D2.5 Report on Estimation of biomass volume at low productivity MLs.

MAIL: Identifying Marginal Lands in Europe and strengthening their contribution potentialities in a CO_2 sequestration strategy

MAIL project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823805; [H2020 MSCA RISE 2018]

Project title	Identifying Marginal Lands in Europe and strengthening their contribution potentialities in a CO_2 sequestration strategy
Call identifier	H2020 MSCA RISE 2018
Project acronym	MAIL
Starting date	01.01.2019
End date	31.31.2021
Funding scheme	Marie Skłodowska-Curie
Contract no.	823805
Deliverable no.	D2.5
Document name	MAIL_D2.5.pdf
Deliverable name	Report on Estimation of biomass volume at low productivity m/ms MLs
Work Package	WP2
Nature[1]	R
Dissemination[2]	CO
Editor	IABG
Authors	Jesús Torralba, Simonas Garsva and Rewanth Ravindran
Contributors	Luis A. Ruiz and Lindner Martin
Date	29-10-2020

[1] **R** = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other

[2] **PU** = Public, **PP** = Restricted to other programme participants (including the Commission Services), **RE** = Restricted to a group specified by the consortium (including the Commission Services), **CO** = Confidential, only for members of the consortium (including the Commission Services).

MAIL CONSORTIUM

Aristotle University of Thessaloniki (AUTH) Greece	Industrieanlagen Betriebsgesellschaft MBH (IABG) Germany
Gounaris N. – Kontos K. OE (HOMEOTECH) Greece	Centrum Badan Kosmicznych Polskiej Akademii Nauk (CBK PAN) Poland
Universitat Politècnica de València (UPV) Spain	Fundación Centro De Servicios Y Promoción FOrestral Y de su Industria De Castilla y León (CESEFOR) Spain

ABBREVIATIONS

Term	Explanation
AGB	Above-Ground Biomass
AIC	Akaike Information Criterion
ALS	Airbone Laser Scanning
CHM	Canopy Height Model
CO_2	Carbon Dioxide
CLC	Corine Land Cover
DBH	Diameter at Breast Height
DEM	Digital Elevation Model
DSM	Digital Elevation Surface
ESA	European Space Agency
EU	European Union
GEDI	Global Ecosystem Dynamics Investigation
GHG	Greenhouse Gas
GPS	Global Positioning System
GRD	Ground Range Detected
GSV	Growing Stock Volume
ha	Hectares
HRL	High Resolution Layer
IMU	Inertial Measurement Unit
InSAR	Interferometric Synthetic Aperture Radar
IWCM	Interferometric Water Cloud Model
km	Kilometres
LiDAR	Light Detection and Ranging
m	Meters

ML	Marginal Lands
MODIS	Moderate Resolution Imaging Spectroradiometer
N	Number of trees
NASA	National Aeronautics and Space Administration
ºC	Degree Celsius
QMD	Quadratic Mean Diameter
r	Plot Radius
rRMSE	Relative Root Mean Square Error
RADAR	Radio Detection and Ranging
SAR	Synthetic Aperture Radar
SNAP	Sentinel Application Platform
SEE	Standard Error of Estimate
TanDEM - X	TerraSAR-X add-on for Digital Elevation Measurement
TCD	Tree Cover Density
WCM	Water Cloud Model

Contents

EXECUTIVE SUMMARY

Large-scale estimation of biomass and carbon content of vegetation is not simple. The traditional methods for forest biomass estimation, generally manual, are not sufficient to cover the need to have a broad and detailed knowledge of the biomass stored in natural environments (forests, shrublands, ML with vegetation, etc.). RADAR and LiDAR remote sensing sensors have the capacity to record large areas and derive from the data obtained, different forest parameters. These sensors can directly measure parameters such as height or number of individuals in a given area, but they can also indirectly estimate parameters such as wood volume, biomass and carbon content.

Task 2.6 “Estimation of biomass volume at low productivity m/sm MLs” studies different methodologies for biomass estimation on marginal lands using RADAR and LiDAR data. To carry out this study we do not have biomass data from marginal lands themselves but have worked with field data acquired in forests at the test sites of Espadán (Spain), Nogueruelas (Spain) and Thessaloniki (Greece). The RADAR data used are free and were provided by the ESA Copernicus Sentinel-1 and the LiDAR data were acquired through a private aerial system and have been provided by the UPV partner through the [CGL2016-80705-R] project financed by the Spanish Ministry and ERDF (European Regional Development Fund).

The biomass estimation for the test areas has been implemented using 3 different approaches:

i) *Water Cloud Method (WCM).* This method uses the Sentinel-1 C-band and analyzes over a forest area the relationship between backscatter generated at the top of the forest canopy and the backscatter generated in the soil gaps. The workflow starts with radiometric, and geometric correction and speckle filtering with the support of a DEM. Then the non-vegetation areas are masked with the help of the CORINE land cover. Areas defined as forest are classified into dense forest and soil with the use of the Tree Cover Density (TCD) High Resolution Layer (HRL) from Copernicus. In these two areas the parameters of forest backscatter and ground backscatter are estimated. Then the β values and the maximum value of Growing Stock Volume (GSV) are selected to apply the WCM equation. Finally, the WCM values are transformed to Above-Ground Biomass (AGB). To reduce noise in backscatter and reduce the error of AGB estimation, different polarizations

and the combination of images acquired at different times of the year were analyzed.

ii) *Interferometric Water Cloud Model (IWCM).* In this method the backscatter is identified in a similar way to the WCM method but generalized to include gaps in the vegetation cover by the introduction of the area fraction covered by vegetation. To perform the interferogram, two images were selected from September 2015, which is the date when the data was measured in the ground. The images were pre-processed to obtain the "backscatter image" and the "coherence image". The height of the vegetation was obtained by subtracting the Digital Terrain Model (DTM) from the Digital Surface Model (DSM) acquired by the LiDAR dataset that was used as "phase heights". Finally, the IWCM model was calculated and the obtained values were converted to biomass values.

I) *LiDAR.* The methodology used allows to estimate the biomass at plot level from aerial LiDAR data. First, the value of the biomass at plot level was calculated from the field data. At the same time, a pre-processing of the Airborne Laser Scanning (ALS) data was carried out, removing the noise, normalizing the heights, and trimming the clouds according to the size and shape of the plots. Afterwards, the height and intensity metrics ALS per plot were obtained. The different ALS metrics were analyzed using the Akaike information criterion to select the relevant predictors for biomass adjustment. With ALS metrics as independent variables and ground truth biomass values as dependent variables, multiple linear regression models were generated for each study area and species. Finally, the accuracy of the different models was evaluated with different statistics by leave-one-out cross-validation.

To implement the SAR *WCM* methodology, Sentinel-1 C-band, CORINE land cover layers and Tree Cover Density (TCD) High Resolution Layer (HRL) from Copernicus are required, which are free and open access, as well as ESA's SNAP processing software. For the *IWCM* SAR methodology, Sentinel-1 C-band (free and open access), DTM, DSM (depending on the resolution are also free at different scales) and biomass field data (usually involve an acquisition cost) are required. In addition, the free SNAP software and commercial MATLAB software are also required, although it could be programmed in other free languages. To implement the *LiDAR* methodology, field data that generally have a cost, airborne LiDAR data that have an acquisition cost or low-density point clouds that in some EU countries are free, are required. For processing the LiDAR data, LAStools and the Fusion software that is distributed free of charge by the US Forest

Service were used, while for the statistical analysis of the data, the free software RStudio was used.

The results with the *WCM* method had a low precision, generally around 30-80% of rRMSE, mainly due to an early signal saturation – short C-band wavelength has limited penetration leading to loss of signal sensitivity at higher biomass levels (above 100 Mg/ha) under non-optimal environmental and meteorological conditions at the time of image acquisition. On the other hand, this method is the only one transferable to all of Europe, although it is very sensitive to the weather conditions in which the different images were taken. The *IWCM* method improved the precision with respect to the WCM method, reaching an rRMSE of 36% - 48.2%% for some stands. This method is more complex to implement, and its transferability depends on the availability of field data. The *LiDAR* methodology was the one that obtained the better precision. It also obtained biomass estimation equations with an R^2_{adj} of 0.69 to 0.83 depending on the test site and the dominant tree species. Regardless of the good results of LiDAR methodology, and its evident usefulness in biomass estimation, these results are not transferable to other test sites. To be transferable, LiDAR and biomass data should have been available to fit the equations to each test site and each set of species.

1. INTRODUCTION AND GOALS

This document describes the methodology and key aspects for the estimation of biomass volume at m/ms MLs using RADAR and LiDAR remote sensing technologies. The main objective of this task is to show and validate methods to quantify biomass and CO_2 in ML with the use of SAR and LiDAR sensors in the reference areas provided by the ***MAIL*** consortium. To achieve this objective, a literature review of the most frequently used methodologies with RADAR and LiDAR for biomass estimation has been carried out and these methodologies have been implemented in the pilot areas. In this case, the methodologies have been implemented on forest areas with marginal zones inserted in them, these are the only areas available with field data, LiDAR and SAR data acquired on the same dates.

2. BACKGROUND

Forests play a key role in the global carbon cycle by capturing 25% of the carbon emitted into the atmosphere by fossil fuel consumption (S. Saatchi, 2019). At European level, forests are one of the most important renewable resources, and provide a wide range of benefits to society. One of the many benefits of the European forests is that they constitute important carbon sinks capable of absorbing and storing about 10% of the total greenhouse gas (GHG) emissions of the European Union. Considering the importance of forests for climate change, the forestry sector, and those areas susceptible to being forested such as MLs, play a key role in ensuring that the potential for carbon sinks[3] is fully exploited. However, these carbon absorption amounts are not sufficient to mitigate the CO_2 emissions of the EU. So, this is where the marginal lands that can be turned into forest can contribute to the absorption of CO_2.

2.1 Biomass

An extensive literature establishes that through actions such as afforestation, forest management and reduction of deforestation, forests are very efficient carbon sinks (Favero, Daigneault, & Sohngen, 2020). Forest structure is a direct indicator of how carbon is stored in the global ecosystems. This carbon stored in the vegetation has

[3] The definition of a carbon sink given by the 1992 Framework Convention on Climate Change refers to any process, activity or mechanism that absorbs a greenhouse gas, an aerosol or a greenhouse gas precursor from the atmosphere.

numerous positive effects on the functioning of the ecosystem (i.e. carbon, nutrients, water). Measuring the vegetation carbon *in situ* is a very complicated task. Generally, carbon measurement is based on vegetation biomass, which is a primary variable correlated to the quantity of carbon flowing in the carbon cycle (Kaasalainen et al., 2015).

Biomass involves above-ground biomass (AGB) and below-ground biomass. Above-ground biomass represents all living biomass above-ground including stems, branches, bark, leaves and seeds of trees and shrubs, while below-ground biomass is made up of all roots except the smallest roots (FAO, 2004). Biomass can be measured in the field using destructive techniques, very accurately, but with a high cost in terms of time and finance (Zhang & Ni-meister, 2014). However, these measurements are the tool to establish predictive models, to evaluate models made with other techniques and to validate the accuracy of estimated biomass values. *In situ* biomass data are obtained by destructive methods applied on an individual tree or on a reference area (plot). This method involves harvesting the plants, drying them, and then weighing the biomass. Once the biomass values are obtained, the allometric relations between the biomass and specific tree attributes are established, essentially using height, Diameter at Breast Height (DBH) or forest cover (Montero, Ruiz-Peinado, & Muñoz, 2005; Ruiz-Peinado Gertrudix, Montero, & Del Rio, 2012). These tree attributes are generally easy to acquire in the field and allow for relatively accurate biomass estimates, and they are typically collected on samples plots designed for a specified study or they are compiled in National Forest Inventories (Zhang & Ni-meister, 2014). However, these measurements are not at forest level and it is difficult to extrapolate the plot estimates to a larger area. In addition, from the small area of plots where biomass is estimated, land use activities, together with increasing climatic disturbances and human pressure on the environment, are changing rapidly the requirements for forest inventories by plots that include more plots and more frequent observations of forest ecosystems (S. Saatchi, 2019).

2.2 Remote Sensing and Biomass

Remote sensing has been widely used as a strong tool in forest structure analysis and biomass estimation because it provides information at local, continental and global scales (Zhang & Ni-meister, 2014). A suite of remote sensing sensors provides measurements of structural and biophysical characteristics of forests based on the interaction of light or microwave energy with forest canopies and woody components (S. Saatchi, 2019). Three types of remote sensing data are commonly used: optical images, radar images and LiDAR data.

Typically, sensors are categorized into passive sensors, which measure the energy reflected or emitted from the earth's surface, and active sensors that generate their own energy and measure the attributes of the energy returning from the surface. Passive sensors measure different wavelength ranges in the optical and microwave spectrum, providing two-dimensional information directly related to the biophysical properties and health status of vegetation (Shugart, Saatchi, & Hall, 2010). In forestry, the reflectance in the optical spectrum is sensitive to the forest structure (position of the trees, tree density, crown size, leaf area), texture and shading, attributes closely related to biomass estimation (Zhang & Ni-meister, 2014). On the other hand, active sensors are designed to operate at a specific wavelength, in particular LiDAR in the visible or near infrared and RADAR in microwave long wavelengths (Shugart et al., 2010). In biomass estimation, radar images link dielectric and geometric properties of a forest (Le Toan et al., 2011). LiDAR is able to characterize the vertical structure of the vegetation and the height of the trees which are variables associated to the biomass estimation (Ruiz, Hermosilla, Mauro, & Godino, 2014; Ruiz, Recio, Crespo-Peremarch, & Sapena, 2018).

Over the last 20 years, remote sensing techniques have been studying in detail to accurately assess the characteristics of the forest and in particular the study of the biomass, e.g. Le Toan et al. (2011); Hermosilla et al. (2014); Zhang y Ni-meister, (2014); Kaasalainen et al. (2015); Silva et al. (2017); Zhao et al. (2018) and Saatchi, (2019). In this document a brief summary of active remote sensing sensors to characterise forest structure and biomass (Figure 1), dividing them into two categories according to the measurement capacity of the sensors has been performed. The first type of sensor refers to direct observation and measurement as performed by LiDAR. The height measurements are the relatively direct from laser altimetry from air or space, as are the relatively direct angle and distance measurements made with clinometers from the ground. There are different studies where it has been proved that tree height can be measured with the same precision, and even better, with LiDAR than with manual measurements, as well as measurement errors can be treated in the same way as field measurements (Dubayah & Drake, 2000; Wulder, Bater, Coops, Hilker, & White, 2008; Wulder et al., 2012). The second category of active sensors would be the RADAR images, which provide indirect measurements of forest volume, biomass and height. In this case the backscatter measurements of RADAR images have a strong sensitivity to forest structure and biomass. The sensitivity of radar images depends on the length of the microwaves used, for example, with L-band the sensitivity is reduced when the biomass increases in the range of 100 to 150 MG/ha and with P-band the sensitivity is lost when the biomass per hectare is between 200 - 300 Mg (Le Toan et al., 2011; S. Saatchi, 2019). If interferometric radar techniques are combined, the sensitivity of the biomass estimate can be increased, as well as the gaps in the crown cover, the structure and the spatial heterogeneity of the forest can be determined with greater precision.

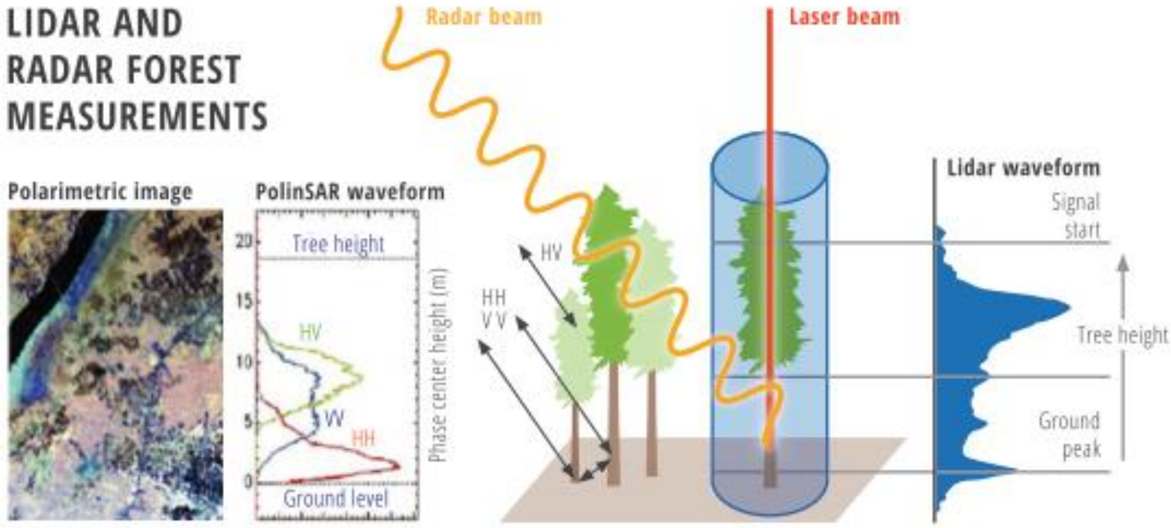

Figure 1. Radar and LiDAR can capture the horizontal and vertical structure of forest ecosystems. Image from Saatchi (2019).

Remote sensing techniques with LiDAR and RADAR are now recognised as the best methods for quantifying and monitoring changes in forest AGB worldwide. Proof of this are the numerous missions of space agencies such as ESA and NASA, which are investing a large amount of their efforts in launching LiDAR (e.g. GEDI) or RADAR (Sentinel-1 and Sentinel-2) sensors for global biomass monitoring. The following subchapters provide a more detailed description of RADAR and LiDAR.

2.3 SAR Remote Sensing Biomass Inventory

Forest above ground biomass (AGB) is mainly estimated by either traditional field measurements or by remote sensing methods (Maurizio Santoro & Cartus, 2018). Ground based field measurements provide the most comprehensive and detailed way of calculating biomass. However, the availability of measurements is limited in extent. Far more extensive data can be obtained from remote sensing methods, such as from multispectral and SAR satellites, over regional or continental scales compared to field plots and providing a more spatially comprehensive measure of forest biomass related parameters. However, direct estimation of biomass is not given by these satellite measurements (Jan I.H. Askne, Soja, & Ulander, 2017), only measurements of some forest or vegetation related signals that can then be used in models to convert to biomass estimation e.g. AGB.

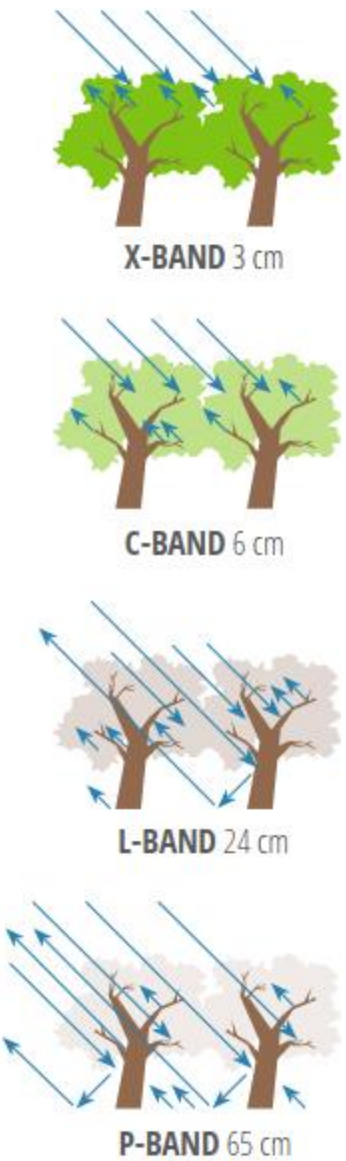

Figure 2. Sensitivity of different SAR wavelengths in the measurement of forest structure and the penetration of the wave through the canopy. Image from Saatchi (2019).

SAR is a type of radar measurements made by satellites in the microwave region of the spectrum, at multiple frequency bands. As an active type of remote sensing, SAR has the ability to penetrate the canopy and interact with the main biomass components, i.e., the tree trunks and branches. SAR backscattering intensity increases as forest biomass increases (S. Saatchi, 2019).

SAR has different sensitivity to forest biomass according to its wavelength.

As the wavelength increases, the scattering saturation value increases, and the correlation between the backscatter and biomass also increases; thus, long wavelength (or lower frequency) bands are more suitable for biomass estimation. Backscatter at low frequencies is sensitive to the major AGB components, i.e. stems and large branches, which is not the case for high radar frequencies (typically, C- and X-band). However, the latter can serve as a complement, in particular for less dense forests, since the backscattering relationship to AGB at low frequencies is complex and is also affected by ground and soil properties (Jan I.H. Askne et al., 2017).

However, radar sensitivity to AGB values changes depending on the wavelength and geometry of the radar measurements and is influenced by surface topography, structure of vegetation, and environmental conditions such as soil moisture and vegetation phenology or moisture (S. S. Saatchi, Le Vine, & Lang, 1994). All algorithms or models used to estimate AGB from SAR measurements must account for all variables that impact SAR measurements. For this, different information types from SAR measurements can be used: Backscatter, Polarimetry, and Interferometry.

2.3.1 Backscatter

The impact of vegetation structure and biomass on SAR data can be investigated by modeling the dominant scattering mechanisms controlling the SAR measurements. A variety of approaches exist for modelling vegetation media, including the characterization of forest vegetation structure, known as scatterers, or scattering components such as stems, branches, and leaves (S. Saatchi, 2019).

SAR backscatter sensitivity to AGB at any frequency depends on

a) Measurement geometry (such as incidence angle and location and size of the image pixels with respect to the size and the orientation of ground plots).
b) Forest structural parameters (such as the size (volume) and density of trees (number per resolution cell), orientation of forest components (leaves, branches, stems), underlying surface conditions (moisture, roughness, and slope)).
c) The dielectric constant that in turn depends on the vegetation water content or specific gravity (i.e., the wood density).

2.3.2 Polarimetry

The backscattering coefficient measurement by SAR systems can be expressed as the combination of three scattering components volume (*vol*) scattering, volume and surface interaction (*vol-surf*), and surface scattering (*surf*) as shown in the following equation:

Equation 2-1 $$\boldsymbol{\sigma^0_{pq} = \sigma^0_{pq-vol} + \sigma^0_{pq-vol-surf} + \sigma^0_{pq-surf}}$$

where *pq* could be HH, HV, *VH* or *VV* polarizations (S. Saatchi, 2019).

Polarization is therefore the key characteristic of radar signals propagating into tree canopies or vegetation volume and scatter from individual vegetation components that collectively contribute to the backscatter energy measured by the radar receiver system. Polarization as the orientation of radar wave vectors (at *H, V,* or any other polarization) interact with vegetation components and backscatter according to the size and orientation of scatterers. For example, a standing live tree with near-vertical orientation depolarizes the incoming waves with different strengths than branches or leaves. Using radars that provide measurements in different polarizations allows separate vegetation with different structures to be reflected in the average size and orientation of different components (S. Saatchi, 2019).

