td>0.56</td>
<td>68</td>
<td>20.6</td>
</tr>
<tr>
<td>(f) 44 S1 + 4 ALOS2 S1 VH-VV Feb-Apr, ALOS2 HV-HH, + deciduous and evergreen info</td>
<td>0.7</td>
<td>46.6</td>
<td>14.1</td>
</tr>
<tr>
<td>(g) 44 S1 + 4 ALOS2 + 12 S2 S1 VH-VV Feb-Apr, B6, B8, B8A</td>
<td>0.7</td>
<td>47</td>
<td>14.2</td>
</tr>
</tbody>
</table>
The visual inspection of a map of AGB absolute values differences (Fig. 8), obtained sub-
tracting the per-pixel values of the maps, revealed larger disagreement (>200 Mg ha⁻¹) in areas
having higher AGB density in the S1-S2 map.

Differences in AGB between the maps and INFC 2005 data were computed for 34 plots for
S1-ALOS2 and 24 plots for S1-S2 because areas with prediction AGB >400 Mg ha⁻¹ were
excluded from this comparison. The resulting AGB differences were significant for both maps,
but the distribution of the error did not show any clear trend at increasing AGB values. The
histogram of the INFC 2005 data (Fig. 7) was more similar to S1-ALOS2 AGB distribution,
as most values were <200 Mg ha⁻¹. The main differences were observed for classes having
low AGB (<100 Mg ha\(^{-1}\)); those classes are less represented in the ground data that were used to train and validate the models. In general, AGB values from both maps were higher when compared with INFC 2005 ground data.

INFC estimates are not available at the provincial level but only at the regional one; the Viterbo province covers 21% of the Lazio region surface, for which the INFC reported for 2005 a total AGB equal to 46010.5 Gg, distributed in a forested surface of 5438.5 km\(^2\) with mean AGB equal to 84.6 Mg ha\(^{-1}\).\(^{57}\)

---

**Fig. 7** Histogram of percentage of plots/pixels occurring in different AGB classes of the INFC data (plots, in yellow) and of the AGB maps (pixels, in red S1-ALOS2 and in green S1-S2).

**Fig. 8** Map of the difference in AGB between S1-S2 and S1-ALOS2 maps.
4 Discussion

At C-band, the contribution from leaves and the upper canopy elements tends to dominate the signal, being the most radar sensitive to features having dimension characteristics of the scale of the probing wavelength. In the present study, the main contribution from crowns determined the S1 sensitivity to seasonal phenological behavior for the deciduous plots, as shown by the Pearson’s correlation values found for this group, which greatly changed according to seasons and polarizations (Fig. 2). Instead, for the evergreen group, the correlation values were always high, with minimal changes throughout the year (Fig. 3). In deciduous plots, it is reasonable to have high VH-AGB values only in the leaves-on season, as in this period the leaves’ presence contributes to crown volumetric scattering, which is missing in leaves-off conditions. The same applies to evergreen plots, as the presence of leaves throughout the year always produces high volumetric scattering. For the deciduous group, the polarization ratio showed higher scores; this ratio can be linked to the presence of small branches, which are exposed in leaf-off months and are related to higher biomass because of their abundance. Additionally, as VV backscatter is usually linked both to water content and surface roughness, the combination of VH and VV possibly reduced the influence of surface scattering and foliar water content, making the S1 signal more sensitive to AGB.

In agreement with these considerations, the “S1VH-VV Feb-Apr” input was always selected in the different regression tests (Table 3), resulting in very high scores for both deciduous and evergreen groups. The winter ratio selection confirms the S1 seasonal sensitivity for one of these groups, a fact previously observed in other research, which found significant differences among the S1 signal in different seasons; also, note that deciduous forests were more affected than evergreen ones by seasonal changes.

The regression analysis in Table 3 indicates that S1 alone moderately explains the AGB variability in the study area, even with temporal time series availability. The integration of ALOS2 L-band to S1 always produced a consistent improvement resulting in increased $R^2$ and reduced RMSE with respect to S1 input alone. The use of multiband SAR information for improving AGB predictions is well documented in the literature for different forests and data types: Ranson and Sun used P- or L-band data in conjunction with C-band data in a northern forest, Ferrazzoli et al. tested the same bands for crop and arborous vegetation in Italy, Mohan et al. used models based on multiband SAR data to improve assessment in India, and Stelmaszczuk-Gorska et al. used L- and C-band data in Russian forests. Considering the limited penetration capability of S1, a benefit from the addition of L-band data was expected for providing information related to larger forest structural elements. For both SAR types, the combination of polarizations showed better results since they can be linked to biomass. This was also observed in previous research, independently by the used SAR bands: polarization ratios obtained from L-band have proved effective in subtropical and tropical forests and ratios from C-band have been used in subtropical forests, as well as in Mediterranean shrublands, where it proved to be less affected by topography effects. However, other studies are needed to clarify the signal mechanisms that lead, in deciduous or evergreen groups, in different forest types, and in varying environmental conditions, to the use of different channels combinations.

