DUE GlobBiomass

D6 – Regional Biomass Maps

Algorithm Theoretical Basis Document

(incorporating the Design Justification File D7)

Prepared for European Space Agency (ESA-ESRIN)

In response to ESRIN/Contract No. 4000113100/14/I_NB

Revision History

Deliverable	**D6, D7 – Regional Biomass Maps**		
Work Package	3000		
Due date	KO+25		
Authors	Shaun Quegan, Yrjö Rauste, Valerio Avitabile, Heiko Balzter, Alexandre Bouvet, Joao Carreiras, Oliver Cartus, Nuno Carvalhais, Sandra Englhart, Johan Fransson, Martin Herold, Agata Hoscilo, Thuy LeToan, Renaud Mathieu, Stephane Mermoz, Henrik Persson, Pedro Rodriguez-Veiga, Danaë Rozendaal, Maurizio Santoro, Matthias Stängel, Krzysztof Stereńczak, Kevin Tansey		
Distribution	FSU: Christiane Schmullius, Evelin Matejka, Carsten Pathe, Christian Thiel		ESA: Frank Martin Seifert; Nathalie Boisard
Reason for change			
Issue			
Revision			
Date			
Release	1		
Version	08		

Contents

List of acronyms	
AGB	Above ground biomass
ALOS	Advanced Land Observing Satellite
APM	Alternating Polarization Medium
ASAR	Advanced synthetic aperture radar
ATBD	Algorithm Theoretical Basis Document
CART	Classification and Regression Trees
CCI	Climate Change Initiative
DEM	Digital Elevation Model
ECMWF	European Centre for Medium-range Weather Forecast
Envisat	Environmental satellite
EO	Earth Observation
ESA	European Space Agency
FAO	Food and Agriculture Organization (
FIA	Forest Inventory and Analysis
GLAS	Geoscience Laser Altimeter System
GM1	Global Monitoring
GSV	Growing stock volume
ICESAT	Ice, Cloud and Land Elevation Satellite
IIASA	International Institute of Applied System Analysis
IMM	Image Mode Medium
IPCC	Intergovernmental Panel on Climate Change
JAXA	Japan Aerospace Exploration Agency
*k*NN	*k*-Nearest Neighbour
LC	Land Cover
MIPERS	Multi-static Interferometric and Polarimetric Electromagnetic model for Remote Sensing
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NBCD	National Biomass Carbon Dataset
NDVI	Normalized Difference Vegetation Index
NPP	Net Primary Production
PALSAR	Phased Array type L-band Synthetic Aperture Radar
RMSD	Root Mean Square Difference
RMSE	Root Mean Square Error
SAR	Synthetic aperture radar
SRTM	Shuttle Radar Topography Mission
USGS	United States Geological Survey
VCF	Vegetation Continuous Fields
WCM	Water Cloud Model
WSM	Wide Swath Mode

1 Introduction and science background

Forest biomass (here understood as above-ground woody biomass [AGB]) is a fundamental biophysical variable describing the amount of woody matter within a forest. It is crucial to human well-being as a source of materials (e.g. for building) and energy (around 90% of the energy consumption in sub-Saharan Africa comes from biomass burning). Its importance in global terms has become even greater in recent years due to international concerns about climate warming. Biomass losses associated with Land Use Change (principally in the tropics) translate into CO2 emissions, and are second only to fossil fuel burning as a source of such emissions (quantified as between 7 and 20% of the total anthropogenic annual emissions). Biomass gains from forest growth remove CO2 from the atmosphere and form an important part of the "residual land sink", thus mitigating climate warming. Indeed, planting of new forests provided the only means of offsetting carbon emissions under the Kyoto Protocol, and preservation and better management of tropical forest to conserve carbon stored as biomass is a fundamental tenet of the UN Reduction of Emissions from Deforestation and Degradation initiative (REDD+). As a result, biomass is recognised as an Essential Climate variable (GCOS 2003). A further important contribution of biomass to climate science is as a means of initialising the land component of couple Earth System Models, quantifying their estimates of carbon fluxes from Land Use Change, and estimating the turnover time of carbon within forest systems (Thurner et al., 2016).

The purpose of the ESA DUE GlobBiomass project is to provide estimates of Aboveground Biomass (AGB) that, if possible, improve on existing products and have better characterised and reduced uncertainties. Biomass products will be developed for five regional sites for the epochs 2005, 2010 and 2015 and there will be one global map for the year 2010. The regional products are intended to provide the best possible estimates of biomass over a varied set of forest types for 2005, 2010 and 2015, making use of all data available to the regional teams, as well as estimates of biomass change between epochs. These will provide a reference against which the global product, which will be a single map for 2010, can be assessed. For all products a crucial second type of product will be maps describing the accuracy of the products.

2 Gap analysis

Until comparatively recently, our knowledge about above-ground forest biomass (AGB) and growing stock volume (GSV) was largely contained in national forest inventories and regional to national scale summaries, such as those contained in FAO's quinquennial Global Forest Resource Assessment (GFRA) reports (FAO 2006, 2010, 2015). However, these data have several important limitations:

- individual plot data are not distributed;
- the uncertainties in the country data are not reported and are likely to be highly variable, particularly for countries without a well-developed forest monitoring infrastructure (as is the case in many tropical forest countries), although the data underlying the individual country estimates can be accessed through FAO;
- inventory data are concentrated in the boreal and temperate zones to meet the needs of commercial forestry, and are extremely sparse in the tropics (Figure 2-1; Schimel et al., 2015), where they are mainly produced by ecological research networks (though this situation is improving, partly as a result of the UNFCCC Reducing Emissions from Deforestation and Forest Degradation initiative [REDD+] fostering the development of national inventories in the tropics);
- the data are not gridded, i.e., they do not represent spatially localised maps of biomass.

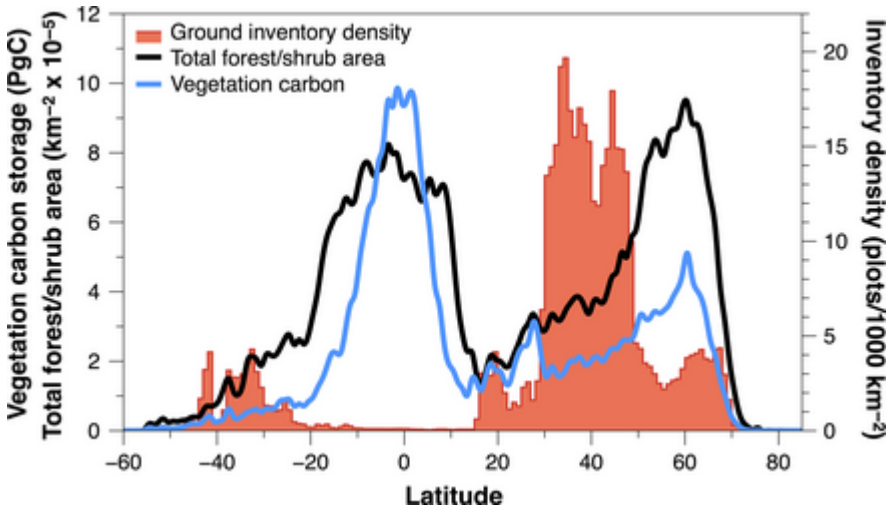


Figure 2-1. Estimates of forest/shrub area, vegetation carbon together with a histogram of the density of inventory data, each shown as a function of latitude (from Schimel et al., 2015).

Although Kindermann et al. (2008) produced global maps of AGB and GSV on a 0.5° grid by extrapolating values from country-level statistics in the FAO GFRA 2005 report with the aid of a global 0.5° map of Net Primary Productivity (NPP) and maps indicating human influence, these have largely unknown and varying errors because of the highly variable quality of the forest data from the individual nations.

This situation began to change within the last decade as remote sensing data from airborne (especially lidar) and spaceborne platforms were employed in biomass mapping. The search for consistent approaches for the tropics prompted use of satellite data calibrated against *in situ* biomass, with special emphasis on the archive of forest height estimates derived from the Geoscience Laser Altimeter System onboard the Ice, Cloud and land Elevation Satellite (ICESat) before its failure in 2009 (Lefsky

2010). This gave rise to two pan-tropical biomass maps (Saatchi et al. 2010; Baccini et al. 2012) at grid scales of 1 km and 500 m respectively. These maps exhibit significant regional differences, although when aggregated to country or biome scale these tend to decrease (Mitchard et al. 2013, 2014).

Gridded biomass datasets have also been generated for other continental-scale regions. Using very long time series of C-band radar data from the ESA Envisat ASAR, Santoro et al. (2011) produced pan-boreal biomass maps at a scale of 10 km; this provided the basis for the carbon stock of Northern Hemisphere forests north of 30°N in Thurner et al. (2013). The National Land Cover Database, the US National Forest Inventory and topographic data from the Shuttle Radar Topography Mission were combined to generate a biomass map for the coterminous USA for the year 2000 (Kellndorfer et al. 2012).

The desire for a more homogeneous global product led to the generation within the EC-funded GEO-CARBON project of a near-global dataset of forest AGB with a posting of 0.01° by assembling and merging the Saatchi et al. (2011) and Baccini et al. (2012) pan-tropical datasets with a northern mid-latitude and boreal dataset (Santoro et al., 2015b) (http://www.geocarbon.net). This is affected by limitations in the input EO data for mapping biomass and approximations in the individual retrieval approaches and the fact that the individual maps are based on data acquired at different times (2000, 2007-2008 and 2010 respectively). The Saatchi et al. (2011) and Baccini et al. (2012) have also been combined into a pan-tropical AGB map at 1 km resolution using an independent reference dataset of field observations and locally-calibrated high-resolution biomass maps (Avitabile et al., 2015).

At country scale, maps of biomass for Sweden have been derived for 2005, 2010 and 2015 using optical data in combination with the Swedish National Forest Inventory, and several countries have generated (e.g. Sweden) or are in process of generating (e.g. Gabon) national biomass maps from airborne lidar data. These are relevant in the context of the regional studies reported in this ATBD.

3 Regional biomass maps

Regional maps will be produced for epochs 2005, 2010 and 2015 in five contrasting regions:

- Poland: Temperate zone
- Sweden: Boreal zone
- Kalimantan (Indonesia): Tropical zone
- Mexico: Tropical-woodland transition
- South Africa: (changed from Northern Congo because of logistical difficulties)

The emphasis in this ATBD is on 2010; different approaches are likely to be needed for the other epochs because of differences in the data available. Maps of change between epochs will also be produced at a later stage. The target for accuracy is a relative RMSE (coefficient of variation) of 20% or better at a spatial resolution of 30-250m. Spatial resolution can, of course, be traded for accuracy (but only if the biomass estimates are unbiased). The biomass maps should be supplied along with associated maps depicting accuracy.

The sections of the ATBD for each regional biomass map follow the same structure consisting of four main sections:

1. General description of each region, to include existing knowledge about biomass, range of biomass encountered, etc.
2. Datasets
    a. Input data (used to drive the biomass algorithm);
    b. Training data (used to estimate parameters, e.g. in a regression equation);
    c. Validation data (set aside data that will be used to test independently if the estimates are accurate);
    d. Data used for accuracy assessment, which will be used by the regional teams to assign accuracy to their product, so may include ground data, but must be separate from data which will be made available to the Validation (WP 7000).
3. Methods
    a. Pre-processing of data
    b. Biomass estimation algorithms
    c. Training methods
    d. Methods to assign accuracies in the uncertainty map
    e. Methods to test the accuracy of the measurements
4. Products

Commonalities and differences between the regional approaches are discussed in Section 4, with Section 5 analysing the properties of the regional biomass maps and Section 6 describing the approach followed to map biomass change across epochs. The relationships between the regional and global approaches are set out in Section 7. Our conclusions are given in Section 8.