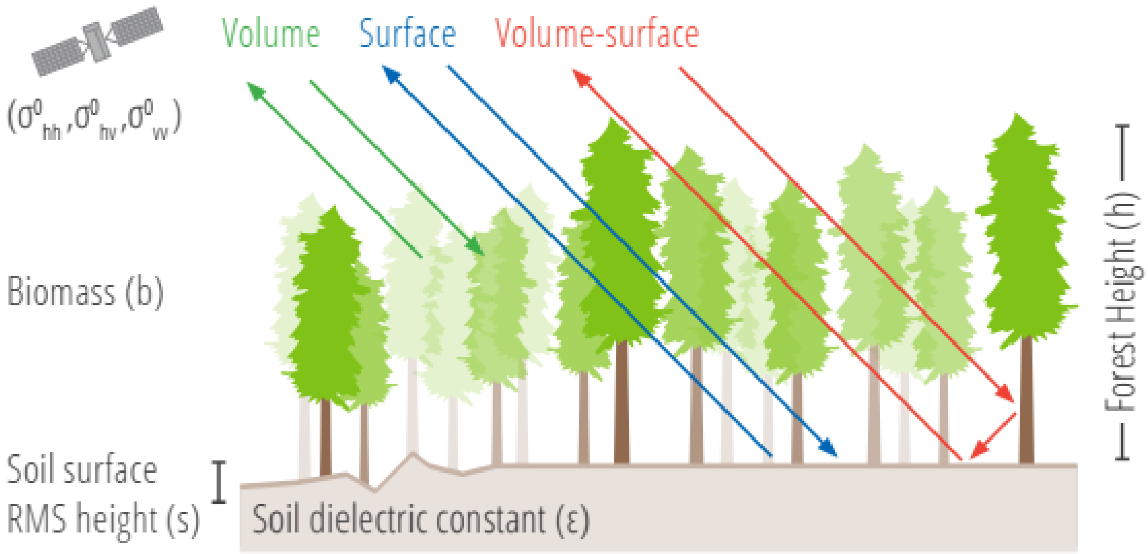

Figure 3. Dominant scattering mechanisms of L-band SAR measurements in a forest area contributing to polarimetric backscatter observations. Image from Saatchi (2019).

2.3.3 Interferometry

SAR Interferometry (InSAR) can be created from the phase information from two subsequent observations by SAR satellites. InSAR seems particularly useful for AGB estimation, for instance where InSAR heights are used in combination with terrain information (DTM) to estimate forest canopy height and, subsequently, AGB (Ulander, Hagberg, & Askne, 1994).

The InSAR height of short wavelengths is often assumed as an approximation of canopy surface height, although a substantial penetration is observed depending on the frequency. Hence, for a precise estimation of the true canopy height, penetration effects need to be corrected (Jan I.H. Askne et al., 2017). Different canopy (e.g., forest structure, moisture, etc.) and acquisition (e.g. incidence angle) parameters affect the penetration, which can differ by a few meters. Consequently, canopy heights estimated from InSAR heights can have substantial bias, and the estimates derived from the same sensor but under different conditions cannot be compared. Therefore, penetration effects must be removed in order to compute consistent and unbiased height estimates that can subsequently be used to estimate AGB and allow a robust estimation of changes in canopy height and AGB. This is done by using physical models for different forest types to estimate the penetration depth and make corrections. In particular, *TanDEM-X* satellite data is mostly considered suitable for this method (Jan I.H. Askne et al., 2017).

Additionally, coherence data from the InSAR is also used in allometric equations and other biomass estimation methods. The relationship between coherence and biomass has been described by statistical, empirical, and physically based models, some of which will be explored in the later sections.

2.4 LiDAR Remote Sensing Biomass Inventory

2.4.1 LiDAR technology brief

ALS (Airborne Laser Scanning) or aerial LiDAR is an active remote sensing sensor that measures the distance to an object using the time-of-flight measurement principle. LiDAR sensors emit a laser pulse and record the time it needs for the energy pulse to impact the object and return to the instrument. The time measurement is then converted into a distance based on the following equation:

Equation 2-2
$$Distance\ (m) = \frac{Speed\ of\ Light \cdot time\ to\ flight}{2}$$

The *x,y,z* position of the object is defined by the known position of the sensor and the precise orientation of the range of measurement between the sensor and the object that intercepts it. Regardless of which platform is used to acquire ALS data, the principles of measurement are identical. The most commonly used platforms in the acquisition of ALS data for forest inventories are fixed wing aircraft or helicopters (Figure 4), although recently the GEDI sensor was launched into orbit with the purpose of studying the evolution of forests, among other things.

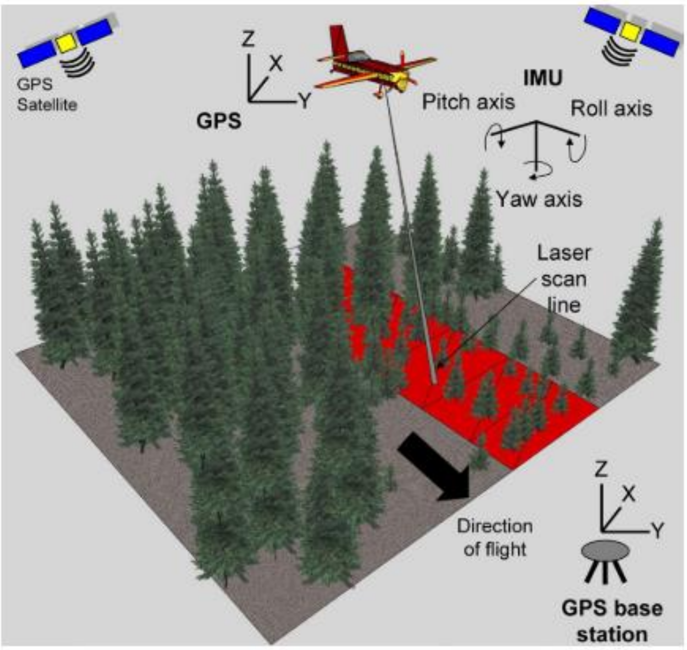

Figure 4. ALS system package. The instruments included are a laser ranging unit; an opto-mechanical scanner; control, monitoring, and recording units; a kinematic global positioning system (GPS) receiver; and an inertial measurement unit (IMU) (Wehr & Lohr, 1999). Image from White et al. (2013).

ALS sensors directly measure the vertical distribution of the forest canopy components as well as the ground topography, resulting in an accurate estimation of the vegetation height and ground elevation. There are two types of ALS sensors working in near-infrared, discrete return, and full-waveform. Full-waveform ALS systems register the reflected (backscattered) energy from each laser pulse as a single or continuous signal (Crespo-Peremarch, Ruiz, Balaguer-Beser, & Estornell, 2018). Discrete ALS converts waveform data into spatially and temporally referenced targets. In this work we will focus on the discrete ALS since it is the available ALS in the pilot areas. A discrete ALS return system records up to 5 returns for each laser pulse it emits. A simple case is when a laser pulse intercepts an object which cannot penetrate such as a very dense forest canopy or an asphalt road, and results in only a return of energy to the instrument. On the opposite, when the pulse intercepts an object that it can penetrate, such as a not excessively dense forest canopy, a first return will occur with part of the energy and the other part of the energy will continue through the canopy and intercept stems, branches, and leaves until reaching the ground (White et al., 2013). This sequence of iterations gives multiple returns for a single laser pulse and produces very useful information about the vertical structure of the forest. Typically, the first returns correspond to the tree cover while the last returns correspond to the ground or objects close to the ground. The intermediate returns would refer to the different parts of the tree depending on the species, height, and physiology condition. When the vegetation is dense, the trunks, branches and leaves tend to cause a multiple dispersion of the emitted laser energy so that less returns are generated from the ground, making it more difficult to generate an accurate DEM. As the density of the vegetation increases, the depth of the canopy and the structural complexity of the forest increase, fewer pulses reach the ground and the DEM accuracy decreases.

The first step of the ALS data post processing consists of generating a single file called “point cloud” where each return has a precise, georeferenced and three-dimensional (x,y,z) location as a result of gathering the data acquired by ALS sensor, the kinematic global positioning system (GPS) and the inertial measurement unit (IMU) (Figure 4). Each of the returns depending on the material against which it impacts, will return to the sensor with a certain energy, called intensity, which is also recorded individually for each return. In addition to the x, y, z position of each return, the intensity will allow to distinguish between the different components of the trees at the same height.

The point cloud is then processed to first identify the ground points and the no-ground points. The ground points generate a precise DTM (representing the height of the ground in relation to some reference) and the no-ground points corresponding to the first returns generate a DSM (representing the heights of the objects in relation to the ground surface). Once both are obtained, the difference between the DTM and DSM values in each cell allows the generation of the digital model of surfaces normalized with respect to the terrain (CHM), also called canopy height model in forestry applications.

These three products derived from the point cloud can be used as inputs in the pre-processing of images acquired with other sensors such as RADAR images. In the same way, DEM is used to normalize the ALS point cloud at heights above ground.

2.4.2 LiDAR Biomass models

Like most forest inventory techniques, ALS measurements can be utilized for mapping and inventory of forest structures. In general, in most forestry applications some measurement of height, height variability or crown cover will be required (Figure 5). Numerous height and density measurements of the forest can be generated from the ALS data due to the strong relationship between the ALS height measurements and the forest parameters. In the same way that AGB is estimated on the ground, the allometric models derived with ALS can vary from one place to another, capturing the differences in tree growth, diameters, and heights in the whole forest. Some allometric models for estimating the AGB using ALS metrics show significant variations for the same forest. This uncertainty in the calculation of the AGB for larger areas can be reduced by using multiple height metrics. However, yet there is no universal model for converting ALS height measurements into full-scale AGB.

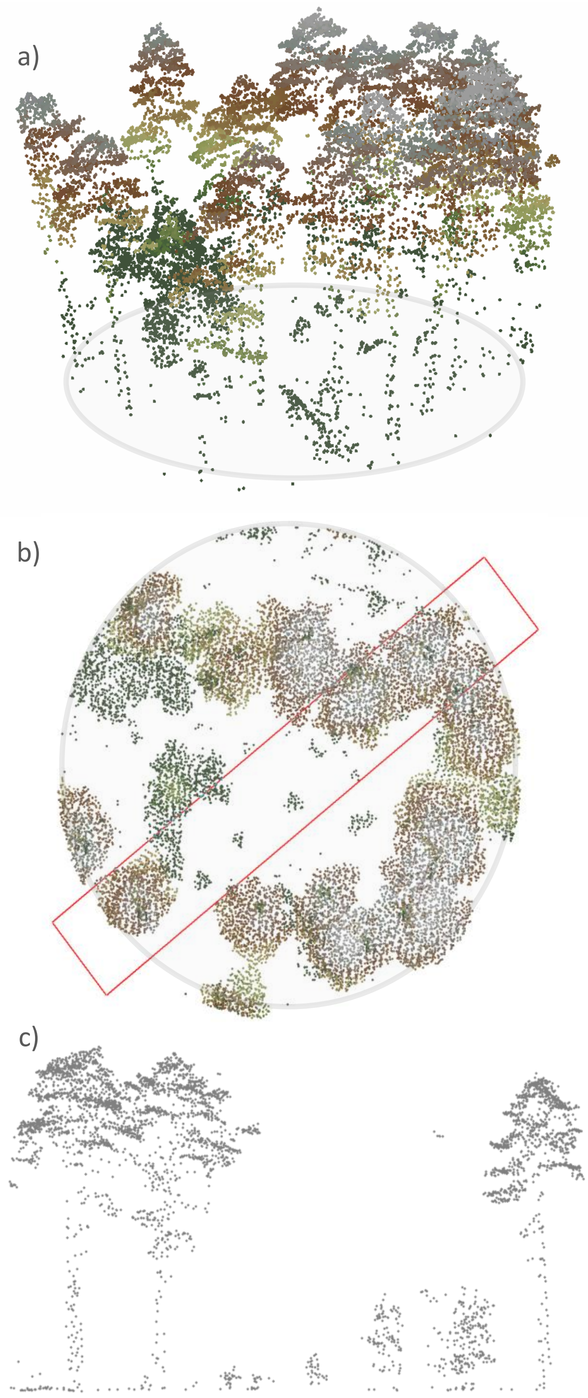

Figure 5. ALS point cloud of plot 15 in the Nogueruelas study area in Spain. a) General view of the point cloud, b) zenith view where the canopy cover can be observed, and c), vertical profile of the forest structure of the plot.

Regardless of the type of LiDAR system used, the estimation of biomass is generally made based on statistical models that relate the AGB to the metrics derived from the ALS system. These assessments are made at different scales, from the estimation at the level of individual trees to the scales of plots and stands. At plot level, the biomass measured on each tree is aggregated into a single value per plot, thus obtaining the reference biomass value for each plot (independent term of the equation). On the other hand, height and intensity metrics derived from ALS data have been prepared. Three main groups of characteristics can be distinguished (Ruiz et al., 2011; Ruiz, 2020):

a. Those *extracted directly from the point cloud* of each plot, which can be either:

  a.1 *Statistical variables of the distribution in height of the points*: as the mean, standard deviation, maximum, skewness and kurtosis of each plot from the normalized point cloud. These characteristics are complemented by the values of the height percentiles. The statistical variables of the height distribution of the LiDAR points provide information about the internal structure of the vegetation.

  a.2 *Variables derived from the density profiles:* From the point cloud, density profiles can also be generated, i.e. height histograms of the set of points within each plot.

b. Variables extracted from the standardized surface model (CHM): The calculated characteristics of the CHM provide information about the maximum height values in each cell or pixel and their spatial distribution within the plot.

c. *The variables derived from the LiDAR intensity data:* In addition to the x, y, z coordinates, each LiDAR return contains information about the intensity of the radiation reflected at that point, that is, spectral information corresponding to the wavelength of the system (NIR). Similar to the height statistical variables, statistical values can be obtained from the distribution of intensities (mean, standard deviation, percentiles, skewness, etc.). In general, the addition of return intensity values to regression models provides a better estimation of biomass than height metrics alone.

ALS metrics can be obtained with different software, for example with the FUSION *cloudmetrics* command (McGaughey, 2016). This command produces more than 90 unique metrics from the heights and intensities measured in the ALS point cloud.

Models can be developed for species or species groups (Domingo, Lamelas, Montealegre, García-Martín, & de la Riva, 2018; Ruiz et al., 2014) or, can be generalized to forest types (Ruiz et al., 2018; Zhao et al., 2018). Different approaches can also be used to build the prediction models, but the most common way to do this is either with parametric methods or with non-parametric methods. Models generated with parametric approaches are characterized by having a finite number of parameters and by making assumptions about the relationship between response and predicted variables (White et al., 2013). This parametric regression approach has been widely employed in the construction of predictive models of forest inventory attributes. On the other hand, the most common non-parametric approach applied to forest inventories with ALS is Random Forests. This method consists of a decision tree approach based on successive regressions (Breiman, 2001). Both approaches have advantages and disadvantages as shown in Table 1, however, the selection of the best method for a given area depends to some extent on the complexity of the forests in the area of interest and the statistical knowledge available for modelling. Finally, the predictions of parametric and non-parametric models should be limited to the range of observed data used to calibrate the model. Both parametric and non-parametric approaches require representative data from the field for the development of a robust model. A Random Forest model cannot extrapolate the predictions, so it requires that the ground samples be distributed evenly in the *X* and *Y* space.

	Parametric Regression	Random Forests
Advantages	-Easy to understand. - The sample size can be determined for certain accuracy and precision requirements. - The model is an equation that clearly quantifies the relationship between the predictor and the predicted variable.	- Category variables can be used as predictors and can also be predicted. - Faster development and implementation than parametric methods - It does not require individual models to be developed based on strata, as long as the calibration data represent the different strata involved - To implement stratum-based models does not require a pre-existing polygon-based inventory

	Parametric Regression	Random Forests
Disadvantages	- The interpretation and application of the regression can be complicated by the transformation of ground measurements or ALS metrics that may be necessary to meet the assumptions of regression-based approaches. - More time and statistical expertise are necessary to create the models - In order to make models of different forest strata it is necessary to have an inventory layer where the different forest types are stratified - "Prediction errors will occur within polygons when individual grid cells do not match the overall strata assignment"	- The development of the models is a black box - No equation output analogous to the parametric regression - As this approach does not extrapolate as a regression it is more difficult to ensure that the full range of conditions is sampled

Table 1. Advantages and disadvantages regarding parametric regression and random forest modelling approaches in the context of calculating forest variables with ALS. Adapted from White et al. (2013)

In this task it was decided to work with parametric models and more specifically with multiple linear regressions due to the characteristics of the pilot sites used.

3. Pilot Sites and Ground Truth Data

The selection of data for this task has been conditional on the availability of data by the consortium partners. For the estimation of biomass with RADAR and LiDAR it is necessary to have biomass data derived from allometric equations using forest parameters with DBH or tree height. To work with LiDAR, data is only available in the pilot areas proposed in Spain, and from those areas only two: Espadán and Nogueruelas. These areas are the only ones that have field and LiDAR data acquired simultaneously. To carry out the work with RADAR, in addition to the Spanish pilot areas, an area in Greece has been added in which biomass data is available. The following is a description of each of the pilot areas and the ground truth data acquired.

3.1 Espadán, Castellón, Spain.

The study area covers 3,741.5 ha and is located in the Natural Park of Sierra de Espadán, in the eastern Spain province of Castellón (Figure 6). “This natural park is a Mediterranean forest with soft and rounded hills, presence of abandoned farming with artificial terraces, and mountain peaks up to 1,100 meters of altitude. The European Environment Agency report from 2016 (Bastrup-Birk, Reker, & Zal, 2016) classified this area as a semi-natural forest with a natural function, composition and structure, but modified by human activities throughout history. Forest type and conditions, and species composition have been influenced by human needs and changes in land use, as well as reforestation of single species policies from the last century” (Torralba, Crespo-Peremarch, & Ruiz, 2018).

This area displays a heterogeneous landscape dominated by pure and mixed native coniferous and deciduous forests. “The most dominant species in the area is *Pinus halepensis* Mill., which mainly forms pure stands with different even aged and densities. Density of *P. halepensis* stands ranges from overstocked stands with small sapling (10,000 to 45,000 trees/ha) to poorly and medium stocked stands with young and high forest (300 to 2,500 trees/ha). *P. pinaster* Aiton is the second most represented species in the area, forming pure stands with densities ranging from 800 to 1,250 trees/ha, and mixed stands with *Quercus suber* L. as codominant species at the upper strata, ranging from 500 to 1,200 trees/ha. *Quercus ilex* L. shows up in punctual places forming pure stands and sometimes mixed with other species such as pines or oaks. In some areas, mixed stands are observed, where *P. pinaster* dominates the upper strata, while *Q. suber* and *Q. ilex*, and *Juniperus thurifera* L. are codominant species with densities between 500 and 800 trees/ha” (Torralba et al., 2018).

“Understory vegetation presence and density are very heterogeneous in this ecosystem and depend on the tree composition. Forest stands dominated by *P. halepensis* have taller and denser understory vegetation than those dominated by *P. pinaster* and *Q. suber.* The most common genera of the understory species are *Erica, Genista, Rhamnus, Pistacia, Juniperus, Rosmarinus, Quercus, Phillyrea, Daphne* and *Thymus”* (Torralba et al., 2018).

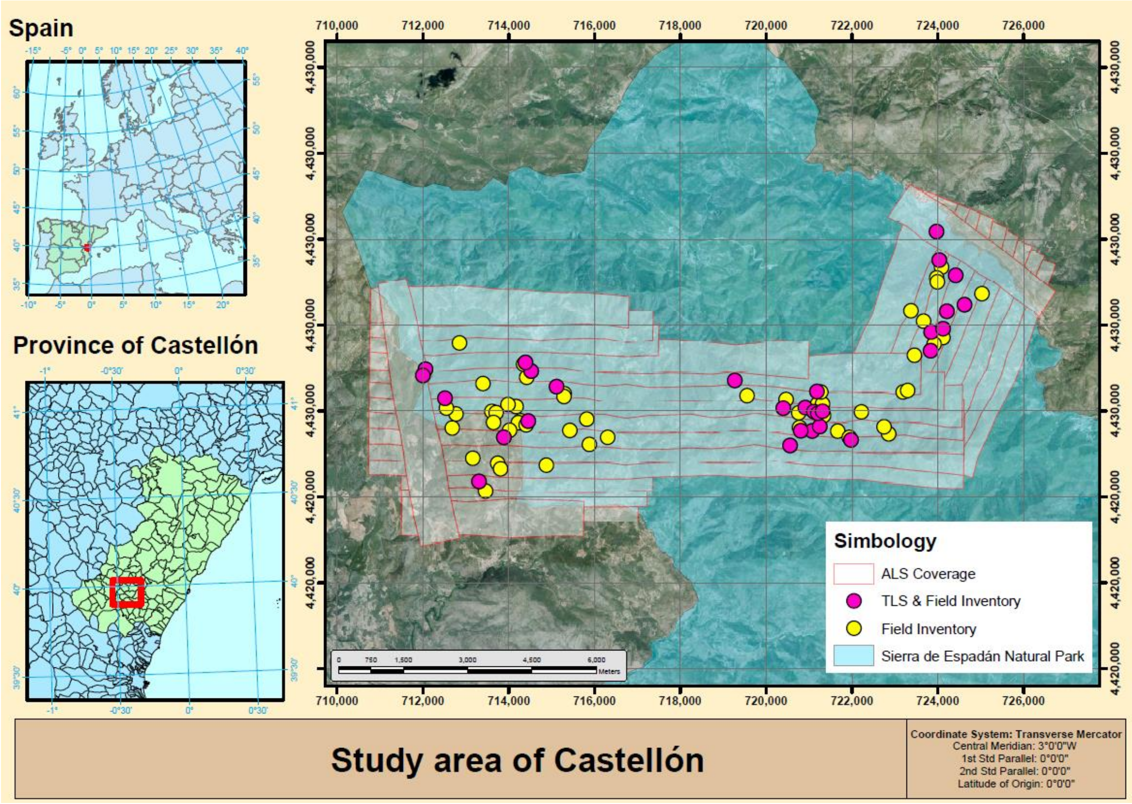

Figure 6. Espadán (Castellón, Spain) pilot site.