Even if the accuracy is improving by fusing C- and L-band data, the tests carried out with the full range of field values (up to 717 Mg/ha) resulted in high RMSE and deviance of predicted values with respect to observed ones. Additional tests carried out using reduced calibration data (upper limit of 400 Mg/ha) lead to more accurate results with a consistent RMSE decrease. Two “upper-bounded” models were selected to produce the AGB Viterbo province maps, having equal prediction accuracy. In the first, based only on SAR inputs, both SAR bands were useful, and the leaf presence information was exploited; in the second, with added S2 data to SAR inputs, ALOS2 was not selected, but instead three bands from the NIR and specific red-edge spectral portion were included. The first model confirmed the usefulness of multiband SAR, while the second showed the complementarity of optical and SAR data, which has been already recognized by a range of biomass studies. Fusing S1 and S2 data, an improvement around 14% in AGB prediction in Mediterranean shrublands was obtained with respect to the usage of single sensors. In other forest studies conducted in China, Italy, and Thailand, the integration of ALOS L-band and Landsat data was usefully employed for biomass assessment. An improvement in accuracy of AGB predictions was also obtained using
SAR in multiple bands together with very high-resolution optical images. The S2 optical spectral region selected in the present research is linked to chlorophyll content and, thus, to both number of leaves and their photosynthetic activity and health conditions. Horler et al. noted that the position of the red edge is affected by the number of leaves layers, which in turn can be related to canopy thickness, leaf area index, and biomass. The relevance of red-edge information for AGB prediction has been previously proved; red edge bands have been used as additional data in LIDAR-based tropical forest biomass predictions and for wetland biomass prediction.

The two produced AGB maps, one based on SAR data and the other on SAR plus optical data, showed reasonable agreement with each other at the per-pixel level, except in areas with high AGB values in the S1-S2 map. When considering the total AGB for the province, the prediction from S1-ALOS2 seems in line with what would be expected if the INFC regional estimates are scaled over the provincial territory and the large occurrence of mature forests in Viterbo area is taken into account. The discrepancy between the two maps can be attributed to various reasons; the most obvious is the intrinsic limits of the models, calibrated up to 400 Mg ha\(^{-1}\) limit, with model scatter plots also showing larger uncertainties at higher AGB values.

The saturation of the SAR signal at high AGB values is likely to be responsible for under-prediction of AGB in the S1-ALOS2 map. SAR saturation has been observed since the first SAR-based biomass prediction researches, with X- or C-band usually associated with saturation at much lower values with respect to other bands. Overprediction of biomass when using optical data is also known to occur, as previously reported with different optical data types and when using joint optical and microwave data, such as in this case.

Given the lack of additional ground truth for further validation, it is difficult to assess which of the maps better approximate the actual AGB values in the different forests of the study area, and it is likely that either overprediction or underprediction occurred. In fact, the slopes observed in Figs. 4 and 5, relating to predicted versus observed values, indicate slight underprediction (being all slopes <1) for both maps while the comparison with INFC-independent sample indicates general overprediction. However, this comparison is influenced by the many differences occurring between the INFC inventory data and the ground data used to generate maps, such as minimum DBH of sampled trees, data collection dates, plot size, and shape. In addition, the maps have a 400-m² pixel that do not match well with the INFC circular plot position and size (530 m²), and the INFC values cannot be safely extrapolated to the larger area covered by a 3 × 3 pixel window used to extract AGB from the maps. The INFC histogram resembles that of the S1-ALOS2 distribution, except for lower AGB classes, which are limitedly represented in the ground truth data used in this study. The scarce number of low AGB plots in the field data is possibly responsible for overprediction at low biomass ranges in both maps and reflects the general difficulty found in research studies to directly collect forest information or to get the availability of ground data from other sources (e.g., inventories). At higher biomass ranges, overprediction may occur, especially when using optical data in addition to SAR due to extrapolation in areas in which crowns show different reflectivity characteristics with respect to plots. The crown spectral characteristics are often influenced by microclimate, water conditions, or other forest health status; future work, such as introducing finer thematic information on leaf presence and forest types into the models, could clarify if this is the case. Presently, we consider that predictions based on SAR data are more conservative than those based on joint optical and SAR, which can generate—at least in this study area—overprediction in AGB values. It is worth noticing that the model based on SAR data required previous knowledge on the deciduous/evergreen dominance, while the one using S2 input did not. This fact gives an added value to S2 data and to the model based on joint SAR and S2, especially considering that this detailed forest type information, which was used here, is not always available.