3.1 Poland

3.1.1 General Description

The regional biomass map for 2010 covers all forested areas in Poland. This amounts to about 91,630 km², which is 29.3% of the area of Poland. More than 77% of the total forest area is managed by the State Forests National Forest Holding, whereas the rest belongs to the national parks, private owners and cooperative owners. Coniferous forest habitats predominate, accounting for 70% of the total forest area, while broadleaved forest habitats account for 49%. In both groups, upland habitats occupy around 6% of the forest area and mountain habitats 8.7%. The forest on average is about 60 years old and average stem volume reaches 270 m^3 / ha. The average stem volume is available at forest stand level over the State Forest (every forest stand is updated in 10 year cycle) and at the plot level as part of the large-scale National Inventory of Forest Condition.

Due to the limited number of biomass allometric equations specific for various tree species estimating above ground woody biomass, especially on a nation-wide level, is generally based on conversion factors applied to the volume of the growing stock. The use of standard conversion factors, as proposed by international guidelines is questionable, as the inventory systems and definitions of growing stock differ from country to country (Jabłoński and Budniak, 2014). In Poland, biomass conversion and expansion factors (BCEF) proposed by the National Center for Emission Management (KOBIZE report, 2014) are used to convert stem volume to stem biomass at national level. There is no forest biomass map for Poland.

3.1.2 Datasets

3.1.2.1 *Input data*

Remote sensing imagery sensitive to forest AGB from different SAR sensors are used. The core dataset is freely available ALOS PALSAR mosaic downloaded from Jaxa (http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/fnf_index.htm). The mosaic contains normalized radar cross-section, γ^0 for HH and HV, mask information (ocean flag, effective area, void area, layover, shadowing), local incidence angle, and total dates from the ALOS launch (Shimada and Ohtaki, 2010). The annual ALOS mosaic is available at 25 m spatial resolution for 2007, 2008, 2009 and 2010.

In addition, the archive ASAR IM PRI, ASAR APP PRI and ERS-2 PRI images acquired over Poland during the period 2009-2010 were obtained from ESA (Cat-1 proposal). In total, 42 ERS-2 scenes on ascending orbits and 179 scenes on descending orbits acquired between March and October 2009 and 2010 were obtained. In addition, 59 ASAR scenes with dual polarization were downloaded.

Landsat cloud-free image composites (circa year 2013) provided reference multispectral imagery from the last available year, typically 2013. If no cloud-free observations were available for year 2013, imagery was taken from the closest year with cloud-free data, within the range 2010–2012. The mosaic is available from http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.1.html (Hansen et al, 2013). Reference composite imagery are median observations from a set of quality assessed growing season observations in four spectral bands, specifically Landsat bands 3, 4, 5, and 7. Normalized top-of-atmosphere (TOA) reflectance values (ρ) have been scaled to an 8-bit data range using a scale factor (g):

$$DN = \rho \cdot g + 1$$

The biomass algorithm has been trained using the ALOS PALSAR mosaics and Landsat cloud-free composite (bands 3, 4, 5 and 7).

ASAR and ERS data sets were not used due to i) limited and scattered distribution of scenes over the entire study area; there were areas, for example on the coastal zone, covered by many images and areas with no data, ii) just a single coverage of C-band data for some places, and iii) the relationship between backscatter coefficient (C-band) and biomass values was rather poor (R2=0.07 – 0.1).

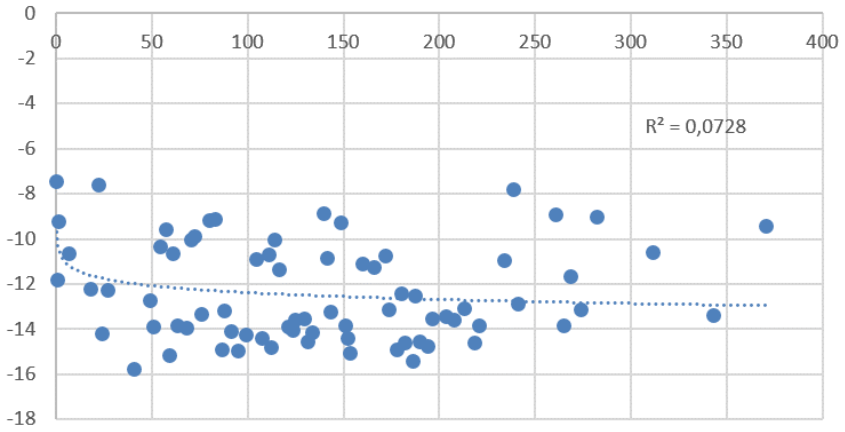


Figure 3-1. γ^0 for VV polarization in 2010 plotted as a function of AGB, using AGB estimated from the 392 plots grouped into 111 homogenous classes.

Table 3-1. EO data used to derive the biomass algorithm for three epochs.

DATASET	SPATIAL RESOLUTION	2005	2010	2015
ALOS PALSAR mosaic	25 m	x	x	
Shuttle Radar Topography Mission-SRTM(1)	30 m	x	x	x
Sentinel-1 Dual Polarisation IW Mode	5 m x 20 m			x
Landsat 5	30 m	x	x	
Sentinel-2	10 m			x

(1) Elevation data from SRTM corresponds to the year 2000. As it is assumed that topography remains constant, so the same dataset will be used for all epochs

3.1.2.2 *Training data*

A large-scale National Inventory of Forest Condition was carried out in 2005, 2006, 2009, 2010, 2011, 2014 and 2016. The aim of the inventory is to assess the condition of the forest under all forms of ownership and direction of the large-scale changes based on carefully selected indicators. The methodology of large-scale inventory of forest condition was developed at the Forest Research Institute, including the results of national and international (Austria, Finland, Germany, Switzerland and the United States) surveys in this field. The location of the forest inventory plots correspond to the ICP Forests (International Co-operative Programme on Assessment and Management of Air Pollution effects on Forests) programme, which monitors forest on a systematic grid of 16 x 16 km. For the purposes of the national forest inventory, within each 16 x 16 km grid, 25 L-shaped groups of sampling

plots located in a systematic grid of 4 x 4 km were established (Figure 3-2). Each L-shaped group of sampling plots consists of five sampling plots located 200 m apart.

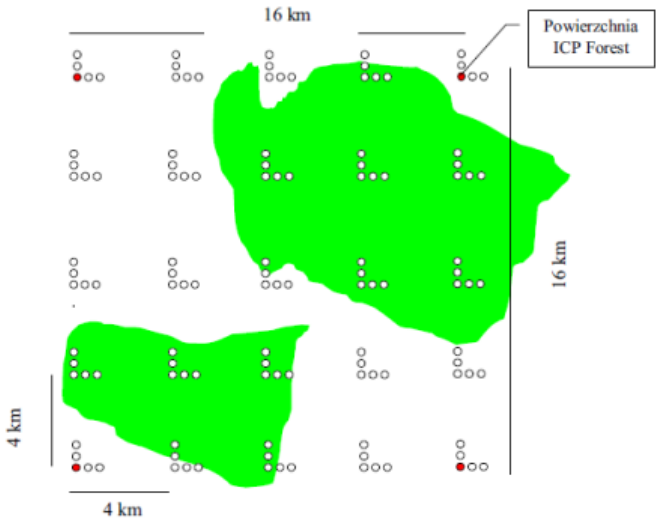


Figure 3-2. Layout of sampling plots.

The forest inventory plots are circular with radius equal to 12.6 m (0.05 ha) meters. Because the location of the plots is determined by the regular grid, some of them are partly covered by forest, but in this project only plots located entirely within the forest were considered. For each inventory year around 4000 plots lie within forest. For the 2010 biomass map, the field data collected in 2009 and 2010 were used.

The parameters measured *in situ*, such as DBH, height and number of trees, were transformed to growing stock volume (m^3/plot). The growing stock volume was converted into woody biomass using the IPCC approach based on biomass expansion factors (BEFs) and wood density (WD) following IPCC guidelines (Penman et al. 2003; IPCC 2006). In IPCC (2006) report the BEFs in the temperate zone are set to 1.3 for coniferous forest and 1.4 for deciduous forest. However, the recent FAO Global Forest Resources Assessment (FAO, 2015) country report for Poland indicates that the BEF should be 1.3 for all tree species, and this value was used in all biomass estimates reported here. Before applying such BEFs, merchantable growing stock volume (m^3) must be converted to dry-weight (tons) by multiplying by a conversion factor known as basic wood density (t/m^3). The basic wood density values for the 10 most common tree species in Poland (3 coniferous and 7 deciduous) according to Poland's National Inventory Report were obtained from the report of the National Center for Emission Management (KOBIZE, 2014). The woody biomass was calculated for each sampling plot based on growing stock volume of the dominant tree species.
For further analysis based on regression equations, the reference plots were selected to meet two conditions:

i. They must be on terrain with slopes less than 5^o (the landform features for each sampling plot are described in the inventory);
ii. The forest surrounding the plot must be homogenous.

To overcome the problem of small plots, the plot size was enlarged to 50 m radius; the homogeneity of each large plot was assessed through analysis of the variation coefficient and then it was visually inspected and compared to the national aerial ortho-photomaps. Only plots that clearly lie in homogenous forest and are located at least 100 m away from the forest edge were considered for biomass estimation.

3.1.2.3 *Validation data*

The reference data were divided into training and validation sets. 70% of sampling plots were used for training of the algorithm and 30% left aside for the validation.

3.1.2.4 *Data used for accuracy assessment*

Since there is no biomass map over Poland, ALS data available for 6 test sites scanned in 2007 and 2015 will be used for the accuracy assessment. In addition, the independent information on GSV for forest stands obtained from the National Forest Inventory Database will be used to assess the accuracy at stand level.

3.1.3 Methods

3.1.3.1 *Pre-processing of data*

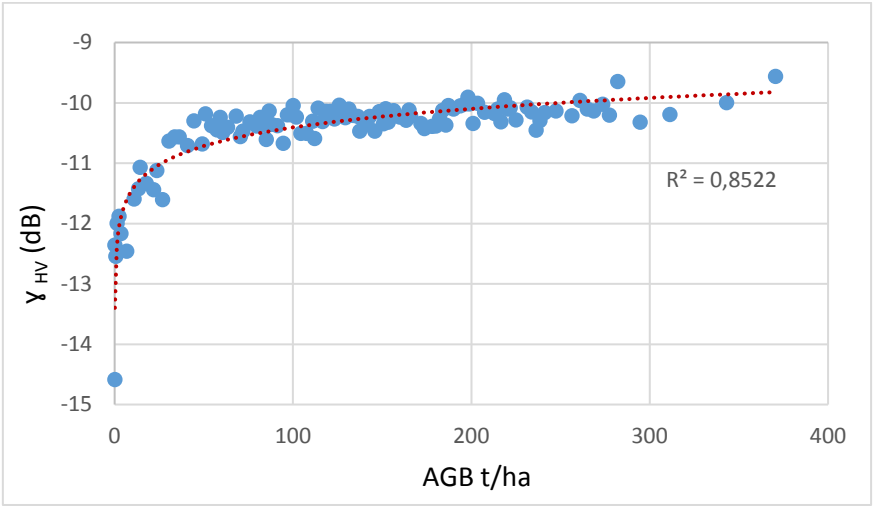
The uncertainties due to speckle effects in ALOS-Palsar mosaics were reduced by applying multi-channel filtering with a 7x7 window (Quegan and Yu, 2001). The ALOS mosaics after filtering were used to calculate additional products, i.e., HH-HV, HH/HV.