3.2 Nogueruelas, Teruel, Spain.

The study area covers 1,900.6 ha and is located north of the municipality of Nogueruelas (Teruel) about 65 km from the city of Teruel (Figure 7). This is an eminently forested area located in the heart of Sierra de Gúdar. The altitude of the study area ranges between 600 and 1,800 meters above sea level. The slopes in the study area are gentler than in the surrounding environment due to the fact that the mountain is located in areas of high plateaus, with the appearance of gentle slopes. The climate is strongly conditioned by the relief of the area. The precipitations are scarce, oscillating between the 500 and something more than 700 mm annual in the summits, many falls in the form of copious summer storms and another part in the form of snow. In general, average temperatures are around 7-9 °C, with short summers and long, harsh and very dry winters.

This area displays a homogeneous landscape dominated by pure coniferous forests where the dominant species is *P. sylvestris L.* (Scots pine). *P. sylvestris* forms pure stands with different even aged and densities ranging from 250 to 1,750 trees/ha, sometimes mixed with other species such as *P. nigra* Aiton. Under the tree canopies, understory vegetation is sparse, but in areas which are open, a dense shrub of *Juniperus sabina* L. (Savin juniper) and *Juniperus communis* L. (Juniper) appear.

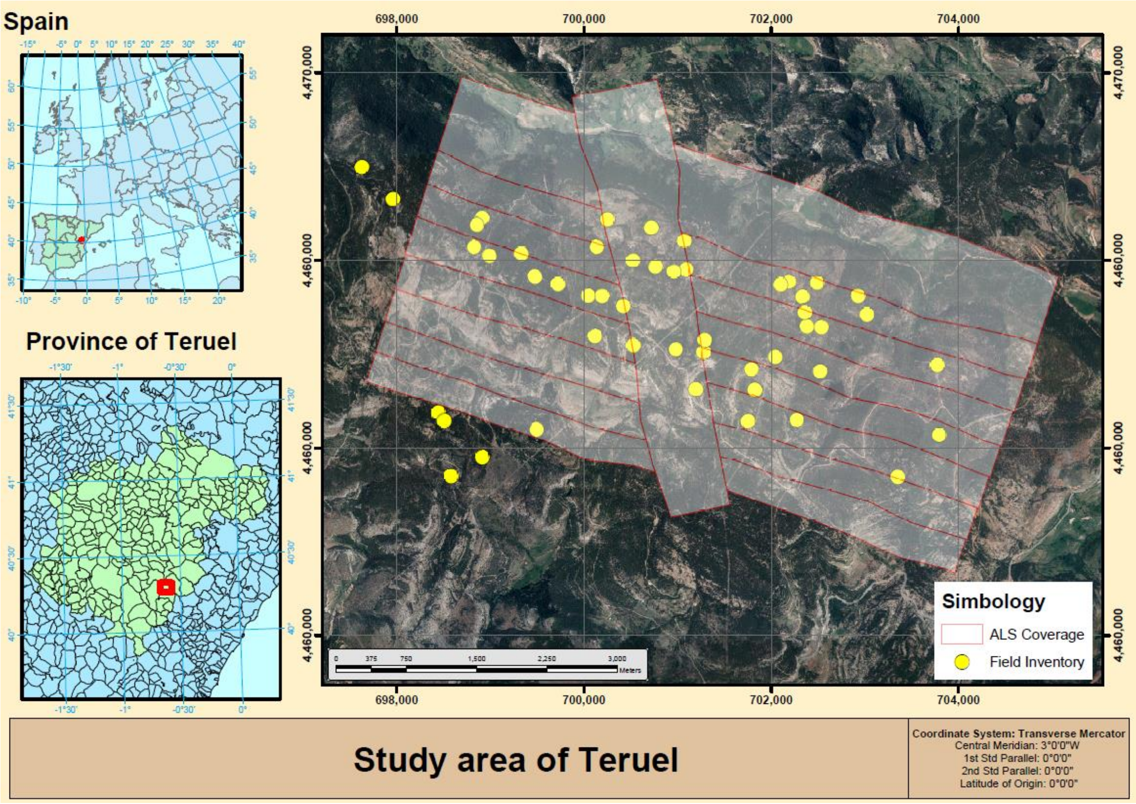

Figure 7. Nogueruelas (Teruel, Spain) pilot site.

3.3 Greece.

The testing site of Thessaloniki covers a total area of 96.63 Km^2 and is located nearly 15 km east of the city of Thessaloniki. The altitude varies significantly from 70 m (the relatively flat lowland area in the south east which includes cultivated areas) to 1,100 m (the mountainous area in the north west which includes low vegetation areas and natural material surfaces) above sea level. The area is mainly characterized by heathland, forest, and cultivated areas.

This study area is divided into a natural forest zone and a reforested forest zone. The natural forest zone is dominated by evergreen broadleaf and in the reforested zone the main species are *Pinus brutia* and *P. halepensis.*

3.4 Ground Truth Data

3.4.1 Spain pilot sites

In the Espadán area, 80 circular plots distributed throughout the study area were collected in September 2015 (r = 15 m, a = 706 m^2). On each plot the following attributes were measured: Diameter at breast height (DBH) from trees with a value above 5 cm, height and canopy base height from the seven dominant trees in each plot, tree species, and percentage of understory vegetation cover. In the other hand, in Nogueruelas were collected 47 circular plots (r = 14.1 m, a = 625 m^2) distributed throughout the study area in October and November 2014. Data collected from each plot included DBH from trees with a value above 5 cm, height from the two dominant trees in each plot, height from five trees representing the average height of the plot, and percentage of understory vegetation cover.

Table 2 shows summary statistics (number of trees, mean, standard deviation, min, and max) of the metric sample plot parameters for both areas.

	Espadán				**Nogueruelas**			
Of sample Plots	80				47			
Plots radio (m)	15 – 3.5				14.1			
DBH range (cm)	5.0 – 82.0				7.5 – 49.5			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
N/Plot (trees/plot)	68	51	2	380	49	21	16	107
N/ha (trees/ha)	3,798	9,093	28	43,914	776	340	256	1,713
QMD (cm)	19.1	6.9	5.2	29.7	20.2	4.1	29.7	12.0
Height (m)	12.1	4.7	1.5	25.5	11.7	2.5	18.6	7.6
Biomass (tons/Plot)	5.81	4.25	0.02	19.40	6.04	2.62	0.75	11.23
CO_2 (tons/Plot)	10.66	7.84	0.05	36.38	11.29	4.90	1.04	20.98

Table 2.Summary statistics of sample plots. N/Plot, number of trees per plot; N/ha, number of trees per hectare; QMD, quadratic mean diameter; Height, height from the dominant trees in each plot; Biomass, tons of biomass per plot; CO_2 tons of carbon dioxide accumulated per plot.

The equations of Montero et al. (2005) were used for the estimation of the aerial biomass for each plot. To apply these equations only the species and the DBH need to be defined. The general equation planted by Montero et al. (2005) to calculate the dry biomass in kg is:

Equation 3-1

$$Biomass = e^{\frac{SEE^2}{2}} \cdot e^{a} \cdot DBH^{b}$$

where SEE is standard error of estimation for each species, *a* and *b* are parameters obtained from Table 2 of Montero et al. (2005) for each species.

From the quantification of the biomass for the tree set, the carbon dioxide stored in each plot was calculated. For this purpose, it is necessary to know the percentage of carbon in the dry matter, which is defined for Mediterranean species in the Iberian Peninsula in Montero et al. (2005). For example, for *P. halepensis* the carbon content of the dry matter is 49.9%, for *P. pinaster* is 51.1%, *P.sylvestris* and *P.nigra* contain both 50.9%, while for *Q.ilex* is 47.5% and for *Q.suber* 47.2%. By means of the ratio among the weight of the CO_2 molecule and the weight of the C atom that composes it, we obtain the ratio that will be used to go from kg of C to kg of CO_2 equivalent (44/12 = 3.67). Thus, multiplying the modular values of biomass by the carbon content and by 3.67 we obtain the modular values of CO_2 for each tree according to the species. Table 3 shows the percentage by weight of carbon contained in the dry matter applied to each species proposed by Montero et al. (2005).

Specie	% carbon	Specie	% carbon
Abies alba Mill.	50.6	*Pinus halepensis Mill.*	49.9
Abies pinsapo Boiss.	50.0	*Pinus nigra Arn.*	50.9
Alnus glutinosa L.	50.0	*Pinus pinaster Ait.*	51.1
Betula spp.	48.5	*Pinus pinea L.*	50.8
Castanea sativa Mill.	48.4	*Pinus radiata D. Don*	49.7
Ceratonia siliqua L.	50.0	*Pinus sylvestris L.*	50.9
Erica arborea L.	50.0	*Pinus uncinata Mill.*	50.9
Eucalyptus spp.	47.5	*Populus x euramericana (Dode) Guinier*	48.3
Fagus sylvatica L.	48.6	*Quercus canariensis Willd.*	48.6
Fraxinus spp.	47.8	*Quercus faginea Lamk.*	48.0
Ilex canariensis Poir.	50.0	*Quercus ilex L.*	47.5
Juniperus oxycedrus L. / J. communis L.	50.0	*Quercus pyrenaica Willd.*	47.5
Juniperus phoenicea L./ J. sabina L.	50.0	*Quercus robur L./Q. petraea Liebl.*	48.4
Juniperus thurifera L.	47.5	*Quercus suber* L.	47.2
Laurus azorica (Seub.) Franco	50.0	Other coniferous	50.0
Myrica faya Ait.	50.0	Other broadleaved	50.0
Olea europaea var. sylvestris Brot.	47.3	Other laurel forest	50.0
Pinus canariensis Sweet ex Spreng.	50.0		

Table 3. Percentage by weight of carbon contained in the dry matter applied to each species. Adapted from Montero et al. (2005).

Due to the great variability of forest structures and densities in the plots in the Espadán study area, it was decided to stratify the plots according to the criteria defined by Torralba et al., (2018). This criterion is based mainly on tree density and basal area resulting in a categorization of the plots in pure and mixed. Table 4 shows the two groups of plots for which the biomass estimation models have been constructed. A group defined by the pure plots of *P. halepensis*, without those plots with a density of more than 10,000 trees per hectare, and another group of pure and mixed plots where the dominant species are *P. pinaster* and *Q. suber.* The 3 pure *Q. ilex* plots have been discarded due to not having a representative sample of plots to elaborate the models. In the Nogueruelas study area it is not necessary to stratify by type of structure since the set of plots maintain similar forest attributes.

Espadán				
ID	**Species**	**Type**	**Nº Plots**	**Plots**
T1	*P.halepensis*	Pure	42	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 30, 33, 34, 45, 46, 47, 48, 49, 50, 56, 58, 59, 62, 63, 64, 65, 69, 70, 71, 72, 74, 75, 77, 80
T2	*P. halepensis* regenerated (>10.000 trees/ha)	Pure	9	51, 52, 54, 55, 57, 60, 66, 67, 68
T3	*P. pinaster and Q.suber*	Mixed	25	11, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 61, 76
T4	*Q.Ilex*	Pure	3	73, 78, 79

Table 4. Clusters of points according to species composition and tree density.

3.4.2 Greece pilot site

The Thessaloniki field data was acquired in July 2007. A total of 42 plots were distributed, 17 in the natural forest areas and 25 in the afforested forest areas. In the plots the species were identified, the height, DBH and age of each tree were measured. The main characteristics of the forest in the area were collected in Table 5.

	Natural Forest				Afforestation			
Of sample Plots	17				25			
DBH range (cm)	3.0 – 7.1				3.7 – 8.4			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
N/Plot (trees/plot)	-	-	-	-	1,225	520	350	2,550
Height (m)	2.5	0.5	1.8	4.0	6.2	1.5	3.7	8.4
Age (years)	8	4	5	20	12	2	8	18
Volume (m^3)	15.3	8.2	5.6	31.4	27.3	21.2	3.7	89.3
Annual increase (cm)	0.8	0.4	0.5	1.9	2.0	0.8	0.4	3.8

Table 5. Summary statistics of sample plots. N/Plot, number of trees per plot; Height, height from the dominant trees in each plot; Age, average age per plot; Volume, volume of wood stocks on the plot; Annual increase, Annual growth in centimetres.

4. METHODS

4.1 Sentinel-1 data-only based method - the WCM

The first biomass estimation method is based on the use of ESA's Sentinel-1's C-band SAR data. The model calibration is aided using additional the Pan-European Tree Canopy Density (TCD) and the CORINE Land Cover (CLC) geo-layers, both available at Copernicus Land Monitoring Services website, to pre-select representative "ground" and "dense forest" areas, which are then used for sampling from the SAR image product, i.e. backscatter intensity image, to find the representative models' coefficients. Considering such rather flexible model, training procedure and that the prediction is based solely on Sentinel-1 data products, makes the method to be a highly transferable solution for biomass retrieval purposes.

The first selected technique, namely the Water Cloud Model (WCM), uses SAR backscatter intensity measurements only. The methodology developed by Santoro et al. (2011) was used as the reference for this variant of the model.

4.1.1 Model description

4.1.1.1 Pre-processing

Sentinel-1 Level-1 Ground Range Detected (GRD) data files need to be processed in order to get reliable good quality backscatter intensity measurements to be used in the WCM prediction. All data processing can be performed using the Sentinel Application

Platform (SNAP). Generally, after applying the precise orbit file to update state vectors and thermal noise removal operators to improve the global image accuracy geometrically and radiometrically, respectively, the radiometric calibration operator must be applied to convert digital pixel values to radiometrically calibrated backscatter. For this purpose, images are calibrated to β^0 values, which represent the most basic radiometric calibration without accounting for the local incidence angles. For speckle filtering, the Multilook operator which simply averages over adjacent pixels, thus reducing the standard deviation of the noise level globally, is used. The radiometric bias caused by topographic land features is accounted for by applying the Terrain Flattening operator (Small, 2011). Finally, the inherent geometric distortions caused by the combination of side-looking SAR acquisition geometry and topographical variations of the scene are accounted for by orthorectifying the image using the Range-Doppler Terrain Correction operator.

GRD data files pre-processing steps using SNAP summary:

1) Application of precise orbit files;
2) Thermal noise removal;
3) Calibration to β^0;
4) Multi-looking 10 x 10 pixels in range and azimuth directions (for 100 m pixel size);
5) Application of Terrain Flattening operator;
6) Application of Range-Doppler Terrain Correction operator.

4.1.1.2 The model formulation

Sentinel-1 operates at the C-band with a central frequency of 5.404 GHz, which corresponds to a wavelength of approximatelly 5.55 cm. Consequently, the backscatter over a forested area can be considered to be a sum of two main contributions: the upper part of the forest canopy and the ground. The high attenuation of the signal in the upper part of the canopy leads to very little to no attenuation at all from the tree-ground double-bounce interaction, or backscatter from the tree branches. Therefore, in a manner similar to the pioneer Water Cloud Model for vegetation by Attema & Ulaby (1978), the total forest backscatter can be described as the sum of direct scattering from the ground through the gaps in the canopy, ground scattering attenuated by the canopy and direct backscatter from vegetation, as given in Askne et al. (1997). Alternatively, it can be expressed as a function of the growing stock volume, V, thus relating forest backscatter to biomass directly, as given in Pulliainen et al. (1994):

Equation 4-1. $$\boldsymbol{\sigma^0_{for} = \sigma^0_{gr} e^{-\beta V} + \sigma^0_{veg}(1 - e^{-\beta V})}$$

Here, the first term in the equation represents the backscatter contribution coming from the forest floor, and the second term accounts for the backscatter portion coming from the forest canopy. The three unknown coefficients that need to be estimated are the backscatter from the unvegetated or little vegetated land surface, σ^0_{gr}, the backscatter from the dense vegetation fields, σ^0_{veg}, and the β parameter is an empirically defined coefficient expressed in ha/m^3. These parameters can be estimated in two ways: by means of the least-squares regression using a reference dataset from the field inventory, or by estimating the coefficients directly from the SAR backscatter intensity image. The latter method is preferred due to its independence from in situ measurements, therefore, disclosing better spatial transferability features.

4.1.1.3 Unknown parameters determination

Parameters σ^0_{gr} and σ^0_{veg} must be defined for every SAR image individually due to the backscatter dependence on soil moisture and local weather conditions. The estimation of the parameters σ^0_{gr} and σ^0_{veg} can profit from forest layer products. Similarly to Santoro et al. (2011) using the MODIS Vegetation Continuous Fields tree cover product with 500 m spatial resolution, we employ forest Tree Cover Density (TCD) High Resolution Layer (HRL) with 20 m spatial resolution, provided by Copernicus Land Monitoring Services. The product covers the majority of Europe including our test-site countries and is well suited for automatic selection of SAR backscatter values to be included in the "ground" and the "dense forest" classes. The estimate of σ^0_{veg} is obtained after correcting the values of the backscatter for "dense forest" pixels, σ^0_{df}, for a residual contribution from the ground. It is reminded here that for this model the TCD product is used only for the creation of the "ground" and the "dense forest" masks for the estimation of σ^0_{gr} and σ^0_{df}, and the actual tree cover density percentage values are not used for biomass estimation. Additionally, the Copernicus' Corine Land Cover (CLC) layer is used to remove all the land cover/land use classes unrelated to vegetation.

As mentioned earlier, the parameter retrieval is done directly from the SAR image, and similarly to the procedure described in Santoro et al. (2011), σ^0_{gr} estimation is based on the following scheme:

1) Masking out vegetation unrelated CLC classes on SAR backscatter image;
2) Creation of TCD < 25% "ground" mask;

3) Mask erosion using N x N pixels sliding window, where N is approx. 10 pixels, but varies from scene to scene;
4) 2-3% of all pixels are assigned to pure “ground” class;
5) σ^0_{gr} is the median value of the histogram.

Differently to σ^0_{gr}, the estimation of σ^0_{veg} requires estimating σ^0_{df} first. Theoretically, σ^0_{veg} represents the backscatter in the case of a completely opaque forest canopy. However, even the densest forest canopy has gaps in it, meaning that some portion of backscatter is coming from the forest floor. To compensate for this contribution coming through the gaps in forest canopy, Equation 4-1 can be inverted to obtain an estimate of σ^0_{veg} from the backscatter of the pixels forming “dense forest” class:

Equation 4-2.

$$\boldsymbol{\sigma^0_{veg}} = \frac{\sigma^0_{df} - \sigma^0_{gr} e^{-\beta V_{df}}}{1 - e^{-\beta V_{df}}}$$

To estimate σ^0_{veg}, Equation 4-2 requires knowledge of 4 other parameters: σ^0_{gr} that is already found, σ^0_{df} to be found next, β to be discussed after that, and a constant V_{df} describing the GSV value representative for the “dense forest” class, which can be set equal to the maximum GSV value expected at the area of interest (Maurizio Santoro et al., 2011).

Similarly, to σ^0_{gr}, the estimation of σ^0_{veg} can be summarized by the following scheme:

1) Masking out vegetation unrelated CLC classes on SAR image;
2) Creation of TCD > 70% “dense forest” mask;
3) Mask erosion using N x N pixels sliding window, where N is approx. 10 pixels, but varies from scene to scene;
4) 2-3% of all pixels are assigned to pure “dense forest” class;
5) σ^0_{df} is the median value of the histogram;
6) σ^0_{veg} is determined using Equation 4-2.

The procedure of σ^0_{gr} and σ^0_{df} parameters estimation is illustrated in Figure 8.

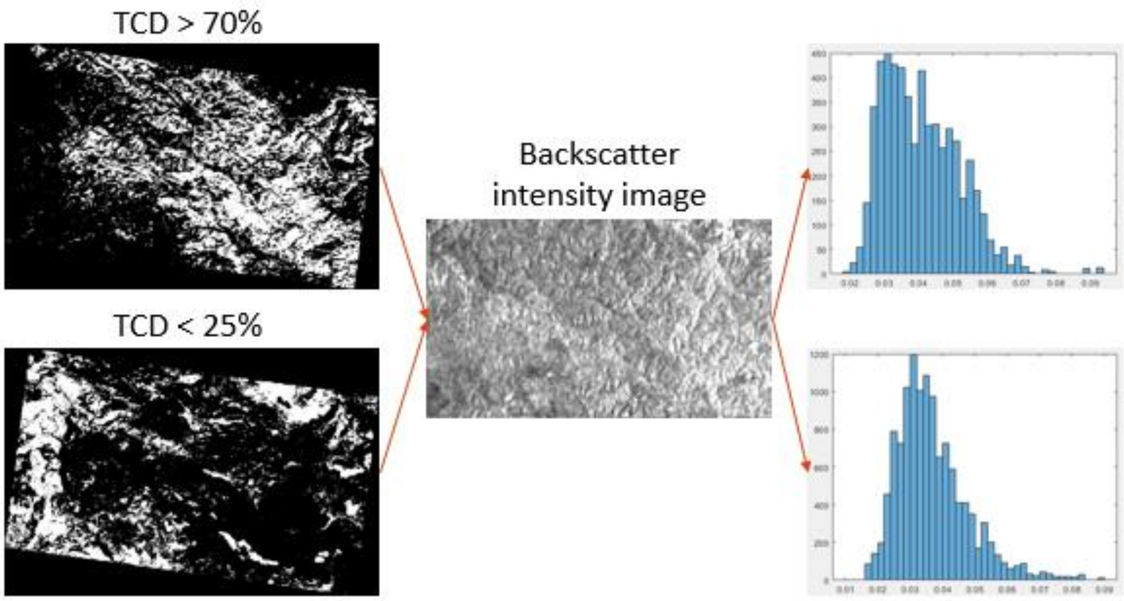

Figure 8. An illustration of σ^0_{gr} and σ^0_{df} parameter estimation. Steps from left to right: mask creation, mask erosion (skipped), pixel selection and parameter estimation from the acquired histogram.

Finally, the β parameter is related to the vegetation dielectric and the forest structural properties, including seasonal effects such as frozen/unfrozen and leaf-on or -off conditions (J.T. Pulliainen, Mikhela, Hallikainen, & Ikonen, 1996). Therefore, it requires to be adapted to the local environmental and forest conditions. Generally, β value varies between 0.004 and 0.012 m^3/ha for mature forest (CCI_BIOMASS_ATBD_V1, by Maurizio Santoro, Cartus, & Lucas (2019)).

4.1.1.4 AGB retrieval from a single image

After the estimation of σ^0_{gr} and σ^0_{veg} and the selection of β parameters, the WCM in Equation 4-1 can be inverted for GSV retrieval. However, special care should be taken when considering differently polarized backscatter intensity measurements. It was noticed that generally typical σ^0_{veg} value exceeds σ^0_{gr}, i.e. backscatter intensity from the vegetation is larger than from the ground, in the case of *VH*-polarized data, but the two parameters might switch places in the case of *VV* measurements. Consequently, the inverted equation sometimes differs for *VH* and *VV* datasets, as shown in Equations 4-3 and 4-4, respectively:

Equation 4-3. $$\hat{V} = -\frac{1}{\beta} ln\left(\frac{\sigma^0_{veg}-\sigma^0_{for}}{\sigma^0_{veg}-\sigma^0_{gr}}\right)$$

Equation 4-4. $$\hat{V} = -\frac{1}{\beta} ln\left(\frac{\sigma^0_{for}-\sigma^0_{veg}}{\sigma^0_{gr}-\sigma^0_{veg}}\right)$$

It may happen that the estimated backscatter value, σ^0_{for}, drops out of the range enclosed by σ^0_{gr} and σ^0_{veg}, thus, leading to negative or infinite biomass estimate in Equations 4-3 and 4-4. In such case, the estimate is assigned with an arbitrary GSV value: 0 m^3/ha in the case of negative estimate and the GSV_{max} value in the case of infinity, as illustrated in the example given in Figure 9. The whole model working principle is outlined in the flowchart given in Figure 10.