## 5 Conclusions

In general, radar has been less frequently used in forestry with respect to LIDAR and optical data. Sentinel-1 is now devised as an effective SAR system to improve forest vegetation remote sensing, and studies exploring radar data potential for forestry applications can facilitate a wider use of SAR sensors.
Overall, this research shows the usefulness of Sentinel-1 data to map forest biomass at a spatial resolution which has rarely been achieved but that is requested for local management activities. Accurate AGB predictions were obtained up to 400 Mg ha\(^{-1}\), which is a value that encompasses most of the forests found in the Mediterranean area. Given the lack of satellite data optimally suited for AGB mapping, the results presented here can be relevant, as they show that, with currently available free SAR data, fine-scale biomass analysis is possible.

In this research, the relatively small number of available ground data did not allow for performing AGB predictions for deciduous or evergreen dominated forests separately; this ground data scarcity is common given the resources needed for field data collection and the restriction imposed on official inventories. The number of available plots thus represents a limit in this research, and the findings reported here have to be considered as preliminary indications, to be expanded for future use of S1 data for biomass mapping purposes.

The results suggest that at C-band the availability of multitemporal information is very important for improved characterization of broadleaf forests. However, the accuracy obtained using S1 data alone was moderate, and the integration of L-band information on leaf presence proved necessary to reduce uncertainty in the AGB predictions. The potential of S1 data in predicting AGB can possibly be improved if models are developed per specific groups (evergreen or deciduous species) or per forest types. The integration of optical S2 data to S1 data showed results that could lead to overestimation; however, other efforts are needed in this respect to confirm this finding. It would be also relevant to evaluate if forest traits other than leaf presence, derived by S2 data, can be useful for improving AGB prediction.

An indication about the optimal periods of satellite data acquisition can also be derived from this research. The time period of late winter and beginning of spring is suggested in the case of SAR data, when the deciduous trees are still leafless and the signal from branches is more evident. Even if this study did not test optical time series, for this data type, the main acquisition issue is related to cloud or haze presence, which even at these latitudes can be consistent in forested areas; the late spring and beginning of summer period, before the leaves dry and start senescence phase, is, therefore, indicated for S2 images.

In addition, this research evidenced, on one hand, the lack of data and the high uncertainty in AGB estimates at the provincial level, where existing maps and inventory data do not provide sufficient information. On the other hand, the limited sensitivity of currently available satellite data for AGB prediction is evident, considering that the best models were those calibrated with an upper AGB bound of 400 Mg ha\(^{-1}\). While waiting for future sensors better suited for forest studies, biomass data can be delivered at fine scale using multiband SAR sensors, as this research showed. Considering that forest areas with very high carbon density are not so diffuse in this region and that for applied purposes it is preferable to maintain as minimum as possible the deviance of predictions, the presented models can be of practical use for analyzing in detail the spatial distribution of biomass in fragmented landscapes of moderate biomass density.

A full open-access data policy, such as the one adopted for Sentinels or Landsat, can favor the integration of optical or multifrequency data to S1 and is urgently needed for other SAR missions to support the full exploitation of SAR potential for forest mapping and inventory. For comparison with inventory data, the collaboration with the National Forest Service allowed the current analysis to be conducted through matching between plots and pixels from the maps created by service staff. However, exact geolocations of plots are not available for security and plot protection reasons; this means that without the support of the National Forest Service this procedure cannot be replicated. We wish that a compromise between data protection and research needs will quickly be found, as well as updating data with more recent inventory information, to help the development of new products based on remote sensing that could also support future field activities and inventory revisits.

**Acknowledgments**

Partial support received from the European Union Horizon 2020 Research and Innovation Program under Grant Agreement No. 640176 (BACI project). G.V.L. acknowledges Japan Aerospace Exploration Agency for providing ALOS2 scenes (P.I. 1012). Data of the second INFI were collected by the National Forest Service (Corpo Forestale dello Stato) and quality checked and processed by the former Forest Monitoring and Management Research Unit of CREA.
References

11. M. A. Lefsky et al., “Lidar remote sensing for ecosystem studies: lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists,” BioScience 52, 19–30 (2002).


Biographies for the authors are not available.