A set of ASAR and ERS-2 data was pre-processed with Gamma software. The geocoding was done using SRTM data. The next step calculated the illuminated area contributing to each individual pixel in slant-range geometry using the SRTM DEM; this gave improved σ^0 and γ^0 normalization areas for scenes exhibiting strong foreshortening and layover. Next, radiometric calibration of multi-look intensity (MLI) images was performed (calibration steps include range spreading loss correction, antenna gain correction, normalization reference area correction, correction for calibration constant). Finally, topographic normalization was done to compensate backscatter effects induced by the terrain (based on the SRTM DEM). The Refined Lee filter was used to reduce speckle in the individual SAR datasets. In the next step, the mosaics of ASAR and ERS-2 will be created over Poland and then used as an additional input to the model.

Landsat cloud-free image composite was downloaded in four tiles, merged, and resampled to 25 meters.

3.1.3.2 *Biomass estimation algorithms*

To generate the biomass map, the relationship relating γ^0 to biomass was first established using logarithmic regression. At present, the biomass map for 2010 is produced using the ALOS mosaics for 2009 (HH, HV) and 2010 (HH, HV) and Landsat reflectance (4 bands). Figures 3-3a-b show the relationships between γ^0 for HV and HH polarisation for 2010 and 2009, respectively, and calculated AGB.



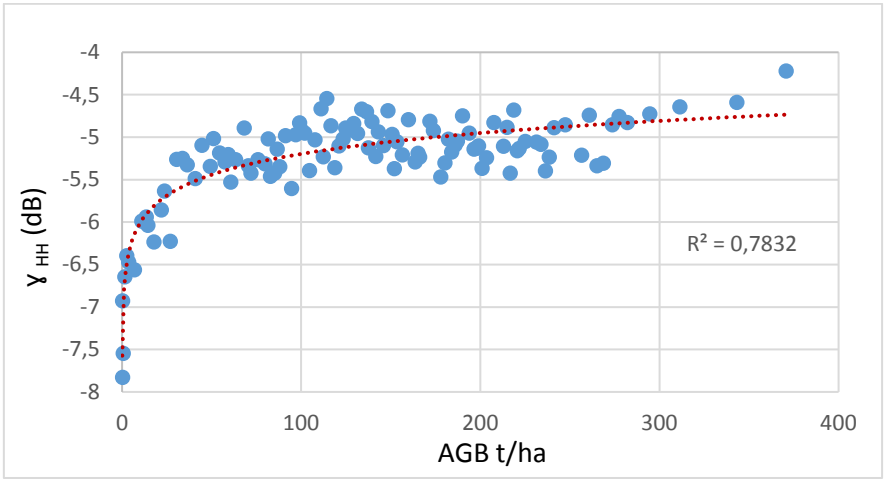


Figure 3-3a. γ^0 for HV and HH polarization in 2010 plotted as a function of AGB*, using* AGB *estimated from the 392 plots grouped into 111 homogenous classes.*

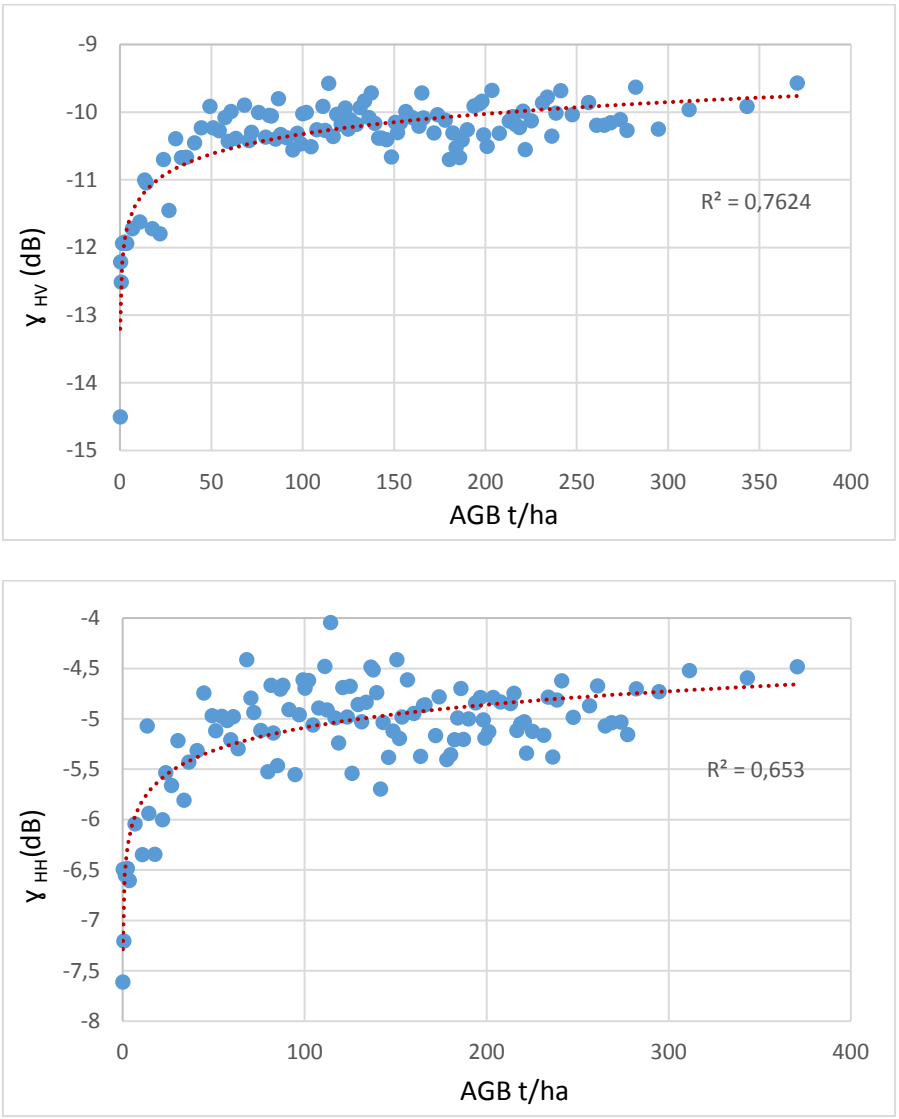


Figure 3-3b. γ^0 for HV and HH polarization in 2009 plotted as a function of AGB*, using* AGB *estimated from the 392 plots grouped into 111 homogenous classes.*

The biomass estimation algorithms for the 2005 and 2010 maps were prepared separately for coniferous forest (dominant in Poland) and broadleaf forest. The final AGB maps were derived by combining both maps using the Copernicus Land Monitoring – Forest type layer (available at 20 m spatial resolution).

3.1.3.3 *Training methods*

The method to be used for regional biomass estimation in Poland will be based on Random Forest (RF: Breiman, 2001). This is a machine-learning algorithm approach that uses multiple self-learning decision trees to parameterize models and uses them to estimate categorical or continuous variables. RF is an ensemble learning technique, where many decision trees are constructed based on random sub-sampling of the given data set. In addition, each node of every tree is split based on another random subset of parameters. A regression model is fitted to the target variable using each of the independent variables. Then for each independent variable, the data is split at several split points. The regression result is aggregated by taking the average of the predictions from all trees.

RF regression relating SAR data to the corresponding field calculated biomass is conducted over the entire country. The RF model is calibrated using a set of training plots located over the entire forested area except mountains.

The parameterization is performed on 200 single trees in the forest. This number of trees provided the lowest mean square error using an out-of-bag set of observations (compare Figure 3-4).

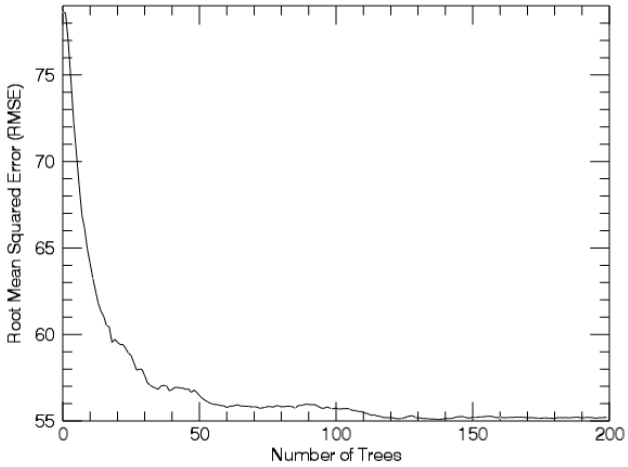


Figure 3-4. Example of the learning curve (out-of-bag-error).

The flow chart for the production of the AGB map for 2010 is presented in Figure 3-5.

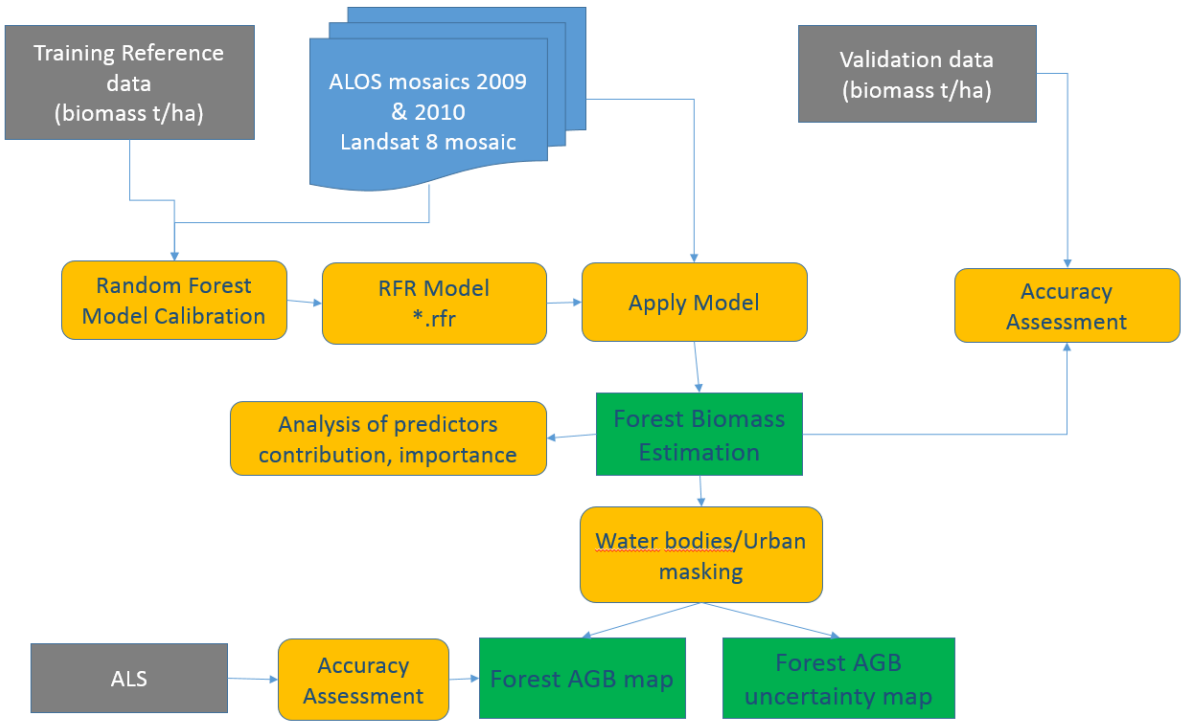


Figure 3-5. Flow chart of the production of the forest biomass map for 2010.