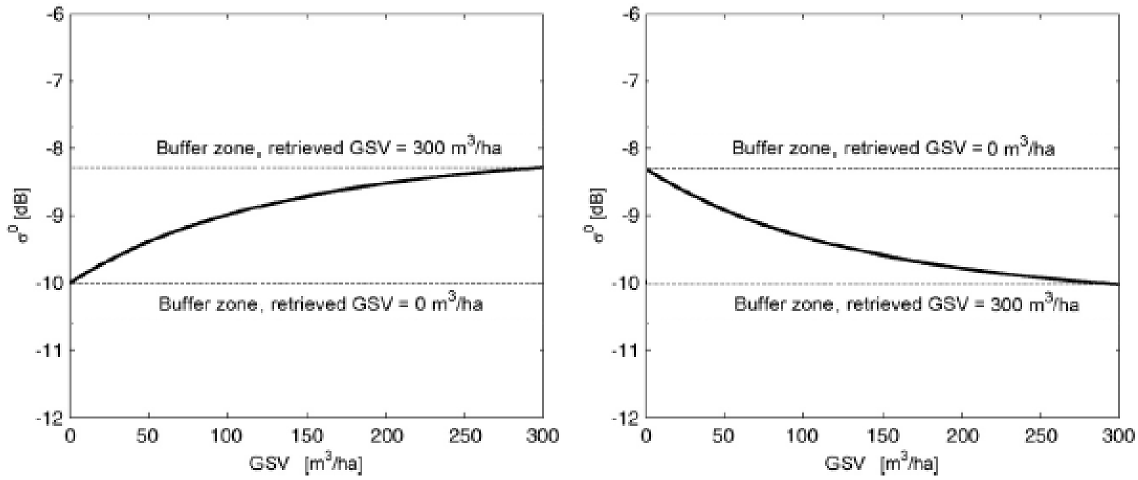

Figure 9. An example of assigning arbitrary GSV estimation values in the case of measured backscatter value falling out of σ^0_{gr} and σ^0_{veg} enclosed range. Left: *VH* case; right: *VV* case.

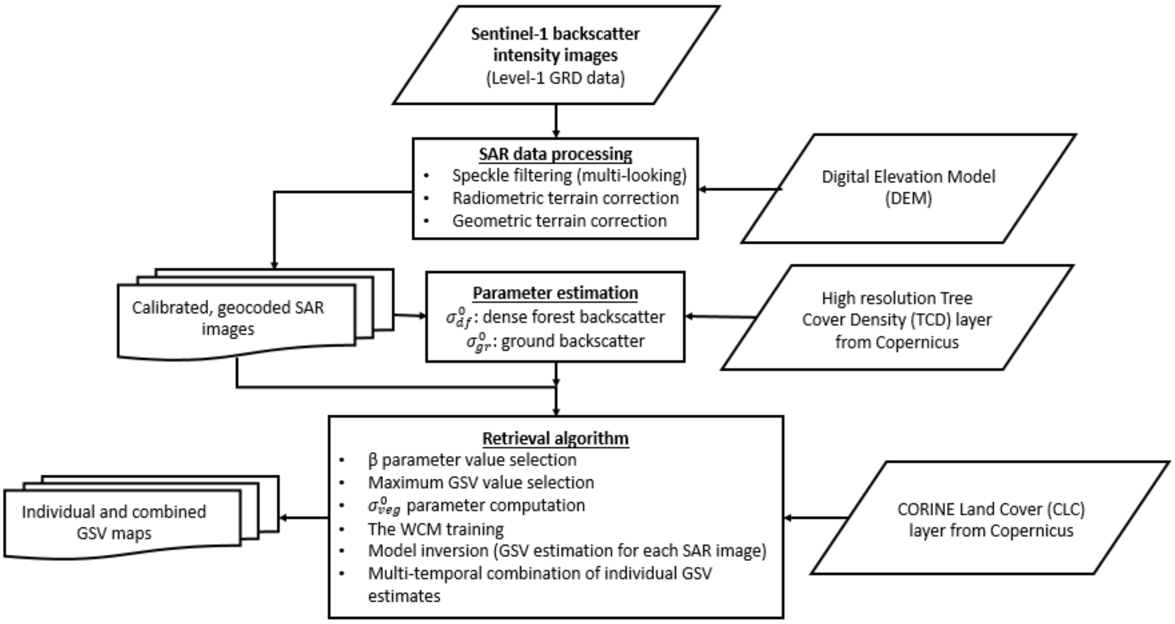

Figure 10. The WCM biomass estimation flowchart.

Finally, in order to have matching estimation result units with the reference data, it may be needed to convert the GSV estimate to AGB, in the case of ground truth provided in mass rather than volume quantity. This can be done by assuming certain wood density from a priori knowledge about the forest, e.g. a GSV value of 300 m^3/ha in boreal forest can be converted to its equivalent AGB value of roughly 150 Mg/ha (conversion factor: 0.5), while the same GSV quantity would yield around 250 Mg/ha AGB in wet tropical forest (conversion factor: 0.85) (M. Santoro et al., 2018).

Finally, it is known that such biomass retrieval procedure using Equations 4-3 and 4-4 is prone to produce significant errors in the case of estimated coefficients σ^0_{gr} and σ^0_{veg} are not being sufficiently distant to be able to accommodate the majority of the backscatter values coming from the vegetation at the reference plots. Based on the model studies, it was found that the optimal gap between the two bounding parameters should be around 4 - 5 dB, therefore, in the case of $\left|\sigma^0_{gr} - \sigma^0_{df}\right| < 4$ dB, the smaller parameter should be further reduced, while the bigger one should be further increased to reach the sufficient distance between the two. Individual estimation settings for each test site and polarization will be presented in section *4.1.2. Results*.

4.1.1.5 Multi-temporal combining of images

At Sentinel-1's C-band data the biomass estimate for individual images is affected by residual speckle noise, a temporally random element caused by environmental conditions and a more systematic kind of error component attributed to C-band backscatter insensitivity to tree stem volume. To decrease the amount of noise caused by these factors, a multi-temporal combination of biomass estimates, similar to Kurvonen et al. (1999) and Santoro et al. (2002), is used:

Equation 4-5.

$$\widehat{B_{mt}} = \frac{\sum_{i=1}^{N} \frac{w_i}{w_{max}} \widehat{B_i}}{\sum_{i=1}^{N} \frac{w_i}{w_{max}}}$$

In Equation 4-5 $\widehat{B_i}$ represents the ith estimate of AGB and $w_i = \left|\sigma_{veg}^0 - \sigma_{gr}^0\right|_i$ is the corresponding weight based on the difference between the two model parameters deputizing for the representative backscatter intensity values of the dense forest canopy and the forest floor for the image i. It is known that images containing forest backscatter measurements with stronger sensitivity to GSV are generally characterized by a larger difference between the two parameters, with the largest one in the stack of N images denoted by w_{max}. It is advised not to include images weighted below the 0.1 dB threshold for optimal performance. Such combination of images should enhance estimation accuracy, or at least provide a more reliable estimate than an average single image would.

4.1.2 Results

4.1.2.1 Espadán

Final test setting:

- Selected plots: *P. halepensis* (55), *P. pinaster* (22), *Q. suber* (7), all (80)
- Time period: March-April, July & September 2015 and January 2016 (4 periods, 10 images each)
- Polarization: *VH* & *VV*
- σ_{gr}^0 and σ_{df}^0:
    - $\sigma_{gr}^0 = (\sigma_{gr}^{00} - 2)$ and $\sigma_{df}^0 = (\sigma_{df}^{00} + 2)$ for *VH* polarization, and
    - $\sigma_{gr}^0 = (\sigma_{gr}^{00} + 2.5)$ and $\sigma_{df}^0 = (\sigma_{df}^{00} - 1)$ for *VV* polarization, where σ_{gr}^{00} and σ_{df}^{00} mark the initially estimated values
- β value: 0.008 ha/m^3
- Maximum GSV value: 200 m^3/ha for *P. halepensis*, 300 m^3/ha for *Quercus* and 420 m^3/ha for *P. pinaster* and all the plots together
- GSV-to-AGB conversion coefficient: 0.66 for all except 0.75 for *Quercus*

Table 6 and Table 7 summarize the final biomass estimation results for Espadán test site using the Water Cloud Model with dual-polarization backscatter intensity data from Sentinel-1 based on the test setting outlined above. It is clear that the overall performance is quite poor as the 50 per cent rRMSE mark was surpassed only by *P. halepensis* stands using *VH* data, with the best results of 43 per cent coming from a combination of images for the periods of March-April and September in 2015. That particular March-April result, also visualized in Graph A in Figure 12, also marks the tied-best achieved Pearson's correlation coefficient of only 0.32 for the whole study.

Overall, *VH* results show stability across all four seasons for each one of the four categories of trees, as the maximum difference is only 7 per cent in rRMSE found for the category involving all tree stands between the periods of March-April (75) and July (82). It can also be noticed that the estimation results for *VH* data always show better results (in rRMSE sense) for the periods of March-April and September over July and January. At the same time the periods March-April and September are attributed with larger mean gaps between σ_{gr}^{0} and σ_{df}^{0} parameters. The always negative bias in Table 6 indicates about gross AGB under-estimation, and is also well illustrated by the left Graph in Figure 11. The same figure also shows that biomass prediction for *P. pinaster* and *Q. suber* types of trees is much worse than that of *P. halepensis*, as it is also reflected in Graphs C, E and A in Figure 12. *P. pinaster* stands in particular demonstrate significant under-estimation as the bias reaches even up to a negative 100 Mg/ha value.

Overall, the situation is not any better in the case of *VV* polarization, as summarized in Table 7 and demonstrated in Figure 11. Although the performance is slightly worse by a couple of per cent in rRMSE sense in the category of all tree stands together as compared to *VH* data solution, the performance drops significantly to some 10 – 15 per cent difference in the case of *P. halepensis* stands, also demonstrated in Figure 12 Graphs A and B comparison. Nevertheless, the accuracy switches sides between the two datasets when comparing the results at *P. pinaster* and *Q. suber* stands, as also illustrated in Graphs C & D and E & F in Figure 12. However, the results for these two particular types of tree stands also show a lot of discrepancies across the three time periods, especially in the case of *Q. suber* where the rRMSE takes on values of 78, 68 and 60 for the periods of March-April, July and January, respectively, as indicated in Table 1Table 7. Interestingly, the gap between the two bounding parameters also seems to be increasing with the improving rRMSE result.

From the Pearson's correlation coefficient perspective, the *VV* results are very confusing and mostly provide even negative coefficient. On the other hand, bias is based mostly around *0* value which is a good sign in an otherwise chaotic environment.

To summarize everything up, the prediction results show reasonable performance only at *P. halepensis* tree stands (up to 43% rRMSE and 0.32 correlation coefficient value), reaching AGB level of approx. 100 Mg/ha, using *VH* dataset, which tends to provide quite stable albeit not very accurate results for all three types of individual tree stands. After reaching 100 Mg/ha mark the under-estimation becomes significant. On the other hand, *VV* dataset solution tends to outperform the other dataset at higher biomass levels (above 100 Mg/ha) in terms of rRMSE, however, Pearson's correlation coefficient demonstrates high level of uncertainty in the results. Overall, the results seem to be improving with larger gap between σ^0_{gr} and σ^0_{df} parameters.

	rRMSE (%)	r	bias (Mg/ha)	std. dev.	gap (dB)		rRMSE (%)	r	bias (Mg/ha)	std. dev.	gap (dB)
All stands MT comb. MAp	75	-0.14	-17	32	4.8	*All stands MT comb. Jul*	82	-0.19	-13	42	4.7
All stands MT comb. Sep	78	-0.14	-8	39	4.9	*All stands MT comb. Jan*	79	-0.14	-14	40	4.5
Halepensis MT comb. MAp	43	0.32	-11	18	4.8	*Halepensis MT comb. Jul*	46	0.26	-7	22	4.7
Halepensis MT comb. Sep	43	0.29	-5	20	4.9	*Halepensis MT comb. Jan*	46	0.29	-10	22	4.5
Pinaster MT comb MAp	76	0.11	-98	18	4.8	*Pinaster MT comb. Jul*	78	0.04	-100	19	4.7
Pinaster MT comb. Sep	73	0.25	-94	18	4.9	*Pinaster MT comb. Jan*	77	0.06	-98	18	4.5
Quercus MT comb. MAp	78	0.12	-54	28	4.8	*Quercus MT comb. Jul*	79	0.19	-59	24	4.7
Quercus MT comb. Sep	75	0.21	-53	20	4.9	*Quercus MT comb. Jan*	80	0.04	-55	25	4.5

Table 6. AGB retrieval statistics of the final WCM prediction test results using *VH*-polarized Sentinel-1 backscatter intensity data assessed by the reference data for Espadán test site on all tree stands together (rows 1 & 2) and separately for *Pinus halepensis* (rows 3 & 4), *Pinus pinaster* (rows 5 & 6) and *Quercus* (rows 7 & 8) at 100 m pixel size in March-April, July & September periods of 2015 and January of 2016. Gap

(dB) is the difference measure between final σ^0_{gr} and σ^0_{df} parameters. MT comb. means it is the weighted multi-temporal combination of all individual estimates for the period.

	rRMSE (%)	r	bias (Mg/ha)	std. dev.	gap (dB)		rRMSE (%)	r	bias (Mg/ha)	std. dev.	gap (dB)
All stands MT comb. MAp	83	0.1	25	54	4.1	*All stands MT comb. Jul*	79	0.17	20	54	4.2
						All stands MT comb. Jan	82	0.18	28	54	4.4
Halepensis MT comb. MAp	57	-0.17	0	24	4.1	*Halepensis MT comb. Jul*	56	-0.1	-3	24	4.2
						Halepensis MT comb. Jan	58	-0.13	1	25	4.4
Pinaster MT comb. MAp	65	-0.33	-1	63	4.1	*Pinaster MT comb. Jul*	63	-0.26	-2	61	4.2
						Pinaster MT comb. Jan	54	-0.06	-1	54	4.4
Quercus MT comb. MAp	78	-0.19	4	48	4.1	*Quercus MT comb. Jul*	68	0.01	0	39	4.2
						Quercus MT comb. Jan	60	0.32	3	47	4.4

Table 7. The same data as in Table 6 but for *VV* polarization data. Note, results for September are unavailable.

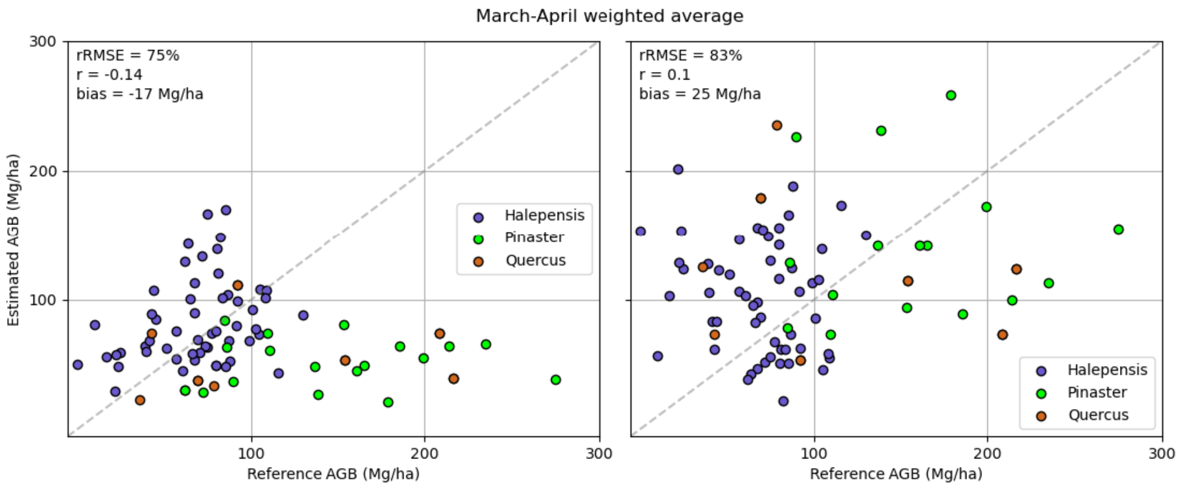

Figure 11. Multi-temporal combination of all individual estimates for all tree stands for the period of March-April, 2015 for Espadán test area. Graph on the left represents *VH* data result, while *VV* data solution is on the right.

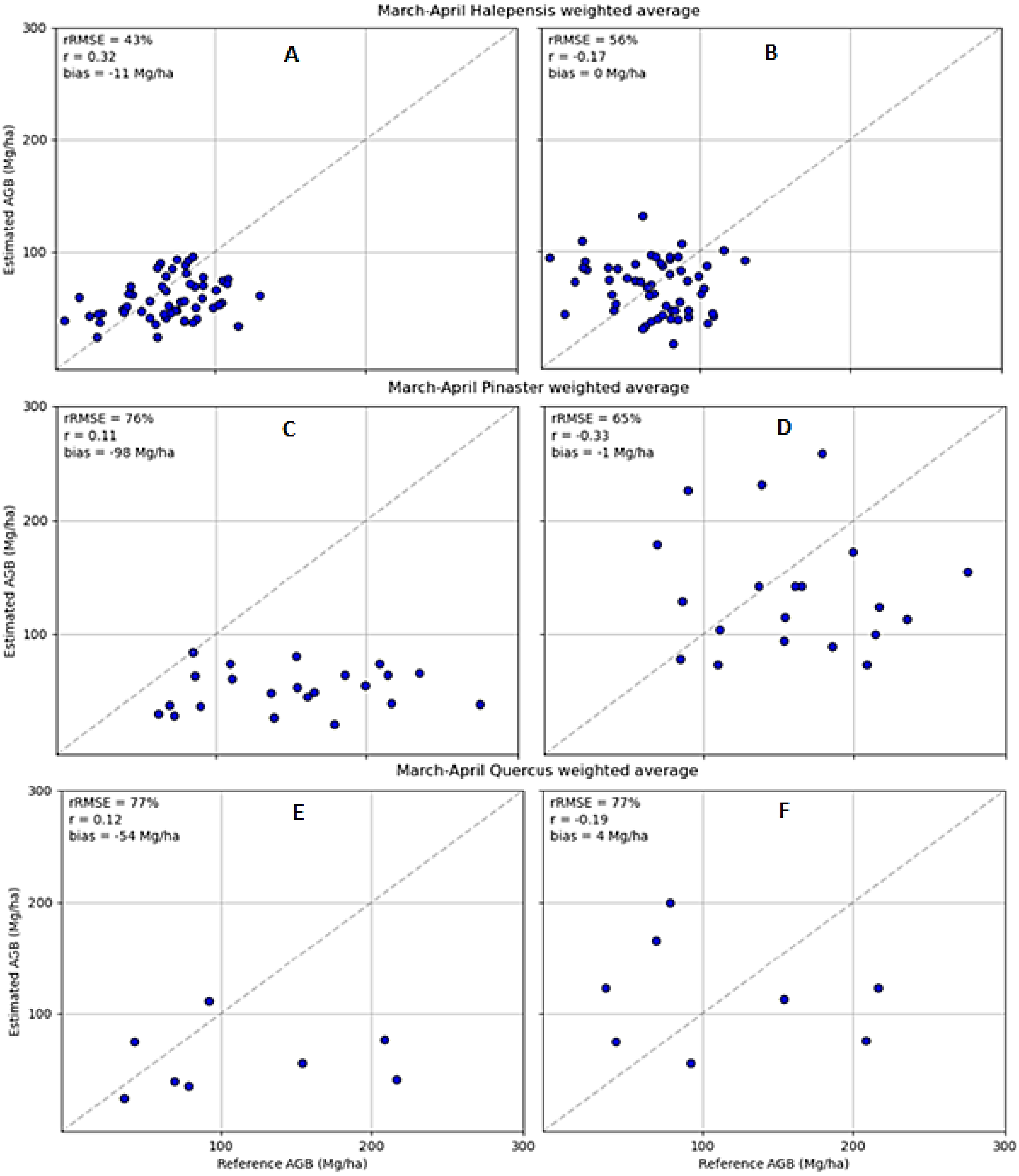

Figure 12. Multi-temporal combination of all individual estimates for the three types of tree stands (top: *P. halepensis*, middle: *P. pinaster*, bottom: *Quercus*) for the period of March-April, 2015 for Espadán test area. Graphs on the left represent *VH* data results, while *VV* data solutions are on the right.

4.1.2.2 Teruel

Final test setting:

- Selected plots: all *P. sylvestris* (51)
- Time period: September 2015 (4 images)
- Polarization: *VH* & *VV*
- σ^0_{gr} and σ^0_{df}:
    - $\sigma^0_{gr} = \sigma^{00}_{gr}$ and $\sigma^0_{df} = \sigma^{00}_{df}$ for *VH* polarization, and
    - $\sigma^0_{gr} = (\sigma^{00}_{gr} - 2)$ and $\sigma^0_{df} = (\sigma^{00}_{df} + 2)$ for *VV* polarization, where σ^{00}_{gr} and σ^{00}_{df} mark the initially estimated values
- β value: 0.008 ha/m^3
- Maximum GSV value: 280 m^3/ha for *P. sylvestris*
- GSV-to-AGB conversion coefficient: 0.66

Figure 13 demonstrates the results of multi-temporal combination of Sentinel-1 backscatter intensity measurements of 4 images over the month of September in 2015 over the study area of Teruel based on the test settings indicated above. It should be noted here that the contrast between σ^0_{gr} and σ^0_{df} parameters at *VH* polarization for the 4 images averaged at 4 dB, therefore the range expansion was not needed. On the other hand, *VV* polarization showed that for these 4 images backscatter intensity from the forest is stronger than the backscatter from the ground by 1.3 dB on average, meaning that this time *VV* backscatter behaves according to the trend seen in *VH* polarization. Therefore, the range expansion was re-adjusted to the one that was initially designed for *VH* polarization in order not to oppose the parameter estimation results.

Interestingly, *VH* data in Figure 13 does not show any signs of saturation as the estimation results keep on going up after the mark of 100 Mg/ha, seen in Espadán study case, is passed, and goes all the way up to 200 Mg/ha. It is also indicated by a respectable correlation coefficient of 0.45 and a bias value of 4 Mg/ha. The same cannot be said about *VV* data results, though, as the rRMSE increases by 21 per cent, correlation coefficient drops by two thirds, and bias grows significantly negative.

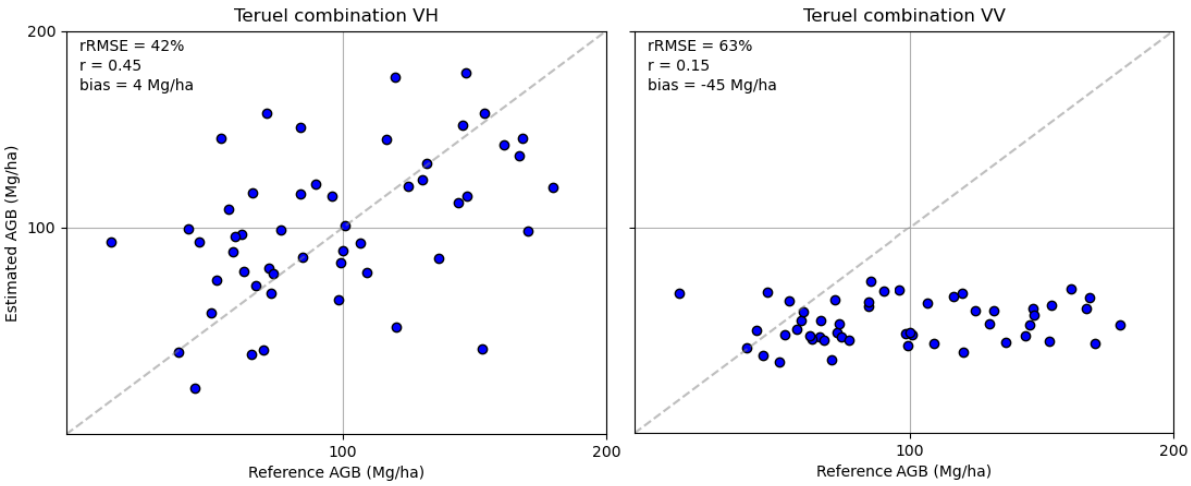

Figure 13. Multi-temporal combination of all individual estimates for Teruel study area. Graph on the left represents *VH* data result, while *VV* data solution is on the right.

4.1.2.3 Thessaloniki

Final test setting:

- Selected plots: all (41)
- Time period: September 2014 (4 images)
- Polarization: *VH* & *VV*
- σ^0_{gr} and σ^0_{df}:
    - $\sigma^0_{gr} = (\sigma^{00}_{gr} - 2)$ and $\sigma^0_{df} = (\sigma^{00}_{df} + 2)$ for *VH* polarization, and
    - $\sigma^0_{gr} = (\sigma^{00}_{gr} + 2.5)$ and $\sigma^0_{df} = (\sigma^{00}_{df} - 1)$ for *VV* polarization, where σ^{00}_{gr} and σ^{00}_{df} mark the initially estimated values
- β value: 0.03 ha/m^3
- Maximum GSV value: 120 m^3/ha
- GSV-to-AGB conversion coefficient: 0.66

Figure 14 demonstrates the results of multi-temporal combination of Sentinel-1 backscatter intensity measurements of 4 images over the month of September in 2014 over the study area of Thessaloniki based on the test settings indicated above. The average initially estimated contrast between the two bounding model parameters for the 4 images was found to be equal to 1 dB for *VH* data and 0 dB for *VV* data, thus range expansion was implemented accordingly. Overall, both graphs in Figure 14 show poor estimation results with the *VV* one being the worst of the two. Disappointment is even further aggravated by the fact that the estimation was performed in the range up to 100 Mg/ha AGB, thus signal saturation is not really expected at this level.

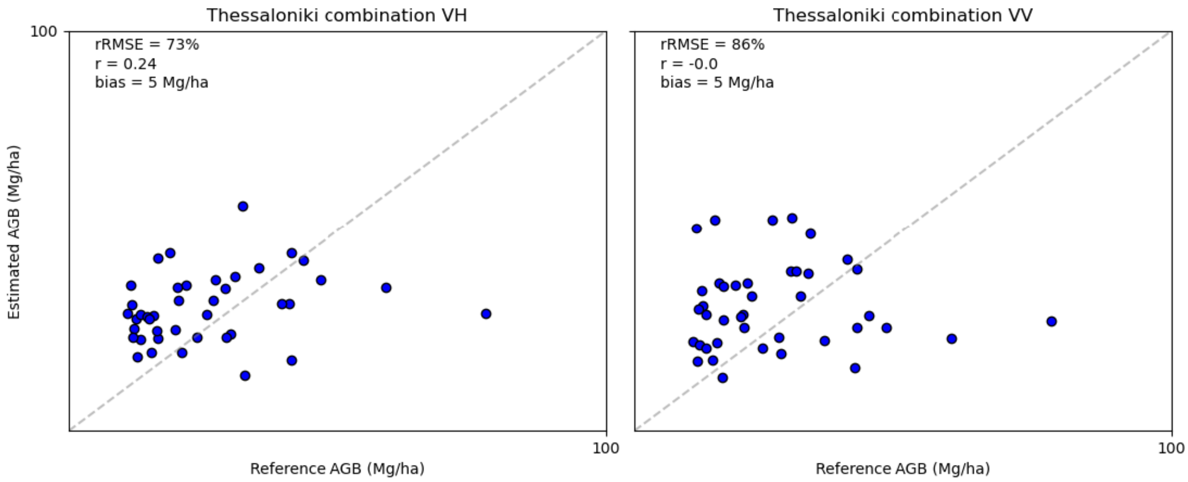

Figure 14. Multi-temporal combination of all individual estimates for Thessaloniki study area. Graph on the left represents *VH* data result, while *VV* data solution is on the right.

4.1.3 Discussion

It was found that in order to prepare the backscatter intensity data from the Sentinel-1 Level-1 GRD data files for the biomass estimation task optimally, the two following factors are crucial: speckle filtering and final pixel size. It was understood that gross multi-looking and large final pixel size allow to achieve optimal forest biomass estimation results, as also agreed upon in Santoro et al. (2011). Both factors, essentially, convey the following message: forest backscatter is a complex system involving multiple backscattering mechanisms, and short Sentinel's wavelength makes the backscatter images over the forested areas particularly speckled, therefore, significant spatial averaging is the most effective way to counter the problem for forest biomass related applications.

Another important factor that can influence biomass estimation performance using SAR backscatter images is the dependence on scene topography due to side-looking SAR geometry. The impact is particularly evident in marginal forests due to typical mountainous scenery, as is also the case for all our study sites. For example, Figure 15 shows the effect of SAR incidence angle on biomass prediction accuracy for the Teruel test site, distinguished by particularly complex terrain. The figure shows rRMSE of AGB estimate for each individual plot assessed by the corresponding incidence angle for both ascending and descending orbits. The linear trend lines indicate that even after applying the state of the art terrain correction algorithm by Small (2011), the effect of terrain topography is still apparent, albeit not excessive, but creating bias in the estimates of individual images nonetheless. However, it is quite clear that the two trend lines show almost exactly the opposite trends caused by looking over the same area from two opposite acquisition angles. Therefore, the best practice when dealing with areas affected by complex topography is to combine an equal number of images from both ascending and descending orbits to achieve the least biased estimation results. An example of such estimation results, combining two estimates from each orbit, was given in Figure 13 for the Teruel test site, yielding an impressive 42 per cent rRMSE using *VH* polarization.

In this study it was seen that a combination of estimates can provide a relatively reliable estimate as compared to individual images, i.e. the result will lie at the very least in the top half of all individual estimates, quite often in the top 10 – 20 per cent, but will rarely be the standalone best result unless large stacks of tens of images over the same area are used, e.g., in Santoro et al. (2011). Nevertheless, if no information is available to assess the accuracy of individual estimates, e.g., no ground truth data, the combined result will always be the most reliable pick.

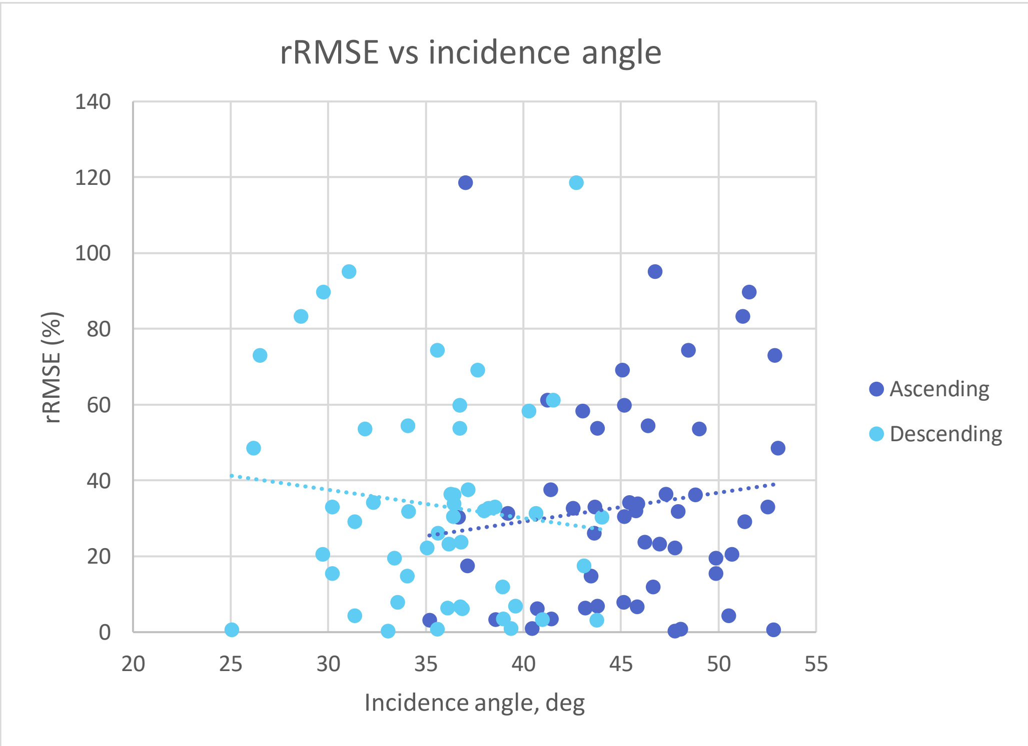

Figure 15. The effect of SAR incidence angle on biomass retrieval accuracy.

Perhaps the most important contributor to the performance of individual estimation results using the WCM is the contrast between initially estimated representative backscatter coefficients from the ground and the dense forest, i.e. $|\sigma_{gr}^0 - \sigma_{df}^0|$. It was illustrated in Figure 12 and Figure 14 for Espadán and Thessaloniki test sites, respectively, that when the two initially estimated parameters are separated only by a small margin, the estimation results are relatively poor in comparison to when the gap is large, as seen in the case of Teruel study site results for *VH*-polarized data illustrated in Figure 13.

The most important causes, besides the physical properties of backscattering objects, directly affecting the contrast between σ_{gr}^0 and σ_{df}^0 are environmental and meteorological conditions, which, in fact, lead to the changes in the physical backscatter properties. In particular, the dielectric constant is susceptible. As a result, the researchers in literature commonly agreed that dry frozen conditions allow to achieve the best biomass retrieval results. In ideal conditions the ground is also covered in dry snow cover (Fransson, Smith, Askne, & Olsson, 2001). In such case, trees can show their true physical properties under stable conditions allowing to precisely distinguish between different biomass levels present at the test site based on backscatter measurements, which are further supported by a constant backscatter from the snow cover on the forest floor, in what is an ideal situation for such backscatter intensity based WCM to flourish. Such conditions, however, cannot really exist, not for a sufficient period of time anyway, in our study sites based in Spain and Greece due to the geographic peculiarities. The strongest effect on the backscatter signal we can observe is caused by rain. On multiple occasions the observations showed relatively better results from individual images obtained during rainy days. Furthermore, in the study over Espadán test site, the investigation was completed over four different periods representing the four seasons, as was shown in Table 8. After checking the weather conditions for the four periods at a nearby weather station, the following relationship was found:

Period (10 images)	Precipitation, mm	Enhanced $|\sigma_{gr}^0 - \sigma_{veg}^0|$, dB
March-April 2015	34.6	4.8
July 2015	5.5	4.7
September 2015	53.4	4.9
January 2016	0.8	4.5

Table 8. Relationship between precipitation and the contrast between the bounding parameters of the WCM for Espadán study site for the four selected periods.

As seen from Table 8, more precipitation generally leads to larger contrast between the two parameters. Moreover, the results in Table 6 confirm that the biomass estimation accuracy for the periods of March-April and September top the results achieved in July and January for all four categories of trees using *VH* data. It can be mostly explained by the fact that trees' needles and, especially, leaves can gather some rain drops on their surface, which strengthens the volumetric backscatter signal (Klaassen, van der Linden, Ballast, & Esa, 1997), particularly notable at higher frequencies. Interestingly, the results for oak category (the only broadleaved trees in the study site) show the estimation accuracy increasing with increasing amount of precipitation for all four periods in what may or may not be a coincidence. However, such findings do not go hand in hand with the literature, as most of the researchers refer to such wet conditions as unstable (Cartus, Santoro, Schmullius & Li, 2011). Therefore, we abstain from giving definite recommendations as our results do show a certain element of uncertainty, but it does invite to look into this subject again.

Meanwhile, other meteorological conditions like temperature or humidity seem to have very little to no effect on the backscatter signal, given that it does not stand for extreme cases, e.g., dry frozen conditions. Of course, colder temperatures can see trees drop their leaves, which obviously changes the signal, but once again it was not the case for our test sites, therefore we could not investigate that effect either.

Finally, throughout the entire WCM *Results* section the two different polarizations were seen to produce very different performances. In the majority of cases, though, the *VH* polarization was seen to top its neighbour both in estimation accuracy and behavior predictability, confirming its superior status in literature too. Even though its superiority was evident in the Teruel results, given in Figure 13, where it exceeded *VV* polarization 42 vs. 63 per cent in rRMSE due to much stronger initially estimated contrast between the bounding parameters, it was not that clear in the case of Espadán where both polarizations showed similar contrasts. While *VH* results showed narrower, more confined estimation results with much larger correlation coefficient, the *VV* results tended to be largely spread and have negative correlation coefficients, although accuracy was sometimes even higher, as shown in Figure 12. The comparison, thus, allowed to make the following conclusions: 1) *VH*-polarized backscatter is more reliable because generally shows larger contrast between the ground and the forest canopy, while *VV* backscatter does not even show consistency in which parameter between σ_{gr}^0 and σ_{df}^0 carries the stronger backscatter, and; 2) in the case of both polarizations showing similarly low contrast, the *VH* polarization is more likely to show reliable results at lower biomass levels (< 100 Mg/ha), but will saturate after that, however, the *VV* backscatter seems to be less affected by saturation and is therefore likely to provide a better estimate above that (> 100 Mg/ha), as suggested in Figure 11 and Figure 12.

We suspect the results in Thessaloniki study area showed no reliability due to very low vegetation biomass levels, as the forests present there are very young and the signal shows no difference between different plots, as well as questionable quality of old ground truth data from 2007 adds to the uncertainty.

Finally, the GSV-to-AGB conversion coefficient and the β parameter selection add further errors to the estimation process. On the other hand, the possibility of adapting to the local scenes by "learning" from the SAR image about typical backscatter values coming from the "ground" and the "dense forest" surfaces boost the WCM transferability properties significantly.

4.2 Interferometric Water Cloud Model – Inversion Model

4.2.1 Introduction

Interferometric Synthetic-Aperture Radar (InSAR) has been used in various applications involving vertical profiling in the landscape such as vegetation heights or land elevation models. This is due to its phase information corresponding to vertical profile information. InSAR has the potential to become an important tool for mapping and monitoring of above-ground dry biomass (AGB), which is a forest parameter important for climate modelling (Koch, 2010). But, deriving forest biomass values from InSAR data is non-trivial and scattering from forests must be understood and the dependence of InSAR observables on biomass, environmental factors, and InSAR system configuration must be modelled (Soja, Askne & Ulander, 2017).

Many such models exist, one of which is the Interferometric Water Cloud Model (IWCM). Similar to the Water Cloud Model (WCM), the IWCM models forest as a random volume with gaps located above a ground surface. The semi-empirical Interferometric Water Cloud Model, IWCM, is based on models for backscattering coefficient, coherence and phase height (J. Askne & Santoro, 2012; J.I.H. Askne et al., 1997).

4.2.2 IWCM Theory and Method Description

4.2.2.1 Overview

In the IWCM method, the backscatter is identified similar to the Water Cloud Model (WCM) method (Attema & Ulaby, 1978), but generalized to include gaps in the vegetation cover by the introduction of the area-fill η, the area fraction covered by vegetation.

Equation 4-6.

$$\sigma^0_{\text{for}} = \eta\left[\sigma^0_{gr} e^{-\alpha h} + \sigma^0_{veg}\left(1 - e^{-\alpha h}\right)\right] + (1-\eta)\sigma^0_{gr}$$

where σ^0_{gr} is the ground backscattering coefficient, σ^0_{veg} is the vegetation layer backscattering coefficient, h is the height of the layer of random scatterers (m), and α is the attenuation coefficient (m^{-1}). Similar to the WCM method described in Section 4.1, these three parameters, σ^0_{gr}, σ^0_{veg} and α are the unknown parameters that need to be estimated.

In order to estimate the unknown variables, a new semi-empirical method using least-squares error optimization introduced by Askne et al. (2017) is used in this section. To achieve this, a set of equations (presented in the next section) that describe the biomass as a function of the backscattering coefficient and coherence data from SAR images is used. The ground and vegetation backscatter coefficients (σ^0_{gr} and σ^0_{veg}), canopy attenuation (α), and zero-biomass coherence (γ_{sys}) in the resulting set of equations are then computed using least squares optimization.

4.2.2.2 Method Description

In the IWCM Method, the complex coherence coefficient is the main interferometric observation used. It is a measure of similarity between two images. Coherence typically has several contributions, including terms for volume decorrelation and for temporal and system decorrelation (J. Askne, Santoro, Smith, & Fransson, 2003; Maurizio Santoro et al., 2002).

The original method as described by Askne et al. (2017) utilizes *TanDEM-X* data, where due to the bistatic observations with zero baseline, the temporal correlation can be neglected, and the coherence equation reduces to

Equation 4-7

$$\bar{\gamma} = \gamma_{\text{sys}} \frac{\bar{\gamma}_{vol} + m}{1 + m}$$

where γ_{sys} is the zero height coherence, γ_{vol} is the volume decorrelation (J.I.H. Askne et al., 1997) determined by α and h, and m is the ground-to-volume scattering ratio:

Equation 4-8 $$\bar{\gamma}_{vol} = \frac{\alpha}{\alpha - jk_z} \frac{e^{-jkzh} - e^{-\alpha h}}{1 - e^{-\alpha h}}$$

Equation 4-9 $$m = \frac{\sigma_{gr}^0}{\sigma_{veg}^0} \frac{1 - \eta\left(1 - e^{-\alpha h}\right)}{\eta\left(1 - e^{-\alpha h}\right)}$$

where k_z = 2π/HoA and HoA is the height of ambiguity. The phase height, i.e. the height of the phase center, z_{est}, are determined by

Equation 4-10 $$z_{est} = -\frac{HoA}{2\pi} \arg\left(\bar{\gamma}\right)$$

In addition to the parameters that can be obtained from SAR images, there are a few that need ground data, making this method only semi-empirical as mentioned before. The area fill factor used in Equation 4-9 is closely related to the forest structure, and hence obtained from ground measurements. A relation of it is expressed by

Equation 4-11 $$\eta(V) = \eta_\infty\left(1 - e^{-\lambda_0 V}\right)$$

where V is the stem volume and η_∞, and λ_0 are factors representing the maximum value of the area-fill and the increase with V. In this method, we will here use η_∞ = 0.9, and λ_0 = 0.01 ha/m^3, similar to the value used in the literature for coniferous forest.

Once the stem volume for each forest stand is calculated, the biomass can then be directly estimated using the formula

Equation 4-12 $$B = BF * V_i$$

To summarize, the IWCM is a semi-empirical model using radiative transfer theory to model penetration through the partially transparent canopy and geometrical optics to model penetration through canopy gaps. The unknown model parameters α, σ_{gr}^0 and σ_{veg}^0 , and γ_{sys} are assumed to be spatially invariant constants with the first three describing the properties of a certain forest type over a large area. Of the three modeled quantities (γ, z_{est}, and σ_{for}^0), z_{est} is a function of α and $\sigma_{gr}^0/\sigma_{veg}^0$. γ is also dependent on γ_{sys}, whereas σ_{for}^0 is dependent on α, σ_{gr}^0 and σ_{veg}^0. The vertical and horizontal structures of the forest are in the model described by h and η.

To solve for these interrelated set of equations, the inverse modeling approach is used as described in the following section.

4.2.3 IWCM parameter Estimation using Inverse Modeling

From the set of equations in the previous section, the IWCM parameters α, σ_{gr}^{0} and σ_{veg}^{0} , and γ_{sys} are estimated using inverse modeling. The observations of phase height, H_i, coherence, Ci, and backscatter, $S_i = \sqrt{S_{1,i} * S_{2,i}}$ where $S_{1,i}$ and $S_{2,i}$ are backscatter coefficient observations from each of the two satellites and the index i refers to each of the stands. The backscattering coefficients and coherence values are computed for each forest stand. Parameters , σ_{gr}^{0} and γ_{sys} correspond to backscatter and coherence for stem volume equal to zero, whereas parameter σ_{veg}^{0} corresponds to backscatter intensity for complete canopy cover, η = 1, and αh ≫ 1. The phase height typically increases with stem volume V, so the first estimates of σ_{gr}^{0} and σ_{veg}^{0} , and γ_{sys} can be obtained by analyzing backscatter and coherence for low and high phase height values.

Besides the IWCM parameters, α, σ_{gr}^{0} and σ_{veg}^{0}, and γ_{sys}, assumed spatially invariant, the stem volumes, V_i, for all stands are also unknown. Since the phase height varies over a large range in relation to the spread of the observations and a large number of looks is used during phase height estimation, we can neglect the uncertainty in phase height estimation and equate the IWCM phase height model z_{est} to the observed phase height, H_i. The coherence and backscatter models are used to obtain estimates of model parameters α, σ_{gr}^{0} and σ_{veg}^{0} , and γ_{sys}, which thereafter will be used to obtain the stem volume and biomass estimates for each stand.