The Copernicus High Resolution Layers (imperviousness and water bodies) available at 20 m spatial resolution will be used to mask out water and urban areas. The SRTM data will be used to prepare a mask for slopes >20 degree.

Strengths, limitations and weaknesses of the method

- The two parameters: number of trees in the forest and the number of variables at each node, are important and should be tuned with the data.
- Resistant to overtraining (overfitting) - over-fitting is less of an issue compared to an individual decision tree and there is no need to prune the trees, which is a cumbersome task (Mishra et al. 2014).
- Saturation in high biomass is observed in the broadleaved forest, which has lower backscatter than coniferous; perhaps applying two separate models, one for coniferous and one for broadleaved forest, may partially overcome the problem with saturation (this needs further investigation).

3.1.3.4 *Methods to assign accuracies in the uncertainty map*

The overall accuracy of the reference biomass data is composed of the accuracy of the forest inventory (measurements), biomass calculation, sampling and prediction, as in the following equation (a modification of that proposed by Saatchi et al., 2011):

$$\varepsilon_{AGB} = (\varepsilon^2_{measurement} + \varepsilon^2_{BEF} + \varepsilon^2_{sampling} + \varepsilon^2_{prediction})^{1/2}, \quad \text{[Eq. 3-1.1]}$$

Here

1. $\varepsilon_{measurement}$: error associated with field data collection, i.e. height, DBH, density and calculation of stem volume at plot level, called a measurement error is assigned to 10% (based on the un published data provided by the team responsible for the National Inventory of Forest Condition).
2. ε_{BEF}: error in calculating biomass using the biomass expansion factors (BEFs) and wood density (WD).The BEF error is assigned to 20% (Forest Research Institute - unpublished data).
1. $\varepsilon_{sampling}$: this error originates from the variability of AGB within the pixel area and depends on the size of the plots used to upscale the AGB measurements to the pixel level. According to Saatchi et al. (2011) and Chave et al. (2003), a sampling error of 20% is assumed.
1. $\varepsilon_{prediction}$: The prediction error includes both the sampling error associated with the representativeness of the training data of the actual spatial distribution of AGB and the model predictions.

The uncertainty of the AGB prediction ($\varepsilon_{prediction}$) is calculated from the RMS error ($\sigma_{\widehat{AGB}}$) calculated for the AGB ranges (Section 5.4, Table 5-1). The following equation is used (Saatchi et al., 2011):

$$\varepsilon_{prediction} = \sigma_{\widehat{AGB}}/\widehat{AGB}\, x100, \quad \text{[Eq. 3-1.2]}$$

The uncertainty layer is provided at the pixel level for each AGB epoch.

3.1.3.5 *Methods to test the accuracy of the measurements*

The accuracy assessment will be done using the ALS data available for 6 test sites scanned in 2007 and 2015. In addition, the independent information on GSV for forest stand obtained from the National Forest Inventory Database will be used to assess the accuracy at stand level.

3.1.4 Products

The final products will be a forest biomass map for 2010 and a map of uncertainties.

3.1.5 Modifications for the 2005 and 2015 epochs

Method

Random Forest Regression is used to estimate biomass for all epochs. The biomass estimation algorithms for the 2005 and 2010 maps were prepared separately for coniferous forest (dominant in Poland) and broadleaf forest. The final AGB maps were derived by combining forest type estimation results using the Copernicus Land Monitoring Forest type layer (available at 20 m spatial resolution). For the 2015 epoch the biomass algorithms were developed without separating forest types.

Data

2005:

Input data

ALOS mosaic for 2007 and Landsat-5 mosaic for 2005-2006.

Reference data

Forest inventory data for 2005-2006.

2015:

Input data

The 2015 AGB map was derived using a time series of Sentinel-1 Dual polarisation IW mode data for the period 2015-2016 pre-processed by the University of Jena team. The pre-processing chain consists of calibration, geocoding (using the SRTM DEM), radiometric normalisation and conversion to γ^0 GeoTiff. The pre-processing was done using the ESA SNAP toolbox. Multi-channel filtering for individual Sentinel-1 tracks was applied to reduce speckle. The multi-temporal sum and median for VV and VH polarisation separately for summer and winter images were then calculated.

A cloud-free mosaic of Sentinel-2 was developed using atmospherically corrected images from bands 2, 3, 4 and 8 available at 10 m spatial resolution.

Reference data

Forest inventory data for 2015-2016.

3.2 Sweden

For Sweden, two alternative methods for regional mapping of AGB for 2010 are proposed: i) *k*-Nearest Neighbour) (*k*NN) estimation based on optical satellite imagery and Swedish National Forest Inventory (NFI) field plot data; and ii) use of the Water Cloud Model (WCM) applied to ALOS PALSAR/ALOS-2 PALSAR-2 data. Selection of the better approach will be made using the newly produced Airborne Laser Scanning (ALS)-based national AGB map of Sweden and field plots from the NFI as reference data.

3.2.1 General description of the region

Sweden is situated almost completely in the boreal forest region, though the southernmost parts are within the hemi-boreal and nemoral regions. Its total land area is 40.8 Mha, of which 23.3 Mha is productive forest land, 5.2 Mha is mountainous vegetation, 5.1 Mha is wetland and 2.9 Mha is farmland, according to the Swedish NFI (Fridman et al., 2014; Skogsdata, 2015). The forested land is dominated by Norway spruce (*Picea abies* (L.) H. Karst.), Scots pine (*Pinus sylvestris* L.) and birch (*Betula* spp.), where pine and spruce are about 80% of the growing stock. In the hemi-boreal and nemoral regions there are also forests dominated by beech (*Fagus sylvatica*) and oak (*Quercus robur*) present. The current total stem volume of growing stock is 3398 million m^3 and the total AGB is more than 2500 million ton. The average growing stock volume and AGB is 106 m^3/ha and 58.6 ton/ha, respectively. The highest plot values (circular plots with either 7 m or 10 m radius) are approximately 1400 m^3/ha for growing stock and 550 ton/ha for AGB.

3.2.2 Datasets

3.2.2.1 *Input data*

Spectral information from satellite images have been used to produce raster datasets. The 2010 dataset used in this project was derived from SPOT-4 HRVIR and SPOT-5 HRG data registered between 2008 and 2010 (approximately 80 percent of the images were registered in 2010). All images were geometrically precision-corrected to the Swedish National Grid, and the pixel size for all bands was resampled to 25 meters using cubic convolution. The *k*NN biomass map is a mapping of forest land only, and existing map data about land use is utilized to discriminate forest land from other land uses.

The ALOS PALSAR dataset used to generate the nation-wide stem volume data product consisted of strips of Fine Beam Dual (FBD) images (HH and HV polarization). All images were acquired during 2010. A gap-filled version using images acquired in the moderate resolution Wide Beam mode (pixel size of 75 × 75 m^2) is also available but not considered in this project because it does not fulfil the requirements on spatial resolution set out in the Statement of Work (< 50 m). The FBD coverage was almost complete, with a 1% gap in the north (Figure 3-6). On average 8 observations were available, with a few more observations (up to 24) locally in the southernmost part. The WB coverage was complete, but the number of observations was lower (on average 4 observations).

The ALOS PALSAR dataset was generated by JAXA EORC and delivered via ftp. In total, 104 FBD image strips covering Sweden were obtained. FBD images were acquired between May and October 2010. The images were provided as multi-looked intensity images in radar geometry (multi-look factors of 1 and 5 in range and azimuth). The ALOS PALSAR images covered a swath of 70 km with a pixel size of approximately 25 m in azimuth and 18 m in slant range, corresponding to 28 m in ground range.

The ALOS PALSAR dataset is unique among all regional case studies because it consists of all images acquired during the summer and fall of 2010. This dataset of long image strips of ALOS PALSAR data was obtained as part of the involvement of SLU in the Science Team of the Kyoto and Carbon Initiative. This multi-temporal dataset is different from the mosaic data publicly available and used by the other

regional teams since it consists of multiple observations of the SAR backscatter. Kyoto and Carbon Initiative data strips are available to a selected number of users being part of the Science Team and are produced on demand. Another advantage of the multi-temporal ALOS PALSAR dataset available to SLU was the possibility to be in full control of the pre-processing, which was not possible for users of the ALOS PALSAR mosaic data. Occasional issues occurring with the mosaic data such as abrupt variation of backscatter along image edges (Shimada & Ohtaki, 2010) did not occur.

The benefit of multi-temporal ALOS PALSAR observations in retrieving stem volume was discussed in Santoro et al. (2015b) where single-image and multi-temporal retrieval where presented and discussed for two test sites in northern and southern Sweden. The rms error of the stem volume retrieved from multiple images was always smaller than the errors associated with the individual single-image retrievals.

A biomass map covering all of Sweden will also be derived using ALOS-2 PALSAR-2 data for the year 2015. So far all requested strip data have been acquired over Sweden and consist of 116 FBD image strips (October 2014 – June 2016), mostly from ascending orbits (Figure 3-6).

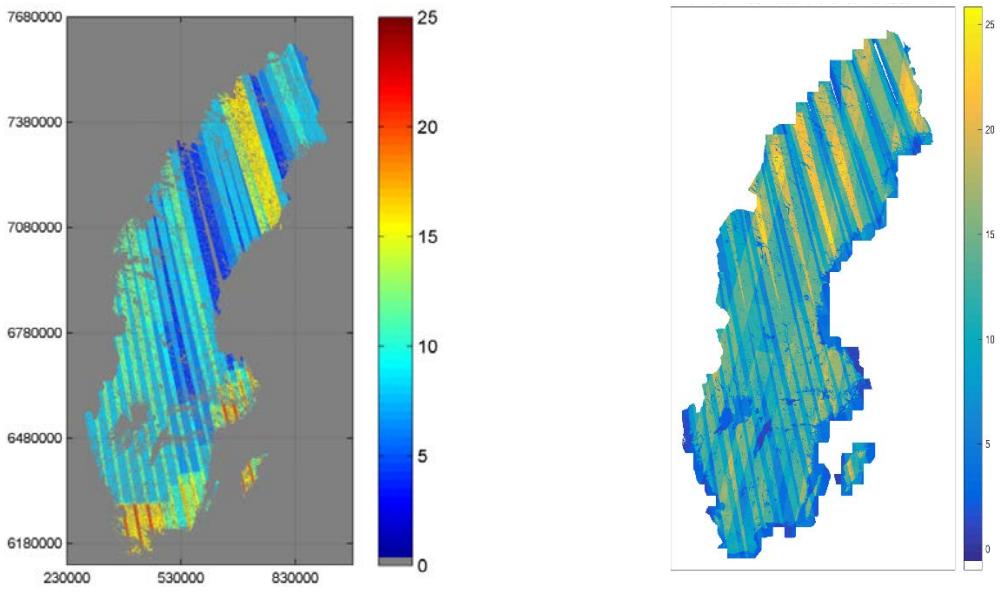


Figure 3-6. Number of ALOS PALSAR backscatter observations over Sweden in FBD mode acquired during 2010 (left) and October 2014 – June 2016 (right). Maps are in the SWEREF 99 projection (UTM 33).