To achieve this, we first solve for the stem volumes V_i as a function of the observed phase height and (so far unknown) values for α, σ_{gr}^{0} and σ_{veg}^{0}, and γ_{sys}. We obtain a function for stem volume estimation such that $V_i = V(\alpha, \sigma_{gr}^{0}/\sigma_{veg}^{0}, Hi)$. With this expression for V_i we use coherence and backscatter to estimate the model parameters α, σ_{gr}^{0} and σ_{veg}^{0} , and γ_{sys} by minimizing the summed squares of the differences between the modeled and measured values of coherence and backscatter for all N stands.

Equation 4-13

$$\boldsymbol{\Delta\gamma\left(\alpha, \sigma_{g}^{0}, \sigma_{veg}^{0}, \gamma_{sys}\right) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\gamma\left(\alpha, \frac{\sigma_{gr}^{0}}{\sigma_{veg}^{0}}, \gamma_{sys}, V\left(\alpha, \frac{\sigma_{gr}^{0}}{\sigma_{veg}^{0}}, H_i\right)\right) - C_i\right]^2}}$$

Equation 4-14

$$\Delta\sigma\left(\alpha, \sigma_{gr}^{0}, \sigma_{veg}^{0}, \gamma_{sys}\right) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\sigma_{for}\left(\alpha, \sigma_{gr}^{0}, \sigma_{veg,}^{0}, V\left(\alpha, \frac{\sigma_{gr}^{0}}{\sigma_{veg}^{0}}, H_{i}\right)\right) - S_{i}\right]^{2}}$$

The IWCM parameters are determined by fitting the model to the observations, Hi, Ci, and Si. When the IWCM parameters are determined the stem volume, Vi, and hence biomass, B = BF·Vi for each stand can be determined.

4.2.4 Study Site and Data

The study site is chosen as the forest sites in Espadán, Castellon and Nogueruelas, Teruel in Spain. The sites are described in detail in *Section 3*.

For SAR data, Sentinel-1 SLC data was used instead of *TanDEM-X* data CoSSC data due to its unavailability. This invalidates the assumption of negligible temporal decorrelation on which the previous equations are based on. However, as an attempt to produce a working chain for this method, Sentinel-1 SAR data was used. Two SAR images from Sentinel-1 from September 2015 covering both the study sites were used to produce the interferogram. Only *VV* polarization is considered in this study to maintain similarity to the original method.

Additionally, DSM and DTM data of the two sites from the LiDAR dataset described in *Section 2.4* of this document were used in the InSAR process.

4.2.5 Application of IWCM Method

4.2.5.1 Data Processing:

Using the Sentinel-1 SAR image pairs, an interferogram was generated using the following steps using TOPSAR tools in SNAP software.

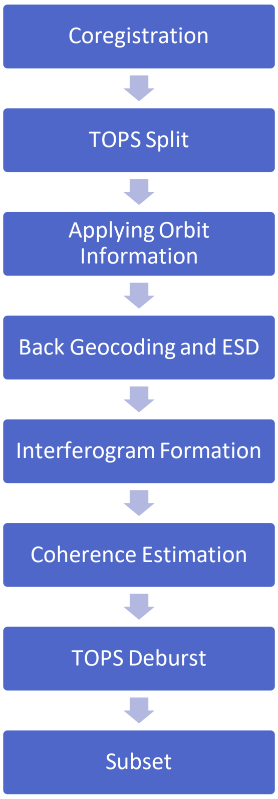

Figure 16. InSAR Processing Steps

However, as expected, the generated interferogram did not produce good results for the InSAR phase height estimation. Only the backscatter and coherence data from the SAR data was used.

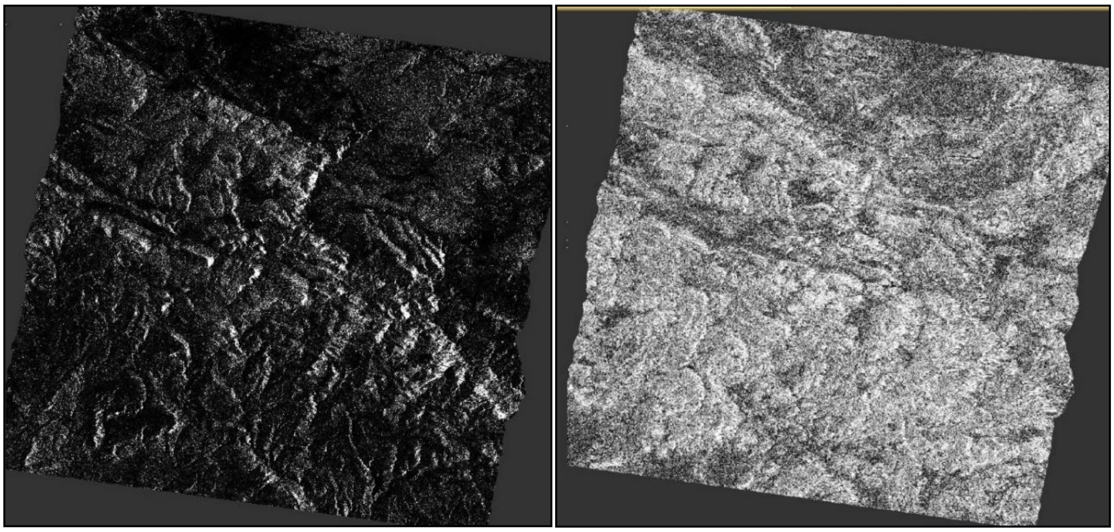

Figure 17.SAR images of Teruel Region. Left: Backscatter, Right: Coherence

Instead here, the LiDAR data is used in the calculation of the vegetation height. Using the QGIS software, a raster calculation operation is performed subtracting the DTM from the DSM, which yields primarily tree heights. Heights obtained at the plot locations from this new layer simulate the phase heights from the original method proposed by Askne et al. (2017).

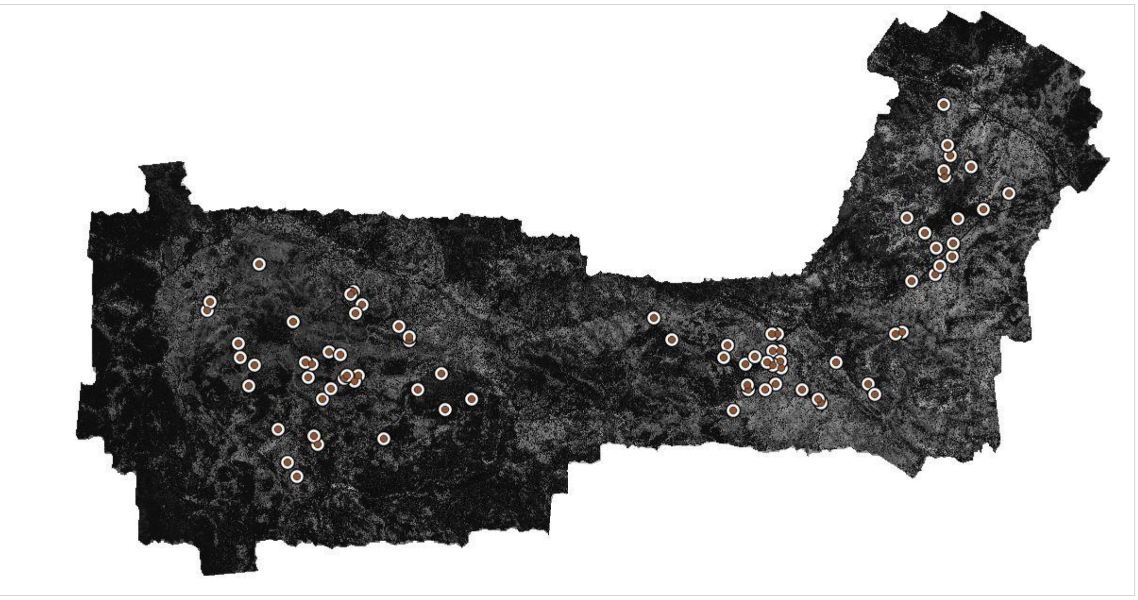

Figure 18. DSM - DTM Raster of Espadán Region, with field plot locations shown.

The coherence, backscatter and “phase height” data is then used in the following steps.

4.2.5.2 Method Implementation:

Using MATLAB, Eq. 4-7 to 4-11 are written as a set of equations of interdependent variables. Since the literature and the chosen site for this method are both coniferous forests, acquired from *TanDEM-X* satellites, it is assumed that the identified values from the site in the literature can be approximated in the case as well. The values of the constants or initial values of variables are used from the literature as follows:

Variables	Values
γ_{sys}	0.75
σ_{gr}^{0}	0.14
σ_{veg}^{0}	0.25
HoA	50
η_{∞}	0.9
λ_{0}	0.01

Table 9. Values of Variables assumed from Askne et al. (2017).

With the above mentioned initial values For α, σ_{gr}^{0}, σ_{veg}^{0}, γ_{sys}, appropriate step sizes were set for these variables within the range of their possible values. The following steps were applied:

1. The modeled phase height z_{est} is set equal to the observed phase height H_i
2. The IWCM parameters are determined by fitting the model to the observations, H_i (phase height), C_i (coherence), and S_i (backscatter) from i^{th} forest plot.
3. The best fit is estimated by finding the best values of α, h and stem volume V for which the least square error of values is minimized, as given in Eq. 4-13 and Eq. 4-14.
4. Based on the best fit parameter values, the stem volume values are identified using Eq. 4-11. Biomass is calculated for each stand using the following formula $B = BF * V_i$ as given in Eq. 4-12 with BF assumed to be 0.512 Mg/ha here.

4.2.6 Results and Discussion

Based on the method described in the previous section, the “best fit” IWCM parameters were estimated to be:

$$\alpha = 0.125,\ \gamma_{sys} = 0.769,\ \sigma_{gr}^{0} = 0.215, \text{ and } \sigma_{veg}^{0} = 0.278$$

Based on these values, the stem volume and subsequently the Biomass values (in AGB) was calculated for each plot using the Eq. 4-12. For both the sites, the predicted biomass is then compared with the reference biomass estimated from field measurements for analyses of the results. However, there are limitations to the analyses.

Firstly, it is expected that there is a large error in the predictions due to the violation of the assumption that there is no temporal decorrelation in Eq. 4-7. The temporal decorrelation is present in the Coherence and backscatter value obtain from the Sentinel-1 SAR acquisition pair, which span over several days, as opposed to the near real-time *TanDEM-X* acquisitions. This major violation of assumption in the implementation of this method due to unavailability of *TanDEM-X* data as mentioned before, severely limits the accuracies achievable in the original literature. Additionally, this also limits the number of type of statistical and geophysical analyses that can be done to explain the results that are obtained using the method, as most observations may not be in any case adequately explained due to the application of a method not suited for Sentinel-1 data used.

The temporal decorrelation error is not insignificant and can mask most other phenomenon with lower amplitudes. Therefore, no spatial analyses have been performed due to the fact such error analyses might be rendered meaningless when compared to significant error from temporal decorrelation. For instance, since the study sites are very hilly, it is expected that layover and shadow effects affect SAR coherence and backscatter even after terrain corrections steps but cannot be satisfactorily explained due to the lower magnitude of its impact compared to temporal decorrelation noise. Similarly, assumed values of field parameters from the study site in the literature, as well as environmental factors such as precipitation and temperature and its impact on the SAR SLC observation are not considered, which are known as considerable factors to be taken into account for the IWCM method in general. Regardless of these limitations in possible analyses, an attempt is made at explaining the general trends and major observations from the resulting statistics.

Site	Mean AGB Reference (Mg/Ha)	AGB Mean Estimated (Mg/Ha)	rRMSE	Mean Bias (Ref - Est) (Mg/Ha)	r
Espadán (all stands)	91.0	76.5 Mg/Ha	42.8%, underestimated	17.1	0.76
P. halepensis (54 stands)	69.1	62.3	37.2%	6.75	0.56
Q. suber (8 stands)	112.3	93.13	44.1%	19.2	0.66
P. pinaster (18 stands)	147.6	109.7	45.2%	37.9	0.73
Nogueruełas (*P. sylvestris*) (47 stands)	95.8	70.5 Mg/Ha	48.3%, underestimated	25.4	0.72

Table 10 AGB Results using IWCM Method

For this, several statistical parameters are used, primarily the rRMSE, along with the mean bias, and Pearson correlation coefficient “r”. The “r” value signifies better correlation between prediction and measurements if closer to value 1, uncorrelated if 0, and negatively correlated if close to -1. The mean bias is the difference between the reference and the predicted biomass, with positive values signifying underestimation and negative values signifying overestimation. The prediction is said to be acceptable if rRMSE is under 30%. The metrics obtained for each site using the IWCM method is summarized in Table 10.

Supporting this expectation, in general, most estimations are not very accurate. The *P. halepensis* stands of Espadán performs best with the highest accuracy and has the lowest mean bias and best rRMSE amongst all stands, at around 36%. The other two stands *P. pinaster* and *Q. suber* in Espadán, and the *P. sylvestris* stand in Nogueruelas site all suffer from large errors, primarily with lower estimated values compared to the reference values, as seen in Fig.19.

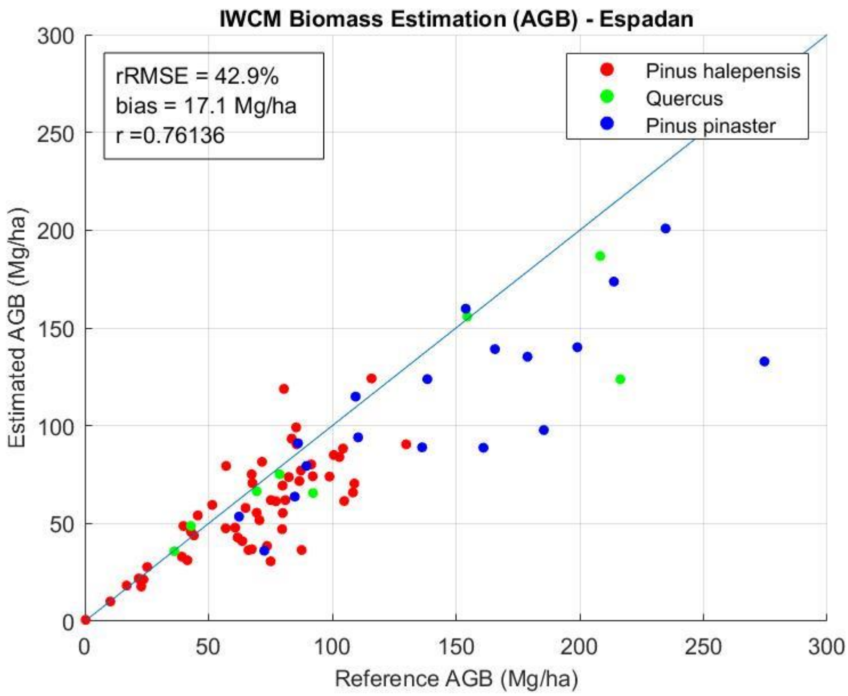

Figure 19 IWCM Biomass Estimation - Espadán

As seen in the Table 10, in all cases, the prediction value is always underestimated compared to the reference values. This is explained by the saturation levels of Sentinel-1 C-band SAR backscatter signals for biomass at values higher AGB values, typically over 100 Mg/ha. Only in the case of the *P. halepensis* site is the mean bias found to be low, which supports the explanation about the saturation, since the mean biomass value of the P. halepensis stands is much lower than the saturation limits of C-band data.

In terms of rRMSE, it can be seen that the rRMSE of both the sites are greater than 40%, which indicates that the estimated biomass poorly relates to the measurements from the field data. The error can largely be attributed to the temporal decorrelation effect, present in the coherence and backscatter values used in the inverse modelling steps. The error is especially high for the *P. pinaster* stands in Espadán and *P. sylvestris* stands in Nogueruelas. This higher error is explained in general by the saturation of C-band at higher biomass values.

In the case of Espadán, P. pinaster stands can be seen to be suffering from the effect of saturation the most, since the stand has high biomass values of well over 100 Mg/ha. A corresponding lowering of the estimated values can be seen in Fig. 19 as the biomass increases. This can be observed in *Q. suber* sites as well, however to a much smaller degree, which is likely due to the low statistical number of the *Q. suber* stands, leading to not representing the full picture. In the end, this results in *Q. suber* stands having a better overall accuracy in rRMSE term and in terms of mean bias. From Fig. 20, it can be observed that this is true in the case of *P. sylvestris* in Nogueruelas site as well.

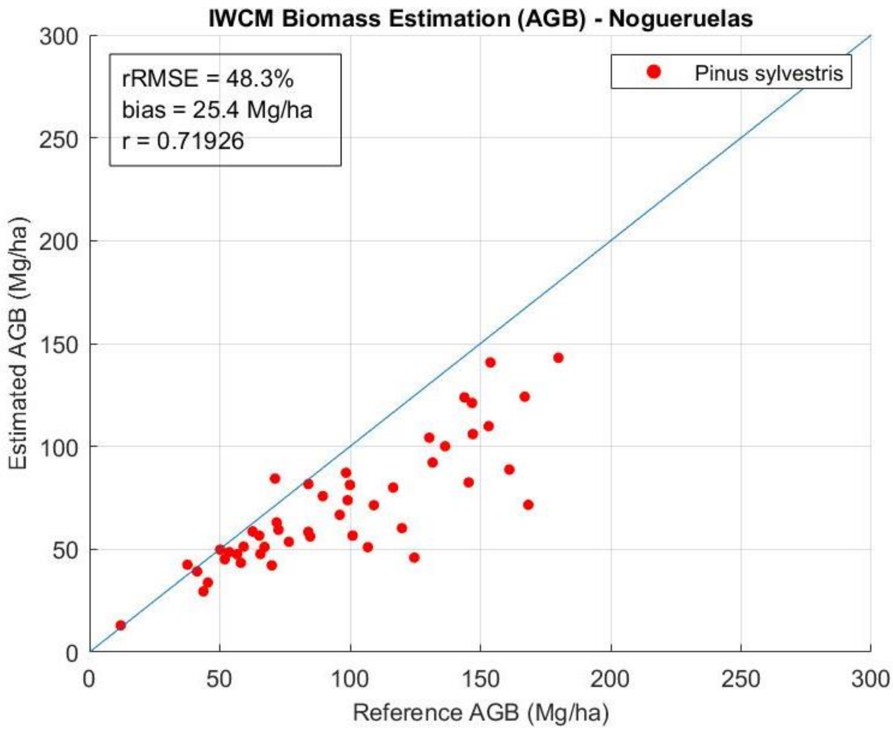

Figure 20 IWCM Biomass Estimation - Nogueruelas

In terms of 'r' value, the two sites perform similarly, with Espadán performing slightly better. In Espadán, although the most frequent *P. halepensis* stands have best rRMSE, the 'r' value is quite low, the lowest amongst all stands. The remaining stands have similar 'r' values, close to the average of the two sites.

Finally, although it is expected that the results of the IWCM method be largely erroneous due to the effect of temporal decorrelations in the measurements, which voids the initial assumptions, the use of Lidar data for the phase height measurements has resulted in lower errors. In summary, the results show that the IWCM method, despite being developed for *TanDEM-X* data, is able to provide biomass estimation for use with Sentinel-1 data along with Lidar data, although with significant errors. The impact of the temporal decorrelation could be mitigated if *TanDEM-X* data is available for a selected study site.

4.3 LiDAR

In this chapter we present the methodology followed for the estimation of the AGB at plot level using airborne LiDAR data from the acquisition and pre-processing of the ALS data to the estimation of the AGB at plot level.

In order to carry out the following methodology, it is essential to have the following data sets:

- Discrete LiDAR data, preferably with high points density per square meter (>4 pts/m^2). To obtain Elevation and intensity distribution ALS metrics.

- Field data from a set of sampling plots, usually circular plots of 15 to 20 meters radius where the tree attributes necessary to estimate the biomass by allometric equations for each species have been collected.

Once the LiDAR point cloud was pre-processed, the AGB values of each plot were calculated and the height and intensity metrics of the LiDAR point clouds for each plot were obtained, we generated the multiple linear regression models. Elevation and intensity distribution ALS cloud metrics were used as independent variables in AGB predictive models. Before the generation of the regression models a selection of the metrics was made using the Akaike information criterion (AIC) (Akaike, 1973) with a maximum of three metrics per combination. Multiple linear regression models were generated for each study area, and for each type of forest, by combining and discarding particular plots, where the dependent variables was biomass. The models were evaluated by comparing the adjusted coefficient of determination (R_{adj}^2) and the root mean square error (RMSE), all obtained by leave-one-out cross-validation (LOOCV). Models have only been generated for biomass because, as explained in section 3.4.1, CO_2 is obtained from biomass values.

Figure 21 shows the workflow for the treatment and classification of the LiDAR points to obtain the Digital Terrain Model (DTM), the Digital Surface Model (DSM) and Canopy

High Model (CHM) for the entire study area and the steps to generate the AGB estimation models.

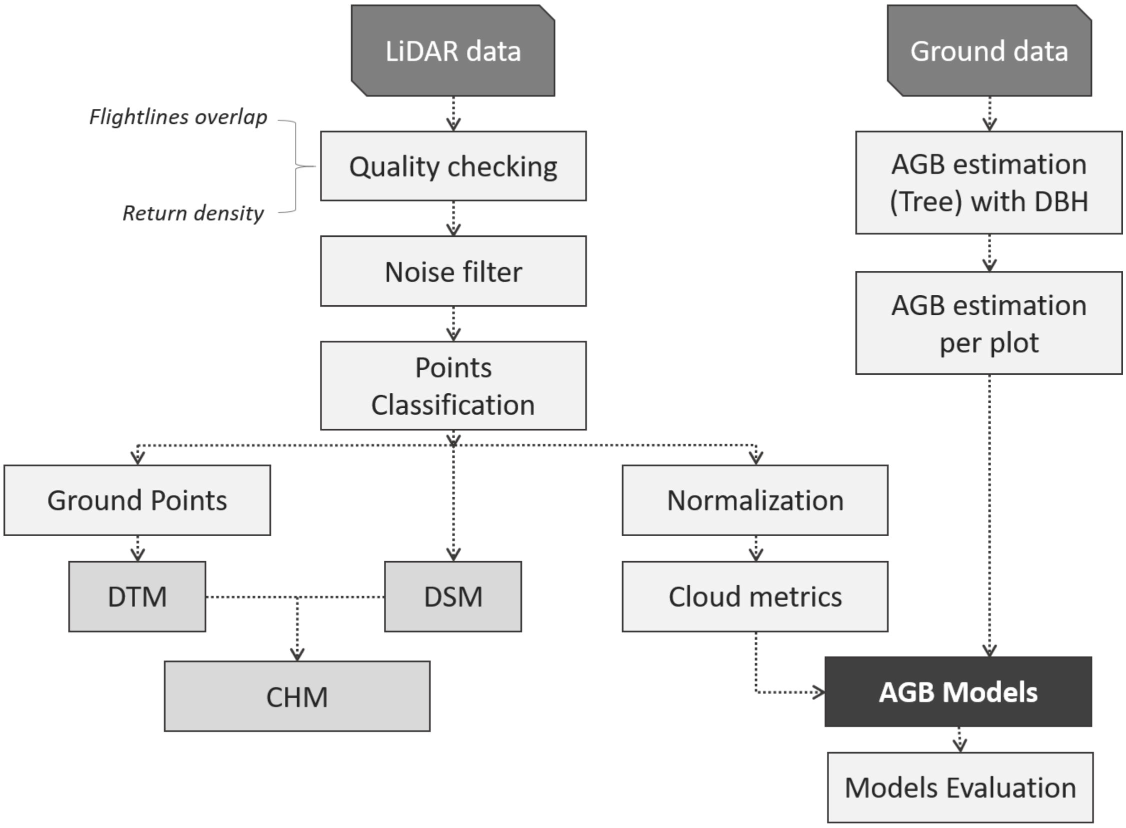

Figure 21. Overview of the methodological approach.

The following sub-sections describe each of the steps in detail.

4.3.1 ALS Acquisition

In both areas the Lidar flights were planned under the same parameters. Both, in Espadán and Nogueruelas the flight was carried out on September 16th, 2015. The sensor used to acquire the data from both zones was a LiteMapper 6800 with an average pulse density of 14 pulses·m^{-2}. In Espadán, the flight altitude ranged from 600 to 820 m above sea level with a minimum overlap of 55% and a maximum of 77% between flight lines. In the area of Nogueruelas the flight altitude was from 1,500 to 1,700 meters above sea level, with a minimum overlap between flight lines of 55% and maximum of 80%. The figures 22a and 22c show the overlap between the flight lines for Espadán and Nogueruelas, respectively. The figures 22b (Espadán) and 22d (Nogueruelas) represent the return per square meter. The return density images clearly show how the density increases in all areas of the flight line overlap.

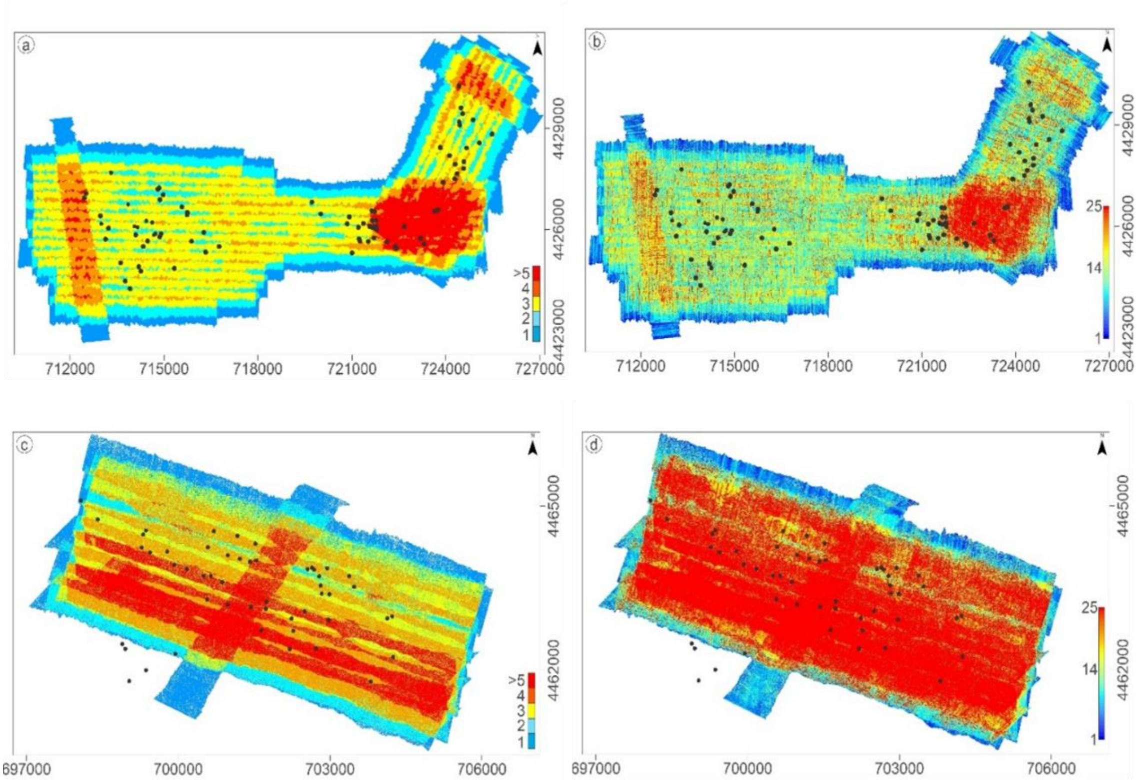

Figure 22. Airborne laser scanning acquisition. Figures 22a and 22b refer to Espadán and figures 22c and 22d represent the area of Nogueruelas. The black points are representing the plots. The figures 22a and 22c show the flightline overlap with blue indicating one flightline, turquoise indicating two, yellow indicating three flightlines, orange indicating four flightlines and red five or more flight lines. The figure 22b and 22d show the return density, dark blue colour means 1 return and red colour means 25 or more return per square meter.

4.3.2 LiDAR data pre-processing

First the filtering of irregular points was carried out. This was done using an adaptive filter that removed all outliers above 2 m from a local neighborhood throughout the study area.

The next step was to determine all the points corresponding to the ground. For this purpose, the points were classified with the *lasground* algorithm of LAStools (Isenburg, 2018) which is a variation of the algorithm described by Axelsson (2000). This algorithm depends on two parameters, a search window and a maximum angle criterion defined by the user. The points classified as terrain generate an area by interpolating these points called Digital Terrain Model (DTM). From the original cloud and establishing a defined cell size according to the density of LiDAR points, the values of the points with maximum elevation in each cell were selected. Once these points were interpolated, a Digital Surface Model (DSM) was generated. The Canopy High Model (CHM) was generated by subtracting the DTM from the DSM.

The DTM, DSM and CHM were all generated with a spatial resolution of 0.5 meters. In Espadán, the average number of points to generate the DTM, DSM and CHM was of 8.2 all returns per square meter and 5.1 last returns per square meter. For the Nogueruelas area the average number of points was 8.1 all returns per square meter and 5.4 last returns per square meter.

To calculate the error obtained in the DTM, the centers of the plots were used, since they were measured in the field by a GPS model Leica GNSS 1200 with an RTK accuracy of ± (10 mm + 1 ppm) and ± (20 mm + 1 ppm) in horizontal and vertical, respectively. The RSME of the Espadán DTM in the Z coordinate is 0.316 m and for the Nogueruelas DTM RMSE is 0.274m.

4.3.3 Metrics extraction

The calculation of the descriptive characteristics or attributes is carried out at plot level, so that it can be compared with the data collected in the field in order to generate the models later (Figure 19).

Using FUSION (McGaughey, 2016) height and intensity statistics from the normalized height point cloud were calculated for each plot. In the extraction of metrics, only the points above 2 m were considered, leaving out the understory vegetation. Table 11 shows the different metrics analysed.

Name	Class	Reference
Total number of returns	*Height or Intensity	McGaughey, Robert
Count of returns by return number (maximum 9 discrete return)		
Minimum value of *		
Maximum value of *		
Mean value of *		
Median value of *(as 50th percentile)		
Mode value of *		
Standard deviation value of *		
Interquartile distance value of *		
Skewness value of *		
Kurtosis value of *		
AAD: Average Absolute Deviation value of *		
MADMedian: Median of the absolute deviations from the overall median value of *		

Name	Class	Reference
MADMode: Median of the absolute deviations from the overall mode value of *		
L-moments (L1, L2, L3, L4) value of *		
L-moments skewness value of *		
L-moments Kurtosis value of *		
Percentil values of *		
Canopy relief ratio ((mean-min)/(max-min))		
Generalized means for the 2nd and 3rd power: Elev. quadratic mean and Elev. cubic mean		
Percentage of first returns above a specified height (canopy cover estimate)	Height	
Percentage of first returns above the mean height/elevation		
Percentage of first returns above the mode height/elevation		
Percentage of all returns above a specified height		
Percentage of all returns above the mean height/elevation		
Percentage of all returns above the mode height/elevation		
Number of returns above a specified height/total first returns * 100		
Number of returns above the mean height/total first returns * 100		
Number of returns above the mode height/total first returns * 100		

Table 11. Description of ALSD metrics (see McGaughey, 2016, for further description).

4.3.4 Modelling

To generate the models, the reference biomass per plot (ground truth) was available. The estimation models were therefore calculated using multiple regression techniques, considering biomass as a dependent variable and the height and intensity metrics derived from the LiDAR data described in section 4.3.3 as independent variables.

Initially, a criterion for the selection of the independent variables was applied by means of step-by-step multiple regression. The intention was to generate models that had a maximum of 3 variables. This method consists of introducing variables into the model according to their relevance in the regression or eliminating them according to their correlation with other variables. Thus, a multiple linear regression of the type (limited to three independent variables) is obtained:

Equation 4-15

$$\boldsymbol{AGB = a_0 + a_1x_1 + a_2x_2 + a_3x_3}$$

Where AGB is the aerial biomass (dependent variable), x_1 the metrics used in the model (independent variables) and $a_1$the coefficients obtained by the ordinary least-squares regression method.

The step-by-step method requires some mathematical criteria to determine whether the model becomes better or worse with each addition or extraction. In this case the criterion of Akaike (AIC) (Akaike, 1973) is used because it is more restrictive and offers a relative estimation of the information lost by a given model: the less information a model loses, the higher the quality of that model.

It was decided to make a single model for Nogueruelas since there is only one dominant tree species. On the contrary, for Espadán three different models were generated because there are different types of plots depending on the dominate species (classification made in sub-section 3.4.1). It was decided to make a general model for the whole area, a model only for the species *P. halepensis* discarding the regenerated *P. halepensis* plots because of their difference in structure and density, and a model for the mixed *P. pinaster* and *Q. suber*. Type T2 (*P. halepensis* regenerated) and T4 (*Q. ilex*) were discarded for not having enough plots to perform the model.

The evaluation of the models is made relating the AGB observed in field and the predicted ones from the LiDAR data, analyzing the determination coefficient R^2, defined as a descriptive measure of the global adjustment of the model that represents the proportion of variance explained by the same one. The R^2 coefficient is calculated as the square of the correlation coefficient of the moment product of Pearson:

Equation 4-16

$$\boldsymbol{R^2 = \frac{\sum_{i=1}^{N}(\alpha_i - \overline{\alpha_i})\cdot(\alpha_{iobs} - \overline{\alpha_{iobs}})}{\sqrt{\sum_{i=1}^{N}(\alpha_i - \overline{\alpha_i})^2 \cdot \sum_{i=1}^{N}(\alpha_{iobs} - \overline{\alpha_{iobs}})^2}}}$$

where *i* is the predicted value from the LiDAR data for plot *i*, *iobs* is the observed value in the field for the same plot, and *N* is the number of plots analyzed.

In addition, the $R^2_{adjusted}$ was also calculated that introduces a penalization to the R^2 value for each predictor that is introduced in the model. The value of the penalization depends on the number of predictors used and the size of the sample, that is, the number of degrees of freedom. $R^2_{adjusted}$ formula is as follows:

Equation 4-17

$$R^2_{adjusted} = R^2 - (1 - R^2) \cdot \frac{n-1}{n-k-1}$$

where *n* the sample size and *k* the number of predictors entered in the model.

The root mean square error (RMSE), which represents the average of the differences between the predicted and observed values, defined by the following equation, was also calculated:

Equation 4-18

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(\alpha_i - \alpha_{iobs})^2}{N}}$$

Since a large number of field plots were not available, the evaluation was carried out using the cross validation procedure, which consists of using (n-1) subsets of plots for the generation of the model and 1 for the validation, repeating the process n times so that the subset of validation plots is always different and independent from the rest.

4.3.5 Results and Discursion

The result for Nogueruelas was a unique model given that only one species is present in the pilot area and the forest maintains a similar structure except for 2 plots with less density of trees which are the plots with less biomass (Figure 23).

The model generated for Nogueruelas (ML1) obtained a R^2 of 0.827 and a R_{adj}^2 of 0.815, both statistics with cross validation leaving one out. The $RMSE_{cv}$ for this model was 1,074.45 kg of biomass.

Equation 4-19

$$Biomass = -3917.56 + 546.27 \cdot Elev.P80 + 465.02 \cdot Percentage\ first\ returns\ above\ mean - 456.80 \cdot Percentage\ all\ returns\ above\ mean$$

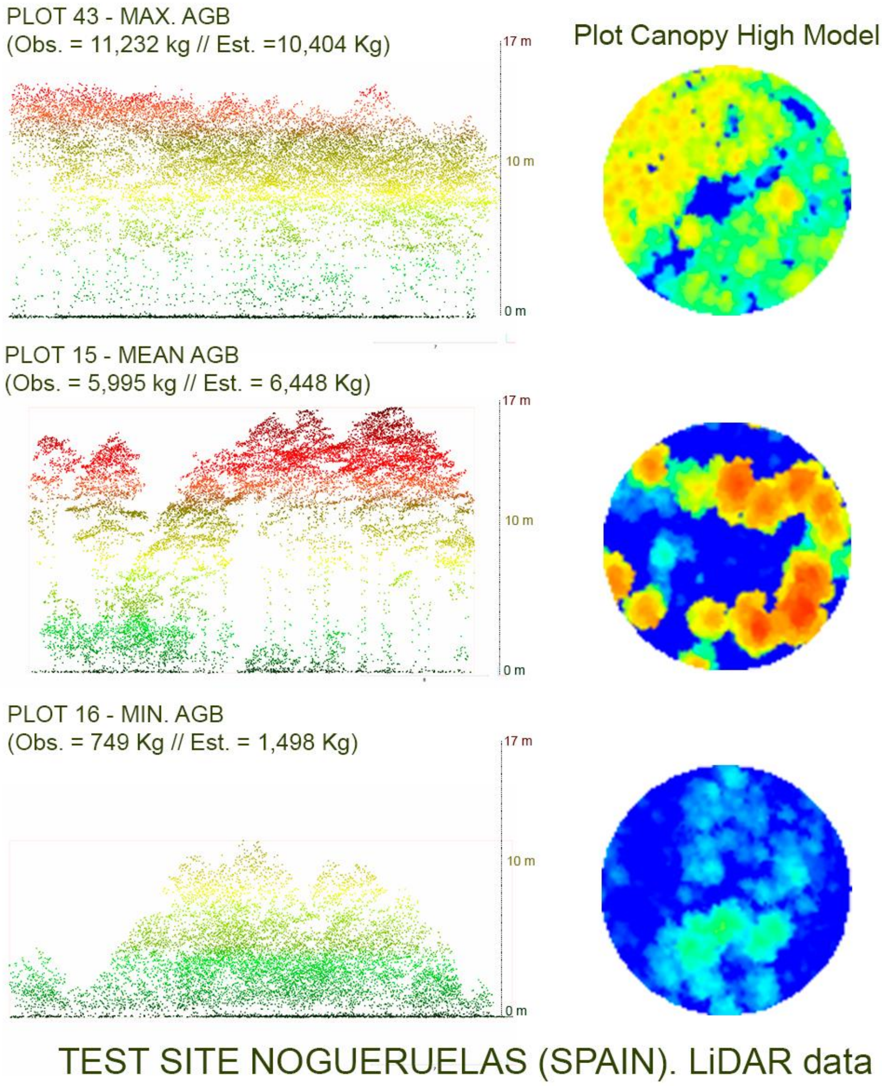

Figure 23. Plots with the maximum, mean and minimum biomass for the Nogueruelas pilot area. On the left is a vertical view of the plots with the observed and predicted biomass. On the right a zenithal view of the distribution of the trees and their canopy cover.

Three different models were generated for Espadán.

The model (ML2) was carried out with 78 plots, where the plots with the highest biomass value (plot 42 = 19,402kg) and lowest biomass value (plot 53 = 24.8kg) were discarded. The R^2 obtained is 0.704, R_{adj}^2 of 0.692 and $RMSE_{cv}$ equal to 2,140.33 kg.

Equation 4-20

$$Biomass = -476.32 - 0.93 \cdot Total\ return\ count + 23.93 \cdot Int.maximun + 2.15 \cdot Return\ 1\ count\ above\ 2.00\ meters$$

The best result was obtained for Espadán using only the *P. halepensis* plots (T1) without the regenerated *P. halepensis* plots. The model (ML3) for biomass was generated with 42 plots and reached a R^2 of 0.843, R_{adj}^2 of 0.831 and $RMSE_{cv}$ equal to 798.60 kg.

Equation 4-21

$$Biomass = -3298.40 + 291.74 \cdot Elev.P75 - 7700.29 \cdot Int.L.Skewness + 114.23 \cdot Percentage\ first\ returns\ above\ mean$$

For the mixed plots of *P. pinaster* and *Q. suber*, the predictive model was generated from 25 plots. For the biomass, the model (ML4) achieved a R^2 of 0.807, R_{adj}^2 of 0.779 and $RMSE_{cv}$ of 1,877.41 kg.

Equation 4-22

$$Biomass = -8662.72 + 40.10 \cdot Int.maximum + 522.62 \cdot Percentage\ first\ returns\ above\ mean - 61049.99 \cdot Int.L.CV$$

Figure 24 shows scatter plots of the plot-level field-based observed vs LiDAR-based predicted variables with linear fits. The worst model obtained corresponds to the model made for the entire Espadán pilot site (Figure 24 – ML2) and achieves a R_{adj}^2 of 0.692. The deficiencies of the model are explained by the diversity of species and the different types of structure present in the area. There are also great differences in density and basal area from one plot to another. Evaluating the scatter plot for this model (Figure 24. ML2) some outliers are evident corresponding to significant under-prediction of the highest observed AGB values.

The models ML3 improve more than 14% when they are carried out only for the species *P. halepensis.* The improvement in accuracy is achieved because they are monospecific plots with a similar vertical vegetation structure. In addition, a continuous range of different tree densities is represented in this set of plots, which improves the fit of the model. However, for the ML3 model the regenerated *P. halepensis* plots are not included, which are plots with a very high density and which generate great differences in the values of the LiDAR metrics.

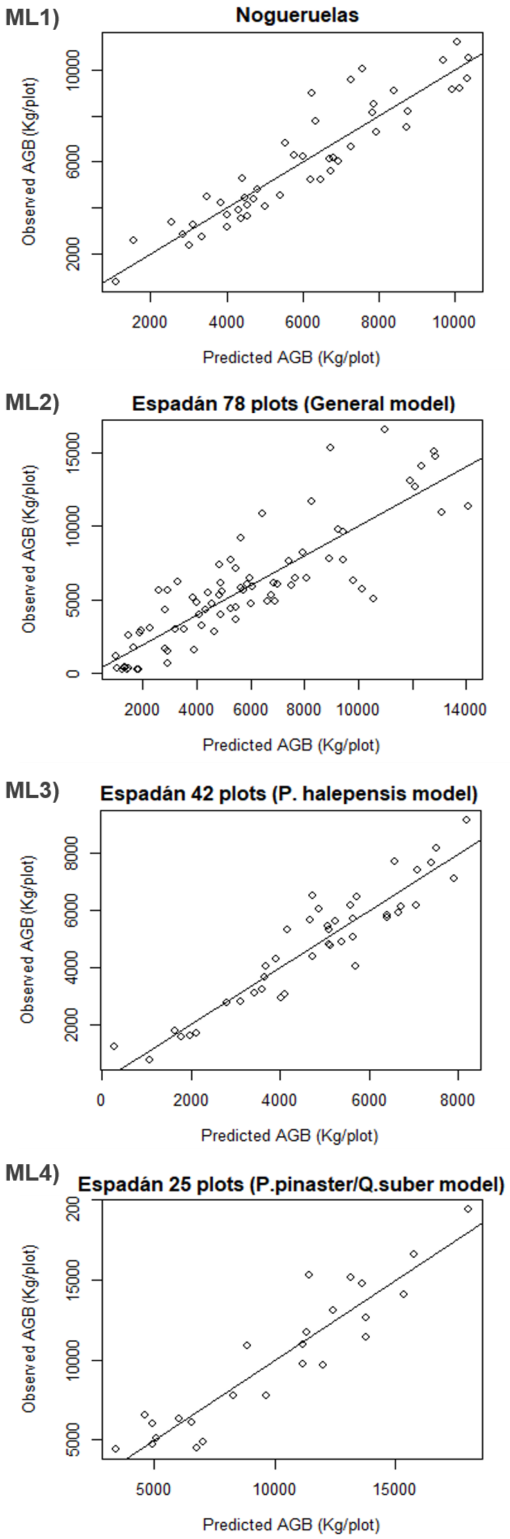

Figure 24. AGB Plot-level observed v. predicted values for the different models. Solid line represents the linear fitting.

On the other hand, the model for the mixed plots, ML4, of *P. pinaster* and *Q. suber* also improves almost 9% with respect to the general model for the whole area of Espadán. This model reduces the accuracy by 5% compared to the *P. halepensis* model, mainly due to the variability of the vertical and horizontal vegetation structure generated by the *Q. suber* species. This species has a lower height than *Pinus pinaster* and also modifies the continuity of the crown coverage generating gaps. These two facts significantly affect the LiDAR metrics and consequently generate a greater variability in the returns. To better adjust the model for this type of plots it would be convenient to have a larger number of plots.

Lastly the model ML1 for Nogueruelas performed on a monospecific area reaches a R_{adj}^2 of 0.815. This study area has plots of different vertical vegetation structure but well represented in the study sample, which makes the model fit with a remarkable precision. Nevertheless, the values of the models for *P. halepensis* in Espadán and Nogueruelas have a R_{adj}^2 between the habitual values obtained by other authors. For example, Montealegre Gracia et al. (2015) reaches a R_{adj}^2 0.89 in biomass estimation for *P. halepensis* monospecific stands with discrete low density LiDAR and parametric methods. Domingo et al. (2018) reaches a R_{adj}^2 of

0.87 with parametric methods for *P. halepensis* stands with significant bush density. García et al. (2010) also generate general models for mixed forest areas (*P. nigra*, *Juniperus thurifera* and *Quercus Ilex*) where it reaches R_{adj}^2 of 0.70 close to the 0.69 reached for the general model of Espadán. They also develop models by individual species where for *P. nigra* it reaches a R_{adj}^2 of 0.84 close to the single species models generated in this work. Also Ruiz et al. (2014) reach R_{adj}^2 values of 0.85 for pure plots of *P. Nigra* and *P.sylvestris* with a radius of 15 meters close to the value obtained in the Nogueruelas model.