Multi-temporal data in the same fashion as from the ALOS PALSAR/ALOS-2 PALSAR-2 sensor for the 2010 and 2015 epochs were also available from the Envisat ASAR sensor. Envisat ASAR operated over the Nordic European countries primarily in the moderate resolution modes Wide Swath and Global Monitoring. High-resolution modes, Image Mode (IM) and Alternating Polarization (AP), were more seldom acquired and in an inconsistent manner. For a display of the ASAR data coverage globally, we refer to Section 7. To adhere with the requirement on spatial resolution set for the regional mapping, the only suitable dataset would have consisted of images acquired in IM and AP modes. In Wegmüller et al. (2013), the temporally densest one-year dataset of ASAR IM (30 images) was investigated at a forest site in southern Sweden to check whether the hyper-temporal approach implemented in the BIOMASAR algorithm could be applied to estimate biomass. The retrieval performed poorly (relative rms error close to 50%) because of the rather small dataset and the lack of many images acquired during wintertime under frozen conditions guaranteeing the largest possible sensitivity of the SAR

backscatter to biomass. It is reminded that C-band based retrieval of biomass was reported to be reliable starting with about 60 scenes (Santoro et. al., 2011).

3.2.2.2 *Training data*

For the *k*NN estimation, data from the NFI in Sweden were used. The NFI has the task of describing the state and changes in Sweden's forests (Fridman et al., 2014). The information collected is used, for example, as a basis for forestry, energy and environmental policy in Sweden. The NFI carries out random sample inventories annually of the Swedish forests and comprises all types of land, but the most thorough inventory is carried out on productive forest land. It is carried out as two independent annual systematic field samples, consisting of either temporary or permanent plots (Figure 3-7), covering all of Sweden annually. The plots are located systematically along the perimeter of square sample clusters, which vary in size for different regions of Sweden (Ranneby et al., 1987; Fridman et al., 2014). In most regions, the clusters contain either 12 temporary plots (7 m radius) or 8 permanent plots (10 m radius). In total, 5000 permanent and 3700 temporary plots are surveyed every year by 15 field teams of surveyors. Approximately 60% of the plots are located on forest land, and on these plots data on both state and changes are collected. Each plot is surveyed by measuring each tree within a 10 or 7 m radius from the plot centre and recording a large amount of data about the vegetation and soil type. The position of each plot is determined by GNSS. The survey data about single trees are used in allometric models to estimate plot-level totals of many variables, such as mean tree height, mean tree diameter, total stem volume and total AGB. The *k*NN estimation is based on plot measured data from the NFI, complemented with satellite image data extracted for each plot based on the corresponding GNSS coordinate, as input to the *k*NN algorithm. In total, about 20,000 NFI plots located on forest land that were field surveyed between 2006 and 2010 have been used to produce each nation-wide raster dataset with estimated forest variables. All estimates were derived image by image and the NFI data used for the estimation were forecasted to the year of the image registration using growth models.

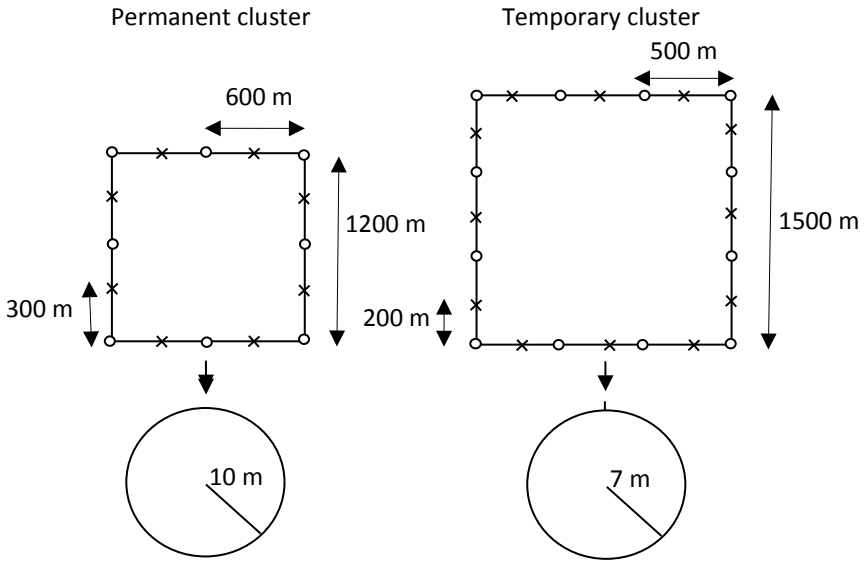


Figure 3-7. Illustration of the distribution of permanent and temporary clusters in the Swedish NFI.

For ALOS PALSAR/ALOS-2 PALSAR-2, an auxiliary dataset of canopy cover was used as training data. Here, we selected the 250 m global MODIS VCF dataset (Hansen et al., 2003) for the year 2010 to

identify "ground" and "dense forest" pixels. To cope with the different resolutions of the MODIS VCF and PALSAR/ PALSAR-2 datasets, the selection of the two types of pixels and therefore the estimation of the model parameters σ^0_{gr} and σ^0_{veg} was done at the scale of the MODIS VCF dataset. Each PALSAR image was multi-looked to 300 m, while the MODIS VCF dataset was resampled (nearest neighbour) to the same pixel size; 300 m represented a trade-off between keeping the spatial distribution of canopy cover and achieving a spatial distribution similar to the canopy cover at the spatial resolution of the PALSAR/PALSAR-2 data. In future, we foresee the possibility of using a Landsat-based dataset of VCF (Sexton et al., 2013), which, at the time of this work, was available only for the year 2000 and exhibited some artefacts of major importance for the performance of the retrieval algorithm.

3.2.2.3 *Internal validation data for national accuracy tests*

To validate the satellite-based models, nationwide NFI field samples in combination with ALS data have been merged to establish accurate wall-to-wall maps of entire Sweden for several common forest variables, including forest biomass.

This raster database for all forest land in Sweden was produced in 2014-2015 using existing ALS data and field data from the Swedish NFI. In total, 11,500 permanent plots from 2009 to 2013 located on productive forest land were used in the project. The NFI plot data were forecasted or hindcasted to the same year as they were laser scanned with the Heureka system developed at SLU (Wikström et al., 2011). Hence, these ALS-based forest maps represent the condition in the forests for the specific scanning year, estimating Sweden's forests from 2009 to 2016.

Since the laser scanning started in 2009, approximately 97% of Sweden's productive forest land has been scanned. The scanning is organised in 387 blocks, usually 25 by 50 km in size. For each block the 350 permanent NFI plots closest to the block centre were chosen. These plots were selected from the block to be estimated as well as nearby blocks scanned with the same scanner type (Leica, Optech or Riegl) and with the same leaf condition (leaf on or leaf off). Our evaluations show that ALS estimates for deciduous forest generally have a lower accuracy than estimates for coniferous forest. Not having leaves on the trees during the ALS campaign does not affect the estimation accuracy much. It is actually more problematic to have leaf-on conditions due to difficulties in defining suitable density measures in the ALS data. Thus, high proportions of deciduous trees (>30%) during leaf-on conditions might lead to biased estimates (over-estimations).

Plots that had been disturbed between scanning and field measurement were removed, using information registered during the field inventory and a statistical outlier elimination procedure. In total, about 15% of the plots were removed.

Linear regression models were used to relate the selected forest variables, or transformations of the variables, to metrics derived from the ALS data. The regression models were then used to predict forest data for all grid-cells within the block. A small set of basic models was used for each variable (see Table 3-2), from which the best model was chosen automatically in each estimated block. The best model was defined as the one with the highest coefficient of determination, R^2. The same small set of alternative models was used to estimate the forest variables (for example stem volume) for all laser-scanned blocks across Sweden. However, the regression parameters were re-estimated for each block using field data from the 350 (approximately) closest permanent NFI plots.

Table 3-2. Model definitions where ElevP80, ElevP90 and ElevP95 are the 80th, 90th and 95th percentiles of the height distribution, ElevStddev is the standard deviation of the height, and H80veg and H90veg are the products of the proportion of 1st returns from higher than 1.5 m above ground and the 80th and

90th percentiles. The associated volume estimations are about 16.9% to 21.8% RMSE at stand level for a few different independent test datasets around Sweden.

Modelled variable	**Explanatory variable 1**	**Explanatory variable 2**	**Explanatory variable 3**
Lorey's height	*ElevP95*		
Basal area-weighted diameter	*ElevP80*	*H80veg*	
Basal area-weighted diameter	*ElevP90*	*H90veg*	
Basal area-weighted diameter	*ElevP80*	*H90veg*	
Basal area-weighted diameter	*ElevP90*	*H80veg*	
Volume$^{0.5}$	*ElevP80*	*H80veg*	*ElevStddev*
Volume$^{0.5}$	*ElevP90*	*H80veg*	*ElevStddev*
Volume$^{0.5}$	*ElevP95*	*H80veg*	*ElevStddev*
Volume$^{0.5}$	*ElevP80*	*H90veg*	*ElevStddev*
Volume$^{0.5}$	*ElevP90*	*H90veg*	*ElevStddev*
Volume$^{0.5}$	*ElevP95*	*H90veg*	*ElevStddev*
Basal area	*H80veg*	*ElevStddev*	
Basal area	*H90veg*	*ElevStddev*	
AGB$^{0.5}$	*ElevP80*	*H80veg*	
AGB$^{0.5}$	*ElevP90*	*H90veg*	
AGB$^{0.5}$	*ElevP80*	*H90veg*	
AGB$^{0.5}$	*ElevP90*	*H80veg*	

3.2.2.4 *Data used for accuracy assessment*

The *k*NN and WCM maps of AGB will be evaluated using the ALS-based maps of AGB and field plots from the Swedish NFI for the epochs 2005, 2010 and 2015.

3.2.3 Methods

3.2.3.1 *Preprocessing of data*

For *k*NN, field plots are updated to correspond to image year by adding the estimated production increment over the time difference. Digital maps are used to define and separate forest land from other land classes and only plots found on forest land are used for classification. Images are preprocessed in several steps. Geometrically correct images are delivered from the Swedish National Land Survey (Lantmäteriet), where they have been corrected to the Swedish national map grid to a nominal accuracy of less than 0.5 pixels. Illumination corrections of the images are then conducted to diminish the topographic effects that otherwise can be prominent in parts of Sweden (Reese et al., 2003; Tomppo et al., 2008).

For ALOS PALSAR/ALOS-2 PALSAR-2, each strip was calibrated using factors provided by JAXA (-83 dB) and terrain geocoded using the national DEM of Sweden with 50 m posting (Lantmäteriet, 2010). To cope with inaccuracies in the orbital data, which could cause a shift in the geocoded geometry of 1-2 pixels, a geocoding refinement procedure was used (Wegmüller, 1999) consisting of a match between the SAR image to be geocoded and an image considered as reference for the output geometry. Here, a simulated SAR image from the DEM was used. In flat terrain, where no match could be found, a mosaic of Landsat images downloaded from the Global Land Cover Facility (GLCF, http://www.landcover.org) was used. The final geocoding accuracy was below 1/3rd of the pixel size. The FBD and WB images were geocoded to 25 m 75 m pixel size respectively. The SAR backscatter was then normalized for slope-induced effects using the approach described in Santoro et al. (2009). Topography-dependent scattering effects in forests due to the local incidence angle were not accounted for. Each image strip was tiled using a pre-defined 30 km grid. The size of the tile was set so that the computing resources could be optimized when estimating stem volume. No additional speckle filtering was applied. The images were provided in multi-looked format, and visual analysis of the geocoded images revealed clear fine-scale textural features that would have been lost with additional filtering. However, the use of the multi-channel filter (Quegan and Yu, 2001) was tested. The spatial distribution of the backscatter changed only minimally because of the strong correlation between ALOS PALSAR/ALOS-2 PALSAR-2 data backscatter observations acquired under unfrozen conditions (Santoro et al., 2009).

3.2.3.2 *Biomass estimation algorithms*

As a complement to ordinary statistics from the NFI, nation-wide full coverage maps with estimated forest variables such as growing stock volume, volume per tree species, tree height and stand age can be created (Tomppo et al., 2008; Fazakas et al., 1999; McRoberts et al., 2010). In Sweden such maps have been produced by combining remote sensing data and field data from the Swedish NFI using the *k*NN method. *k*NN is a non-parametric classification technique that is widely used for assessment of forest attribute maps (McRoberts et al., 2010). The Swedish production is semi-automated, using a set of predefined steps for preparation of field plots and images. Using a combination of NFI plots and spectral information from SPOT-4/5 images, raster maps of different forest parameters with 25 m pixel size have been produced for 2000, 2005 and 2010 (Nilsson and Olsson, 2010). Maps are planned to be produced approximately every fifth year and will be available free of charge via the Internet. They cover approximately 95% of all forest land, forested wetlands and mountainous forests in Sweden.

Some small areas are expected to be missing due to lack of suitable satellite imagery, cloud cover being the usual culprit.

Forest variables are estimated as weighted averages of observed variable values for the *k* most similar plots in a feature space defined by the spectral bands in, for example, SPOT-4/5 imagery. Similarity is defined by the feature space distance *d* from pixel *p* to field plot *i*, and the weight $w_{i,p}$ for field plot *i* when estimating forest variables for pixel *p* is defined as:

$$w_{i,p} = \frac{1}{d_{i,p}^{t}} \div \sum_{j=1}^{k} \frac{1}{d_{j,p}^{t}} \qquad \text{(Eq. 3-2)}$$

for the *k* plots with the shortest feature space distance to pixel *p*, otherwise $w_{i,p} = 0$. The estimate of a forest variable ($\hat{y}$) for pixel *p* is defined as:

$$\hat{y}_p = \sum_{i=1}^{k} w_{i,p} y_{i,p} \qquad \text{(Eq. 3-3)}$$

Together these steps create a chain of functions that classifies image by image and, once combined, result in a mosaic, or map, of forest parameters (Figure 3-8).

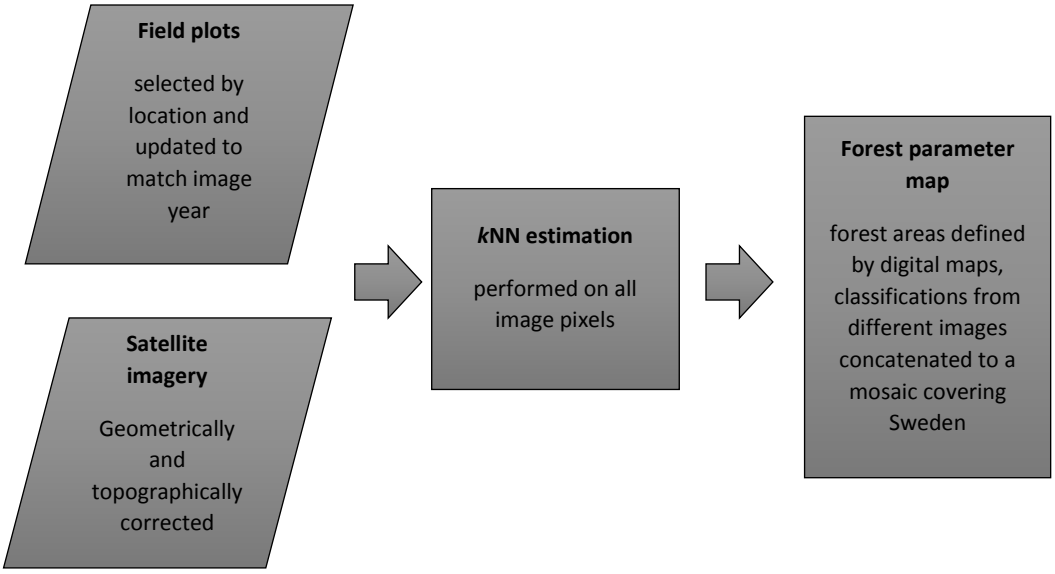


Figure 3-8. Illustration of the production chain used for the kNN based forest products.

AGB estimation from ALOS PALSAR/ALOS-2 PALSAR-2 data uses a model-based approach exploiting a WCM with gaps (Askne et al., 1997; Fransson and Israelsson, 1999; Askne et al., 2003; Santoro et al., 2006; Santoro et al., 2015b) to retrieve forest stem volume from each date in a multi-temporal sequence of dual-polarized backscatter observations. The estimates of stem volume obtained from each SAR observation by inverting the trained model are then combined using a linear weighting scheme (Cartus et al., 2012; Santoro et al., 2006; Santoro et al., 2015b). The performance of this approach to retrieving forest variables from L-band backscatter was compared with other parametric and non-parametric approaches in Tanase et al. (2014), and showed no significant shortcomings, although the comparison was not undertaken in boreal forest. The approach is based on the

BIOMASAR algorithm, which is also applied for the global biomass mapping. Nonetheless, the implementation of the estimation of the model parameters is somewhat different due to different spatial resolutions, EO datasets and auxiliary datasets available for Sweden and at global scale.

The WCM with gaps assumes that the forest backscatter consists of a component coming from the canopy and a component from the ground that reaches the sensor either through gaps in the canopy or, attenuated, through the canopy. Double-bounce and multiple interactions are not considered because in managed boreal forest these terms were found in previous studies to be negligible with respect to direct scattering (see Santoro et al., 2006, and references therein). Polarimetric decomposition of the fully polarimetric data over two test sites in south and north Sweden was further used to support our assumption (unpublished work). The total forest backscatter could be explained as a contribution of a surface and a volume component, and the double-bounce component was negligible. As a result of literature reports and our local analysis, we decided to neglect a double-bounce term in the forest backscatter model because of no evidence that such term is necessary. This, however, does not imply that double bounce should always be neglected as specific forest structures or growth stages in combination with soil conditions may be characterized by a considerable multiple scattering component. For this to be taken into account, one would need a detailed description of forest and soil conditions. As a result of our choice to neglect some scattering types in the inversion model, we may propagate errors to the estimates of biomass.
The original forest backscatter model expresses the total forest backscatter, σ^0_{for}, as a function of two forest variables, forest height, h, and a measure of canopy closure from a microwave perspective referred to as the area-fill factor, η.

$$\sigma^o_{for} = \left[(1-\eta)\sigma^o_{gr} + \eta\sigma^o_{gr}e^{-\alpha h}\right] + \sigma^o_{veg}\left(1-e^{-\alpha h}\right) \qquad \text{(Eq. 3-4)}$$

The total forest backscatter is modelled in terms of a ground component and a vegetation component, where σ^0_{gr} and σ^0_{veg} express the backscattering coefficient of the ground and the vegetation, respectively. The first term of (Eq. 3-4), corresponding to the ground backscatter, takes into account that the scattered waves return to the radar either through canopy gaps or attenuated by the foliage. The two-way tree transmissivity is expressed in (Eq.3-4) by an exponential function including a coefficient for the two-way attenuation per meter, α, and the tree height assumed to correspond to the depth of the vegetation.

The inversion of such a model is cumbersome because it requires knowledge of one of the two forest variables. Inverting for height would require knowledge of the area-fill factor, i.e. a measure of the canopy closure seen from the perspective of the radar, which is hardly quantifiable. While it can be argued that the canopy closure seen by short wavelength radar is close to the optical canopy closure, there is a difference because the microwaves require larger gaps than the visible wavelengths to penetrate. Following the explanation of the area-fill factor, an inversion for area-fill is not of interest to foresters.

The expression for the total forest backscatter can be simplified by assuming a fairly simple relationship between area-fill factor, tree attenuation and forest transmissivity. This expression has been derived by comparing the WCM of Equation (3-4) and a WCM expressed as a function of stem volume as in Pulliainen et al. (1994). A model for area-fill has been introduced in Santoro et al. (2002) as:

$$\eta = \frac{1 - e^{-\beta V}}{1 - e^{-\alpha h}} \qquad \text{(Eq. 3-5)}$$

In (Eq. 3-5), the empirical coefficient β is introduced to model the two-way forest transmissivity, $e^{-\beta V}$, where V represents the stem volume (Pulliainen et al., 1999; Santoro et al., 2002). The coefficient β depends on forest structure and the dielectric properties of the canopy. At L-band, β was estimated usually to lie between 0.003 and 0.007 in boreal forest (Santoro et al., 2006; Santoro et al., 2015b).

By replacing in (Eq. 3-4) the area-fill factor from (Eq. 3-5) and assuming an allometric function between height and stem volume (see e.g., Santoro et al., 2002; Askne and Santoro, 2012), the WCM can be expressed as a function of stem volume only.

$$\sigma^o_{for} = \sigma^o_{gr} e^{-\beta V} + \sigma^o_{veg}\left(1 - e^{-\beta V}\right) \qquad \text{(Eq. 3-6)}$$

The three model parameters in (Eq. 3-6), σ^0_{gr}, σ^0_{veg} and β, are unknown a priori and need to be estimated to invert the model to retrieve stem volume. Model training will be discussed in the next Section.

Given a measurement of the forest backscatter, σ^0_{meas}, and the corresponding estimates of the three model parameters, the inversion of the model in (Eq. 3-6) is straightforward and allows the stem volume, $\hat{V}$, to be estimated:

$$\hat{V} = -\frac{1}{\beta}\ln\left(\frac{\sigma^o_{meas} - \sigma^o_{veg}}{\sigma^o_{gr} - \sigma^o_{veg}}\right) \qquad \text{(Eq. 3-7)}$$

Assuming that N measurements of the SAR backscatter are available for the same unit (i.e., pixel, plot or polygon), individual estimates of stem volume can be combined as a weighted linear combination to obtain a new estimate referred to as multi-temporal stem volume, V_{mt} ((Eq. 3-8)), with better accuracy than each of the individual estimates:

$$V_{mt} = \frac{\sum_{i=1}^{N} w_i \hat{V}_i}{\sum_{i=1}^{N} w_i} \qquad \text{(Eq. 3-8)}$$

To maximize the contribution of estimates derived from images with strong backscatter contrast between unvegetated and dense mature forest, the weights w_i are defined as the vegetation-to-ground backscatter difference in dB for the ith image, σ^0_{veg} - σ^0_{gr}, normalized by the maximum of the N weights:

$$w_i = \frac{\sigma^0_{veg,i} - \sigma^0_{gr,i}}{\max\left(\sigma^0_{veg,i} - \sigma^0_{gr,i}\right)} \qquad \text{(Eq. 3-9)}$$

The plausibility of this definition was demonstrated in boreal forests (Santoro et al., 2011).

3.2.3.3 *Training methods*

The *k*NN dataset is produced using the non-parametric *k*NN estimation (training) algorithm applied on NFI field plots. Hence, no separate training is needed (see Section 3.2.3.2).

The estimation of σ^0_{gr} and σ^0_{veg} is typically undertaken by fitting the model in (Eq. 3-6) to pairs of measurements of the backscatter and *in situ* stem volume forming a training dataset (see e.g., Santoro et al., 2015b). To capture the spatial variability of the backscatter, ideally several local test sites would be needed to obtain local representations of the modelled backscatter. Using one representation of the model based on single test site for a large area would introduce a retrieval error due to imperfect characterization of the backscatter away from the test site in the case of spatial variability of the environmental conditions. Given the lack of a dense network of samples everywhere in Sweden suitable for training, we consider the approach implemented in the BIOMASAR algorithm, which derives the two model parameters of the backscatter without the need for training data and assumes a certain value for the transmissivity.