The $RMSE_{cv}$ value for Nogueruelas is 1,074 kg. If we apply the formulas of Montero et al. (2005) for the species *P. sylvestris* the root mean square error of the estimation would correspond to approximately 5 trees of 25 cm DBH in a plot with 49 trees of mean density. If we make the same comparison for the $RMSE_{cv}$ of the general model of Espadán which is 2,140 kg it would be an error of 14 *P. halepensis* of 25 cm of DBH in a plot with a mean density of 68 trees. For the Espadán *P. halepensis* model error would be an equivalence of 5 trees and for the mixed model of 12 and 11 trees if they were *P. pinaster* or *Q. suber* respectively.

As a conclusion, a methodology for the estimation of AGB based on metrics obtained from discrete LiDAR data on the Spanish pilot site has been described and evaluated. Geographically closest works have been reported using discrete LiDAR with different degrees of success, in different Mediterranean areas, species and forest structure (Domingo et al., 2018; García et al., 2010; Montealegre Gracia et al., 2015; Ruiz et al., 2014). In our case, however, there was significantly more species variability: mean and standard deviation of the AGB parameters were highly compared to the other studies. This variability may make it difficult to define more precise models. For the lower species and forest structure variability, better biomass estimation models are fitted. The results show that the use of discrete lidar-based methodologies is accurate for estimating AGB. However, these models cannot be extrapolated to other areas with different species, density, or forest structure.

5. Comparison of methods

Having evaluated each of the three models' ability to estimate biomass at our study sites, it is now possible to compare the effectiveness of the three techniques to be able to draw some conclusions and provide recommendations for the future. In order to do so, we selected three criteria as the basis for the comparison: modelling technique accuracy, transferability and dependence on conditions variability at test site.

From the estimation accuracy point of view, both IWCM using *TanDEM-X* data and LiDAR data models showed high accuracy in the range between 75 and 85 per cent for both Espadán and Nogueruelas test sites in Spain. Additionally, LiDAR point cloud modelling allows to describe forest structure with a great precision. On the other hand, Sentinel-1's C-band biomass estimation using WCM performance was limited mainly by the signal saturation at higher biomass levels, thus, producing better estimation results at lower biomass levels, with the lowest rRMSE value of 42 per cent achieved at Nogueruelas test site.

Conversely, the WCM using Sentinel-1's backscatter intensity data is the most transferable approach of the three as it does not require any ground truth data for model calibration, therefore, theoretically, it can be applied to any marginal lands in Europe, or even globally. Relatively simple LiDAR methodology can be adjusted to any test site having ground truth data available to calibrate the model. However, airborne LiDAR data does not have global coverage (GEDI currently provides satellite LIDAR data on a global scale). High-density LiDAR data is not free. There are free sources of LiDAR data in some countries, e.g., Spain, but with lower density of points per square meter. Finally, the IWCM using *TanDEM-X* data besides using the ground truth data to calibrate the model also uses LiDAR retrieved DSM and DTM layers to estimate vegetation height as an additional parameter in the methodology. Therefore, it has the worst spatial transferability of the three, mainly, due to the fact that none of *TanDEM-X* and high point density LiDAR data is freely available.

Finally, the WCM and the IWCM methods using short-wavelength spaceborne SAR data which are sensitive to local environmental conditions at the time of image acquisition as the electromagnetic waves are strongly affected by the physical properties of the sensed vegetation. Therefore, the meteorological conditions should always be checked and kept in mind when performing biomass estimation using these methods. Airborne LiDAR technology is usually used only when environmental and meteorological conditions allow to do so, therefore, there is no additional requirement to adjust the measurements for biomass estimation once the LiDAR point cloud is retrieved. Additionally, all three methods show some sensitivity to different types of trees, but only the WCM using solely Sentinel-1 C-band data is severely affected by increasing biomass due to signal saturation at high biomass levels.

To conclude, airborne LiDAR technology-based methodology showed the best biomass estimation performance in the presented study areas, shortly followed by *TanDEM-X* and LiDAR data fusion exploiting method, leaving the WCM using only Sentinel-1 C-band data behind. However, the latter shows higher transferability features, which shall be exploited further by the future spaceborne SAR missions in 2020s.

5.1 Summary of Pros and cons of each methodology

5.1.1 WCM Method

Pros:

- ✓ Transferability – model calibration on demand (can be done individually for each image).
- ✓ Free and available everywhere (at least Europe-wide) – Sentinel-1 data free and available globally, Copernicus Tree Canopy Density (TCD) & CORINE Land Cover (CLC) available Europe-wide, but there are alternatives for global solutions too.

Cons:

- Poor accuracy – generally somewhere around 30 – 80% rRMSE.
- Early signal saturation – short C-band wavelength has limited penetration leading to loss of signal sensitivity at higher biomass levels (above 100 Mg/ha) under non-optimal environmental and meteorological conditions at the time of image acquisition.
- Strong sensitivity to weather conditions – meteorological conditions can influence the retrieval accuracy significantly over the same area.

5.1.2 IWCM Method

Pros:

- ✓ Better accuracy: The method implemented here achieves better accuracy when compared to Sentinel-1 only methods such as WCM. The original method using

TanDEM-X data is even more promising and is able to achieve 15.8% to 21.2% rRMSE, which is better than many SAR based biomass estimation methods.

Cons:

- Availability of data: *TanDEM-X* data is not freely available. Lack of freely available data restricts the application of this method severely.
- Transferability: Some of the parameters used in this method have initial values and ranges that are site and forest type specific. For higher accuracy, these parameters need to be obtained from field data which may not be available for all regions.
- Complexity of method: Inverse modelling approach is complex, and depends heavily on initial conditions, i.e. in this case the observations taken into account, which might not be representative of the general conditions of the site. Need to derive model parameters for every site separately taking into consideration the site conditions during observation.

5.1.3 LiDAR

Pros:

- ✓ LiDAR technology allows to describe the forest structure of the forest with great precision.
- ✓ Good accuracy: The models generated with discrete LiDAR reached an accuracy (R^2_{adj}) of 0.82 for Nogueruelas and an accuracy of 0.83 and 0.78 for the models developed by species in Espadán.
- ✓ Free software is available for the processing of LiDAR data.
- ✓ The Methodology is simple to elaborate and can be extrapolated to other study sites, as long as field data and LiDAR data from the area in question will be available.

Cons:

- Transferability: The models generated for this project are not transferable to other areas with different species, density or forest structure because the models are based on field data from a specific forest type.
- Airborne LiDAR data does not have global coverage (GEDI currently provides satellite LIDAR data on a global scale). High-density LiDAR data is not free. There are free sources of LiDAR data in some countries (e.g. Spain) but with

lower density of points per square meter. For the MAIL project, only the areas of Nogueruelas and Espadán with high density LiDAR data were available.

- The accuracy of the biomass models is directly related to the variability of species and structures of the study forest, as can be seen in the general model carried out for the entire Espadán area. The lower species and forest structure variability, better biomass estimation models are fitted.

6. RECOMMENDATIONS

We have observed that the LiDAR methodology obtains better results than the WCM and IWCM methodologies. The IWCM method can produce marginally better accuracy using Sentinel-1 data, hence it is recommended only when *TanDEM-X* data is available, along with relevant field data. Only the WCM methodology allows extrapolation of the result to the entire EU and does not require more data than the Sentinel-1C image. This methodology has a relatively low accuracy, so the use of this methodology should be subject to the accuracies obtained for different areas of Europe, depending on the type of forest, density (number of trees), and meteorological conditions of image acquisition.

To implement the LiDAR methodology LiDAR data and biomass data as ground truth are necessary. We recommend not using the LiDAR biomass estimation equations for other test sites than those where the equations have been configured. To apply the LiDAR methodology in other forest areas it is recommended to adjust the equations with field data and with LiDAR data acquired on the area under study. For future studies it is recommended to work with acquired data on MLs. Another methodology that could be explored would be based on satellite LiDAR data (GEDI) that has global coverage. To use this satellite sensor it is also recommended to have ground truth data on the pilot sites under study.

On the other hand, the biomass estimation equations with LiDAR have been configured on forest areas and not on marginal lands so it is not extrapolated to the marginal areas inside the pilot sites either. However, the results obtained with these methodologies can help define the objective and future forest that is planned to be established in the MLs. These biomass values can be considered as the reference values in task 4.2, as this task aims to define the plantation modules on the marginal areas and perform a future short, medium and long term estimate of the biomass and CO_2 stock that will accumulate in the afforestations. In consequence, knowing the amount of biomass and carbon stored in the forests existing today in the test sites can be a great help in estimating the biomass that will be accumulated by the afforestations established on the marginal lands of the same test sites. Not forgetting the limitations of soil, climate, slope, etc. of the MLs defined in tasks 2.1 and 2.3.

To determine the current status of the biomass in the MLs defined in the pilot sites, we recommend using the global product Climate Change Initiative – Biomass (CCI-Biomass) from ESA climate office with a resolution of 100 x 100 m, i.e. one pixel has the size of 1 ha, equivalent to the minimum size used to define marginal lands (defined in task 2.3). This product is available for three different periods 2010, 2017 and 2018, is free and open access from ESA climate office webpage. This product has an exhaustive methodology and evaluation of the accuracy obtained in the biomass layers of the different years. On the other hand, NASA provides the Global Aboveground and Belowground Biomass Carbon Density Maps for the Year 2010 (Spawn, Sullivan, Lark, & Gibbs, 2020). This dataset provides temporally consistent and harmonized global maps of aboveground and belowground biomass carbon density for the year 2010 at a 300-m spatial resolution. Nevertheless, according to the definition of MLs, most of these lands should be areas devoid of vegetation cover.

The methodologies proposed in Task 2.6 for the estimation of biomass on marginal lands, given the spatial and temporal resolution at which these can work, can be considered as a good tool for the future monitoring and evaluation of the forest stands established in the MLs.

REFERENCES

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In N. Petrov B & F. Csaki (Eds.), *Proceedings of the 2nd International Symposium on Information.* (pp. 267–281). Budapest, Hungary: Akadémiai Kiadó.

[2] Askne, J., & Santoro, M. (2012). Experiences in Boreal Forest Stem Volume Estimation from Multitemporal C-Band InSAR. In *Recent Interferometry Applications in Topography and Astronomy*. https://doi.org/10.5772/35102

[3] Askne, J., Santoro, M., Smith, G., & Fransson, J. E. S. (2003). Multitemporal repeat-pass SAR interferometry of boreal forests. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2003.813397

[4] Askne, J.I.H., Dammert, P. B. G., Ulander, L. M. H., & Smith, G. (1997). C-band repeat-pass interferometric SAR observations of the forest. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(1), 25–35. https://doi.org/10.1109/36.551931

[5] Askne, Jan I.H., Soja, M. J., & Ulander, L. M. H. (2017). Biomass estimation in a boreal forest from TanDEM-X data, lidar DTM, and the interferometric water cloud model. *Remote Sensing of Environment*, *196*, 265–278. https://doi.org/10.1016/j.rse.2017.05.010

[6] Attema, E. P. W., & Ulaby, F. T. (1978). Vegetation modeled as a water cloud. *Radio Science*. https://doi.org/10.1029/RS013i002p00357

[7] Axelsson, P. (2000). DEM Generation from Laser Scanner Data Using adaptive TIN Models. *International Archives of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2005.10.005

[8] Bastrup-Birk, A., Reker, J., & Zal, N. (2016). European forest ecosystems: State and trends. In *EEA Report n° 5/2016*. https://doi.org/10.2800/964893

[9] Breiman, L. (2001). Random forests. *Machine Learning*. https://doi.org/10.1023/A:1010933404324

[10] Cartus, O., Santoro, M., Schmullius, C., & Li, Z. (2011). Large area forest stem volume mapping in the boreal zone using synergy of ERS-1/2 tandem coherence and MODIS vegetation continuous fields. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2010.12.003

[11] Crespo-Peremarch, P., Ruiz, L. Á., Balaguer-Beser, Á., & Estornell, J. (2018). Analyzing the role of pulse density and voxelization parameters on full-waveform LiDAR-derived metrics. *ISPRS Journal of Photogrammetry and Remote Sensing*, *146*, 453–464. https://doi.org/10.1016/j.isprsjprs.2018.10.012

[12] Domingo, D., Lamelas, M., Montealegre, A., García-Martín, A., & de la Riva, J. (2018). Estimation of Total Biomass in Aleppo Pine Forest Stands Applying Parametric and Nonparametric Methods to Low-Density Airborne Laser Scanning Data. *Forests*, *9*(4), 158. https://doi.org/10.3390/f9040158

[13] Dubayah, R. O., & Drake, J. B. (2000). Lidar Remote Sensing for Forestry Applications. *Journal of Forestry*, *98*(6), 44–46. https://doi.org/doi.org/10.1093/jof/98.6.44

[14] Favero, A., Daigneault, A., & Sohngen, B. (2020). Forests: Carbon sequestration, biomass energy, or both? *Science Advances*, *6*(13), eaay6792. https://doi.org/10.1126/sciadv.aay6792

[15] Fransson, J. E. S., Smith, G., Askne, J., & Olsson, H. (2001). Stem volume estimation in boreal forest using ERS-1/2 coherence and SPOT XS optical data. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431160010006872

[16] García, M., Riaño, D., Chuvieco, E., & Danson, F. M. (2010). Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sensing of Environment*, *114*(4), 816–830. https://doi.org/10.1016/j.rse.2009.11.021

[17] Hermosilla, T., Ruiz, L. Á., Kazakova, A. N., Coops, N. C., & Moskal, L. M. (2014). Estimation of forest structure and canopy fuel parameters from small-footprint full-waveform LiDAR data. *International Journal of Wildland Fire*, *23*(2), 224. https://doi.org/10.1071/WF13086

[18] Isenburg, M. (2018). *LAStools - Efficient tools for LiDAR processing. (Version 180409) obtained from http://rapidlasso.com/LAStools*. Alemania: Rapidlasso GmbH.

[19] Kaasalainen, S., Holopainen, M., Karjalainen, M., Vastaranta, M., Kankare, V., Karila, K., & Osmanoglu, B. (2015). Combining Lidar and Synthetic Aperture Radar Data to Estimate Forest Biomass: Status and Prospects. *Forests*, *6*(12), 252–270. https://doi.org/10.3390/f6010252

[20] Klaassen, W., van der Linden, M., Ballast, A., & Esa. (1997). SAR sensing of vegetation wetness: The first results. In *Third Ers Symposium on Space at the Service of Our Environment, Vol 1*.

[21] Koch, B. (2010). Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2010.09.001

[22] Kurvonen, L., Pulliainen, J., & Hallikainen, M. (1999). Retrieval of biomass in boreal forests from multitempotal ERS-1 and JERS-1 SAR images. *IEEE Transactions*

on Geoscience and Remote Sensing. https://doi.org/10.1109/36.739154

[23] Le Toan, T., Quegan, S., Davidson, M. W. J., Balzter, H., Paillou, P., Papathanassiou, K., … Ulander, L. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, *115*(11), 2850–2860. https://doi.org/10.1016/j.rse.2011.03.020

[24] McGaughey, R. (2016). FUSION/LDV: Software for LIDAR Data Analysis and Visualization. *United States Department of Agriculture, Forest Service, Pacific Northwest Research Station*, p. 211. Retrieved from http://forsys.cfr.washington.edu/fusion/FUSION_manual.pdf

[25] Montealegre Gracia, A. L., Lamelas Gracia, M. T., De la Riva Fernandez, J., Garcia-Martin, A., & Escribano Bernal, F. (2015). Cartografía de la biomasa aérea total en masas de Pinus halepensis Mill. en el entorno de Zaragoza mediante datos LiDAR-PNOA y trabajo de campo. *Análisis Espacial y Representación Geográfica: Innovación y Aplicación*, 769–776.

[26] Montero, G., Ruiz-Peinado, R., & Muñoz, M. (2005). *Produccion de biomasa y fijación de CO2 por los bosques españoles* (Instituto). Madrid: Instituto Nacional de Investigación y Tecnología Agraria y Alimenta. Ministerio de Educación y Ciencia.

[27] Pulliainen, J.T., Mikhela, P. J., Hallikainen, M. T., & Ikonen, J.-P. (1996). Seasonal dynamics of C-band backscatter of boreal forests with applications to biomass and soil moisture estimation. *IEEE Transactions on Geoscience and Remote Sensing*, *34*(3), 758–770. https://doi.org/10.1109/36.499781

[28] Pulliainen, Jouni T., Heiska, K., Hyyppa, J., & Hallikainen, M. T. (1994). Backscattering Properties of Boreal Forests at the C and X-Bands. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/36.312892

[29] Ruiz-Peinado Gertrudix, R., Montero, G., & Del Rio, M. (2012). Biomass models to estimate carbon stocks for hardwood tree species. *Forest Systems*, *21*(1), 42. https://doi.org/10.5424/fs/2112211-02193

[30] Ruiz, L. Á. (2020). *Obtención de variables forestales a partir de datos LiDAR*. Retrieved from http://hdl.handle.net/10251/147476

[31] Ruiz, L. Á., Hermosilla, T., Godino, M., Almonacid, J., Fernández-Sarría, A., Recio, J. A., … Mauro, F. (2011). Procedimiento para la estimación de variables dasométricas a partir de datos LiDAR. *Actas Del XIV Congreso Nacional de La Asociación Española de Teledetección: "Bosques y Cambio Climático,"* 129–152. Mieres del Camino (Principado de Asturias).

[32] Ruiz, L. Á., Hermosilla, T., Mauro, F., & Godino, M. (2014). Analysis of the

influence of plot size and LiDAR density on forest structure attribute estimates. *Forests*, *5*(5), 936–951. https://doi.org/10.3390/f5050936

[33] Ruiz, L. Á., Recio, J. A., Crespo-Peremarch, P., & Sapena, M. (2018). An object-based approach for mapping forest structural types based on low-density LiDAR and multispectral imagery. *Geocarto International*, *33*(5), 443–457. https://doi.org/10.1080/10106049.2016.1265595

[34] Saatchi, S. (2019). SAR Methods for Mapping and Monitoring Forest Biomass. In A. Flores, K. Herndon, R. Thapa, & E. Cherrington (Eds.), *SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation.* https://doi.org/10.25966/hbm1-ej07

[35] Saatchi, S. S., Le Vine, D. M., & Lang, R. H. (1994). Microwave backscattering and emission model for grass canopies. *IEEE Transactions on Geoscience and Remote Sensing*, *32*(1), 177–186. https://doi.org/10.1109/36.285200

[36] Santoro, M., Cartus, O., Mermoz, S., Bouvet, A., Le Toan, T., Carvalhais, N., … Seifert, F. M. (2018). A detailed portrait of the forest aboveground biomass pool for the year 2010 obtained from multiple remote sensing observations. Geophysical Research Abstracts, 20, EGU2018-18932. *Geophysical Research Abstracts*.

[37] Santoro, Maurizio, Askne, J., Smith, G., & Fransson, J. E. S. (2002). Stem volume retrieval in boreal forests from ERS-1/2 interferometry. *Remote Sensing of Environment*, *81*(1), 19–35. https://doi.org/10.1016/S0034-4257(01)00329-7

[38] Santoro, Maurizio, Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., … Wiesmann, A. (2011). Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2010.09.018

[39] Santoro, Maurizio, & Cartus, O. (2018). Research pathways of forest above-ground biomass estimation based on SAR backscatter and interferometric SAR observations. *Remote Sensing*. https://doi.org/10.3390/rs10040608

[40] Santoro, Maurizio, Cartus, O., & Lucas, R. (2019). *Algorithm Theoretical Basis Document. Biomass Project*.

[41] Shugart, H. H., Saatchi, S., & Hall, F. G. (2010). Importance of structure and its measurement in quantifying function of forest ecosystems. *Journal of Geophysical Research: Biogeosciences*, *115*(G2), n/a-n/a. https://doi.org/10.1029/2009JG000993

[42] Silva, C., Hudak, A., Vierling, L., Klauberg, C., Garcia, M., Ferraz, A., … Saatchi, S. (2017). Impacts of Airborne Lidar Pulse Density on Estimating Biomass Stocks and Changes in a Selectively Logged Tropical Forest. *Remote Sensing*, *9*(10), 1068.

https://doi.org/10.3390/rs9101068

[43] Small, D. (2011). Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2011.2120616

[44] Soja, M. J., Askne, J. I. H., & Ulander, L. M. H. (2017). Estimation of Boreal Forest Properties from TanDEM-X Data Using Inversion of the Interferometric Water Cloud Model. *IEEE Geoscience and Remote Sensing Letters*. https://doi.org/10.1109/LGRS.2017.2691355

[45] Spawn, S. A., Sullivan, C. C., Lark, T. J., & Gibbs, H. K. (2020). Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Scientific Data*. https://doi.org/10.1038/s41597-020-0444-4

[46] Torralba, J., Crespo-Peremarch, P., & Ruiz, L. Á. (2018). Assessing the use of discrete, full-waveform LiDAR and TLS to classify Mediterranean forest species composition. *Revista de Teledetección*, (52), 27. https://doi.org/10.4995/raet.2018.11106

[47] Ulander, L. M. H., Hagberg, J. O., & Askne, J. (1994). ERS-1 SAR interferometry over forested terrain. *Proceedings of the Second ERS-1 Symposium*, 475–480. Hamburg: ESA.

[48] Wehr, A., & Lohr, U. (1999). Airborne laser scanning - An introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/S0924-2716(99)00011-8

[49] White, J. C., Wulder, M. A., Varhola, A., Vastaranta, M., Coops, N. C., Cook, B. D., … Woods, M. (2013). A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach. *The Forestry Chronicle*, *89*(06), 722–723. https://doi.org/10.5558/tfc2013-132

[50] Wulder, M. A., Bater, C. W., Coops, N. C., Hilker, T., & White, J. C. (2008). The role of LiDAR in sustainable forest management. *The Forestry Chronicle*, *84*(6), 807–826. https://doi.org/10.5558/tfc84807-6

[51] Wulder, M. A., White, J. C., Nelson, R. F., Næsset, E., Ørka, H. O., Coops, N. C., … Gobakken, T. (2012). Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment*, *121*, 196–209. https://doi.org/10.1016/J.RSE.2012.02.001

[52] Zhang, X., & Ni-meister, W. (2014). Remote sensing of Forest Biomass. In J. M. Hanes (Ed.), *Biophysical Applications of Satellite Remote Sensing* (p. 236). https://doi.org/10.1007/978-3-642-25047-7

[53] Zhao, K., Suarez, J. C., Garcia, M., Hu, T., Wang, C., & Londo, A. (2018). Utility of multitemporal lidar for forest and carbon monitoring: Tree growth, biomass dynamics, and carbon flux. *Remote Sensing of Environment*, *204*, 883–897. https://doi.org/10.1016/J.RSE.2017.09.007

ANNEX I: TABLE OF FIGURES

ANNEX II: LIST OF TABLES