BIOMASAR has been developed and validated for C- and L-band multitemporal backscatter observations and the same type of model as in (Eq. 3-6) (Santoro et al., 2011; Cartus et al., 2012). The parameters σ^0_{gr} and σ^0_{veg} are assumed to be equal to an average backscatter for unvegetated surfaces (but belonging to a vegetation land-class) and a very dense canopy, respectively. The coefficient β is set constant and equal to 0.004 ha/m^3 as a first approximation, given the lack of direct measurements of attenuation. While it is understood that such an assumption may introduce errors in the retrieval, experimental results at two test sites in Sweden revealed that the estimation error introduced by assuming β = 0.004 ha/m^3 is negligible when compared to other error sources (e.g., weak sensitivity of the backscatter to stem volume in dense forest). It is here worth noticing that the approach used in the regional mapping of biomass of South Africa derives the estimate of a coefficient of forest transmissivity from a set of *in situ* observations. While appealing, the method has not been tested in boreal forest yet and the level of information required by this approach to perform is unclear. This aspect, however, will be formally investigated in the second year of this project.

The assumption of a constant value for β is currently being challenged in the context of the algorithmic development for the global biomass mapping part. The regional mapping of Sweden and the global mapping share the same forest backscatter model but slightly different model training approaches to cope with the different scopes the models are used (regional mapping in managed boreal forest vs. global mapping). The algorithm to retrieve biomass from L-band data (ALOS PALSAR for 2010 and ALOS-2 PALSAR-2 for 2015) will be updated if the algorithmic development of the global biomass mapping part leads to a significantly different characterization of the β coefficient.
In the remainder of this Section, we outline the estimation of σ^0_{gr} and σ^0_{veg} and the procedure to retrieve stem volume as currently implemented. Improvements are discussed at the end. It is herewith assumed that we have a stack of multi-temporal SAR backscatter images, co-registered to a common reference geometry.

Estimation of σ^0_{gr}

The estimate of σ^0_{gr} is equal to the average backscatter for pixels labelled as "unvegetated" within a window of finite size. The unvegetated pixels are selected based on an estimate of canopy cover fraction from an auxiliary dataset and are defined as a canopy cover fraction below a given threshold. A reliable dataset of canopy cover fraction contemporaneous with the PALSAR data acquired in 2010

is the 250 m global MODIS Vegetation Continuous Fields dataset (Hansen et al., 2003) for the year 2010. To cope with the different spatial resolutions of the MODIS VCF dataset and the PALSAR dataset (25 m), the selection of unvegetated pixels is done at 300 m, so the backscatter is extracted from a version of the PALSAR image multi-looked from 25 m to 300 m (and a VCF resampled from 250 m to 300 m). The reason for using 300 m is provided in Fransson and Santoro (2014).

As a trade-off between precision of the model parameter estimates and computational burden, the estimation window is set to 30 × 30 km^2. This choice was supported by previous evidence that the L-band backscatter has a strong temporal consistency (Santoro et al., 2009) and does not present marked spatial variability at kilometric scale. The threshold for labelling pixels as unvegetated is adaptive, starting from 15% and reaching at most 30%. The threshold is increased when the fraction of pixels labelled as unvegetated within the estimation window is less than 0.3%. In this way, we could achieve estimates of σ^0_{gr} without artefacts for almost all 30 × 30 km^2 subsets into which the SAR images available over Sweden were divided. A higher VCF threshold or a smaller fraction of unvegetated pixels within the estimation window were often characterized by unrealistic estimates of σ^0_{gr} which were then discarded.

With this approach, a grid of σ^0_{gr} estimates spaced by 30 km in easting and northing is obtained for each SAR image. A raster image matching the geometry of the SAR backscatter image is then derived by bilinear interpolation of the sample estimates.

Estimation of σ^0_{veg}

To obtain an estimate of σ^0_{veg}, the mean backscatter for pixels labelled as "dense forest" within a window of finite size (taken to be 30 × 30 km^2) is first computed. Dense forest pixels correspond to pixels in the dataset of canopy cover fraction with a canopy cover above 85% of the maximum value within the estimation window. An estimate is derived when the proportion of dense forest pixels within the window is above 0.5%. As in the case of σ^0_{gr}, a grid of the backscatter for dense forest, referred to as σ^0_{df}, spaced by 30 km in easting and northing is obtained for each SAR image. The estimate of σ^0_{veg} is then obtained from σ^0_{df} by compensating for the backscatter component originating from the ground. The compensation is achieved by inverting (Eq. 3-4) in order to express σ^0_{veg} as a function of the remaining parameters in the model:

$$\sigma^o_{veg} = \frac{\sigma^o_{df} - \sigma^o_{gr} e^{-\beta V_{df}}}{1 - e^{-\beta V_{df}}} \qquad (3\text{-}10)$$

For the estimation of σ^0_{veg}, a representative value of stem volume for dense forest is required, V_{df}. This is defined as the 90th percentile of the stem volume distribution within the estimation window (i.e., over 30 × 30 km^2). The estimation of V_{df} implies a priori knowledge on the spatial distribution of stem volume, which can be obtained for example from inventory data or a raster dataset of spatially explicit values of stem volume (see e.g., Santoro et al., 2015a). For Sweden, V_{df} is estimated from the *k*NN Sweden 2010 dataset, thus implying that an incorrect representation of the spatial distribution of stem volume in this dataset translates to a retrieval error using the SAR data. An assessment of modelling and retrieval at the Remningstorp and Krycklan test sites revealed underestimation of stem volume in the densest forests when utilizing this retrieval approach (Fransson and Santoro, 2014). A possibility would be to complement the kNN dataset with data from the Swedish NFI. This will be investigated during the second year of this project. By comparing retrieval results from 2010 and 2015, it will be

possible to spot errors due to an incorrect modelling (including the parameterization of V_{df}) and, therefore, gain confidence in the estimation of this parameter.
Taking into account that the estimation of the model parameters is done at reduced spatial resolution and stem volume is scale-dependent, it is necessary to use a value of V_{df} at the same spatial resolution as the SAR and canopy cover datasets. For this reason the *k*NN Sweden 2010 dataset was averaged from 25 m to 300 m and an estimate of V_{df} was derived at the same pixel size simply to aid estimation of σ^0_{veg}.

From the 30 × 30 km^2 grid of point-wise σ^0_{veg} estimates for a given image, a raster matching the geometry of the SAR backscatter image is then obtained by bilinear interpolation.

Retrieval of stem volume
The retrieval of stem volume in (Eq. 3-6) is constrained to generate estimates between 0 and a maximum stem volume, which is set empirically to V_{df} + 50 m^3/ha. Here, the value of V_{df} for the original spatial resolution of the SAR data is considered, i.e., 25 m. The stem volume estimate for backscatter measurements below the smallest modelled backscatter is set to 0 m^3/ha. Conversely, if a backscatter measurement is up to 2 dB greater than the maximum modelled backscatter, the corresponding estimate is equal to the maximum retrievable value. For backscatter measurements more than 2 dB above the maximum modelled backscatter, the estimated stem volume is set to not-a-number.

The multi-temporal combination in (Eqs. 3-8, 3-9) is applied to all estimates of stem volume corresponding to a backscatter difference ($\sigma^0_{veg} - \sigma^0_{gr}$) larger than 0.5 dB. Images with weak backscatter contrast between dense forest and unvegetated areas may be detrimental to the final estimate. The impact of such an assumption (as yet unverified at L-band) is negligible since the forest backscatter difference at L-band is often greater than 1 dB in Swedish forest (Santoro et al., 2006; Santoro et al., 2009; Santoro et al., 2015b).

3.2.3.4 *Methods to assign accuracies in the uncertainty map*
The common uncertainty approach available for all three years originates in the estimated biomass for the epoch maps, which were classified into 30 tons/ha classes, 0-210 tons/ha. Then each estimated pixel was assigned the uncertainty (RMSE) derived class-wise, when the epoch biomass maps were evaluated against national forest inventory plot data.

In addition, a further developed uncertainty approach is available for the 2015 product, since the 2015 product is based solely on the BIOMASAR algorithm. The quantification of uncertainties for the ALOS PALSAR-2 based estimates of stem volume is adapted from the procedure applied to quantify uncertainties of growing stock volume (GSV) estimates derived from hyper-temporal ASAR data and the BIOMASAR algorithm. It is therefore referred to Section 7.2.10.1 and Santoro et al. (2015a). The procedure is in line with the guidelines developed by the University of Leicester to provide a common framework for determining uncertainties, deviating where other data or other ways of quantifying individual uncertainties are used. For completeness, a brief overview is given; the quantification of the individual uncertainty components below will be addressed during the second year of the project.

The precision of a stem volume estimate obtained from an observation of the PALSAR backscatter is determined by propagation of error of (i) the measured SAR backscatter, σ^0_{meas}, and (ii) the estimates of the forest backscatter model parameters σ^0_{gr}, σ^0_{df}, β and V_{df} . Assuming that the errors in the five variables are uncorrelated and overall small, the variance of the GSV estimate obtained from a single

SAR backscatter observation, δ_V, corresponds to the sum of the individual variances (uncertainty of a SAR backscatter measurement, $\delta_{\sigma^0_{meas}}$, uncertainty of the model parameters, $\delta_{\sigma^0_{gr}}$, $\delta_{\sigma^0_{df}}$ and δ_{β}, uncertainty of the GSV of dense forest, $\delta_{V_{df}}$) weighted by the related partial derivatives (Santoro et al., 2015a):

$$\delta_V = \sqrt{\delta^2{}_{\sigma^0_{meas}}\left(\frac{\partial V}{\partial \sigma^0_{meas}}\right)^2_{\sigma^0_{gr},\sigma^0_{df},\beta,V_{df}} + \delta^2{}_{\sigma^0_{gr}}\left(\frac{\partial V}{\partial \sigma^0_{gr}}\right)^2_{\sigma^0_{meas},\sigma^0_{df},\beta,V_{df}} + \delta^2{}_{\sigma^0_{df}}\left(\frac{\partial V}{\partial \sigma^0_{df}}\right)^2_{\sigma^0_{meas},\sigma^0_{gr},\beta,V_{df}} + \delta^2{}_{\beta}\left(\frac{\partial V}{\partial \beta}\right)^2_{\sigma^0_{meas},\sigma^0_{gr},\sigma^0_{df},V_{df}} + \delta^2{}_{V_{df}}\left(\frac{\partial V}{\partial V_{df}}\right)^2_{\sigma^0_{meas},\sigma^0_{gr},\sigma^0_{df},\beta}} \quad \text{(Eq. 3-11)}$$

The uncertainty of the multi-temporal stem volume estimate, $\delta_{V_{mt}}$, is then modelled as a linear combination of the single-image stem volume uncertainties, $\delta_{\hat{V}_i}$, assuming that the original weights, w_i, are the best estimate of the individual variances of $\hat{V}_i$ across the time series of observations, Equation 3-8. Nonetheless, the covariance between different observations needs to be taken into account since L-band observations are highly correlated over vegetated terrain. This aspect has not yet been formalized for this regional case but we expect to implement the procedure being developed for the global case using the BIOMASAR-L algorithm (Section 7.2.10.2), this algorithm being very similar to what has been used here to derive the estimates of GSV for Sweden from the ALOS PALSAR dataset.

3.2.3.5 *Methods to test the accuracy of the measurements*

The accuracy assessment for both the *k*NN and WCM AGB maps will be carried out on a pixel level and also aggregated to larger area units using the ALS-based maps of AGB for the epoch 2010.

3.2.3.6 *Methods of merging*

The best combination product is created by applying a multiple linear regression where national field inventory plots are to be explained with the kNN and WCM map-based pixels. From this combination, the formula is applied to the entire rasters in order to create the merged product.

3.2.4 Products

The following AGB maps will be included and evaluated using the ALS-based maps for the epoch 2010:

- AGB map based on *k*NN estimation and NFI field plot data,
- AGB map based on WCM based on ALOS PALSAR data.
- Merged map with the best combination of the kNN map and the WCM based map.

The maps are produced image by image and the combined to provide mosaics, or maps, of AGB with a pixel size of 25 m by 25 m for the entire country.

It is noted that retrieval based on ALOS PALSAR mosaic data (i.e., a single backscatter observation) is known to perform worse than multi-temporal retrieval and is, therefore, not addressed in this study.

3.2.5 Modifications for the 2005 and 2015 epochs

Methods

*k*NN estimation applied to optical SPOT-4/5 images for the 2005 product. Inversion of WCM for nationwide ENVISAT ASAR data (150 m pixels) for the 2005 product. Moreover, the combination of these maps is evaluated using linear regression, with the respective maps as explanatory variables. Inversion of WCM for nation-wide ALOS-2 PALSAR-2 data for the 2015 product. Additionally, *k*NN estimation applied to optical Sentinel-2 images for a subset of Sweden for 2015. Also inversion of WCM for this subset of Sweden, using Sentinel-1 data. The combination of the two methods will be evaluated for the subset, initially using linear regression, as with the 2005 and 2010 products. If we can derive an improved weighting of the maps from the separate data sources, we will use that instead.

Data

2005:

Input data
*k*NN based AGB map with 25 m pixel size (based on NFI+SPOT-4/5 imagery) and WCM based stem volume map (based on ENVISAT ASAR).

Reference Data
NFI plots from 2005/2006.

2015:

Input data
ALOS-2 data with 25 m pixel size for nationwide product. Sentinel-1 and Sentinel-2 data for subset of Sweden.

Reference Data
NFI plots and additional test site data at plot- and stand-level for the subset region Remningstorp. An ALS-based map of AGB/stem volume is available for the period 2009-2015, where large blocks of estimated data are available for each year during this period.

3.3 Kalimantan (Indonesia)

3.3.1 General description of the region

The regional area of Indonesia is represented by the Indonesian part of Borneo, which is named Kalimantan, and covers 73% of Borneo's land mass. The north of Borneo comprises the Malaysian states of Sarawak and Sabah and the small independent Sultanate of Brunei Darussalam. Borneo is the third largest island in the world and the largest landmass in the Sundaic area. The island lies in a region (between latitudes 7°N and 4°S) of frequent rainfall and high temperatures throughout the year, which are ideal conditions for plant growth (MacKinnon et al. 1996). The pattern of rainfall is linked to the "dry" southeast monsoon from May to October and the "wet" northwest monsoon from November to April. The area of Kalimantan covers approximately 540,000 km² and is depicted in Figure 3-9.

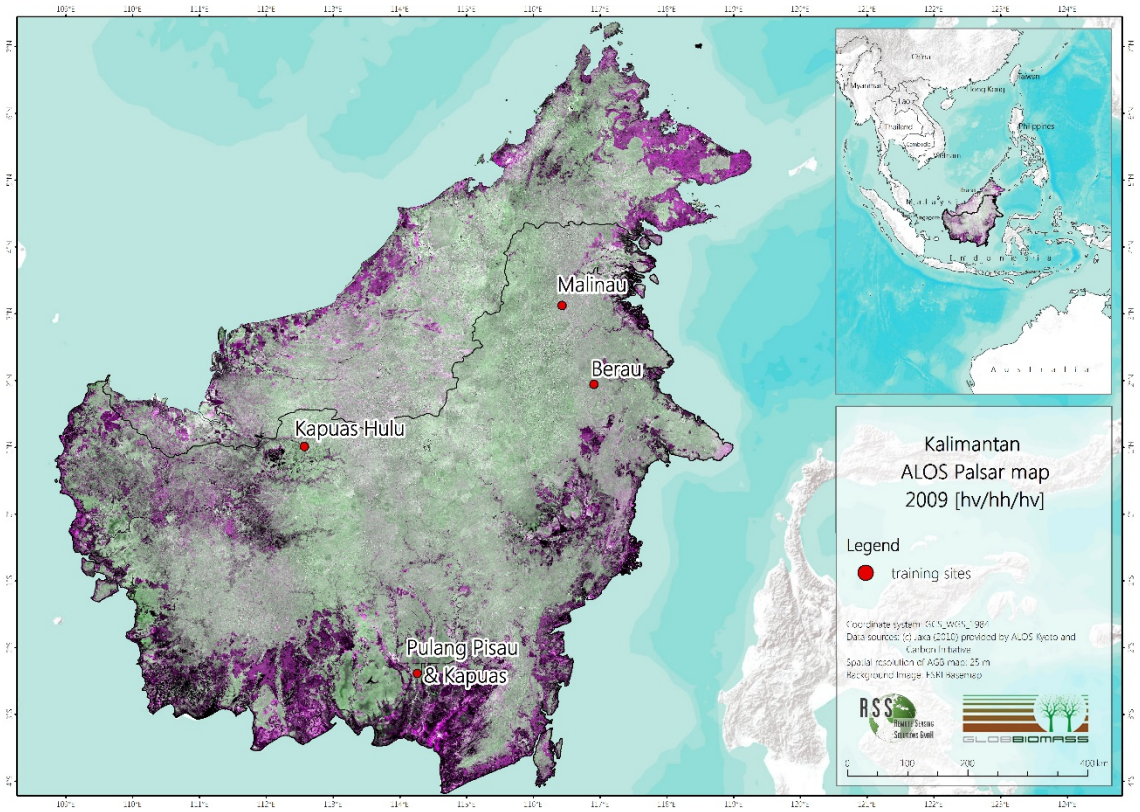


Figure 3-9. ALOS PALSAR false color composite of Kalimantan: red – HV, green – HH, red – HV.

The ecosystem of Kalimantan includes different forest types: mangrove forests, peat swamp and freshwater swamp forests, the most extensive extent of heath forests in Southeast Asia, lowland dipterocarp forests, ironwood forests, forests on limestone and ultrabasic soils, hill dipterocarp forests and various montane formations (MacKinnon et al., 1996). Peat swamp forest has developed over a large sedimentary plain that extends south to the Java Sea. Extensive coastal swamps have developed over this plain, mainly during the past 10,000 years, creating massive peat domes, and elevating the land surface. Peatland is the most significant carbon store and sink in this area. Lowland dipterocarp and peat swamp forests can usually be well discriminated in the field by means of species composition, average tree height, tree crown diameter, and canopy closure, with lowland dipterocarp forest being more diverse with taller trees and more closed canopy (MacKinnon et al., 1996). Different stages of forest degradation are present in Kalimantan, resulting from present and former logging activities

(over the last three decades) and fire events facilitated in anthropogenic degraded forests during strong El Niño events.

Three training sites with available AGB reference data were selected across Kalimantan in order to calibrate the regional AGB models:

- Central Kalimantan: This training area is located in the Pulang Pisau and Kapuas districts, in the heart of Central Kalimantan and is characterized by peatlands up to 16 m deep (Vepakomma et al., 2011). The most severe impact was caused by the Ex-MRP (Mega Rice Project), conceived by the Indonesian government in 1995 to convert an uncultivated area of 988,568 ha through the construction of about 6,000 km of drainage and irrigation channels between 1996 and 1997. By the time the MRP was abandoned, in 1999, more than 0.5 Mha of peatlands were deforested, 4,600 km of drainage canals were constructed to drain the peatlands, and an estimated 60,000 migrants moved to the area to live and work in the project area (Moeliono et al., 2009). These conditions, combined with a severe El Niño event in 1997, resulted in extensive forest and peatland fires that burned around 5.2 million ha, mostly on peat (Siegert et al., 2001). According to Page et al. (2002), Indonesia emitted between 0.81 and 2.57 Gt of carbon in 1997 as a result of burning peat and vegetation. Illegal logging activities in Central Kalimantan usually create only small scale impacts on the forest canopy. Timber trees are felled, cut into manageable lengths and then dragged along narrow skid trails (mostly about 3 to 10 meters wide) to the nearest river, or are floated through small canals (0.5 to 1.5 meters wide), which have been cut into the peat layer, to larger channels for further downriver transportation. This procedure causes much less visible impact on the forest than industrial logging operations, which create extensive infrastructure. After removing commercially valuable timber through industrial logging in the 1990s and recent illegal logging, human induced fires are observed spreading from drainage channels that are initiated for land speculation (Franke et al., 2012).

- West Kalimantan: This training site is located in the district of Kapuas Hulu in the central part of Borneo and belongs to the West-Kalimantan province. The district covers a total area of 3 million ha. The major part of the district lies in a basin consisting of a complex pattern of peat domes, lakes, freshwater wetlands and the floodplain of the Kapuas, Borneo's largest river. These lowlands have an elevation between 25 m a.s.l. and 100 m a.s.l. and have generally flat topography, with the exception of some rocky outcrops. In the east and south-east of the district the terrain becomes more rugged, and the landscape gives way to the Müller mountains (south-east) and Kapuas Hulu mountain range (north-east). There are several threats that cause deforestation or forest degradation, such as large-scale estate plantations, small holder agriculture, selective logging and mining (Potter and Lee, 1998).

- East Kalimantan: This training area is located in the districts of Berau and Malinau in East Kalimantan. Berau was relatively unaffected by the major fires of 1982-83 and 1997-98 and 75% of the area is still primary or secondary forest. Malinau is part of the Heart of Borneo, a conservation agreement initiated by the World Wildlife Fund for Nature to protect 220,000 km² of forested region in Borneo, and covers a total area of 4 Mha. Terrestrial ecosystem variation in both districts remains extremely high, including nearly all of the major ecosystem types known for Borneo (Whitmore, 1984). Coastal regions are dominated by mangrove, estuarine and mixed freshwater and peat swamp ecosystems. Further inland of these swamps,

extensive areas of lowland mixed dipterocarp or hill dipterocarp forest are present (Whitmore, 1984).

3.3.2 Datasets

3.3.2.1 *Input data*

The regional map for Indonesia is based solely on SAR imagery due to the important advantage of daylight and weather independence as the signal can penetrate through clouds, haze and smoke. This is of special importance in the frequent cloudy tropical areas. Hence, multispectral imagery was not taken into account.

Previous work on AGB estimation in the tropics show that X and C-band data has found up to present only little application (Lu 2006). Pandey et al. (2010) found a correlation between C-band ENVISAT ASAR backscatter and AGB up to 250 t/ha. Most AGB studies in tropical forests were conducted on the basis of L-band SAR data (Hamdan et al. 2011; Mitchard et al. 2011; Ryan et al. 2012; Wijaya et al. 2009). Sarker et al. (2012) successfully predicted AGB in a subtropical forest on the basis of L-band ALOS PALSAR texture. P-band backscatter has proven to allow more accurate AGB predictions than L-band backscatter (Saatchi et al. 2011b). Based on these results and on data availability, this regional AGB estimation will be based on C- and L-band SAR data.

L-band SAR

ALOS PALSAR data used in this study are 25 m resolution mosaics provided by the Kyoto & Carbon Initiative which are openly available and free of charge. The original ALOS PALSAR FBD data with 12.5 spatial resolution were not available over the whole regional area due to data costs. AGB estimation based on ALOS PALSAR mosaic data has already been successfully performed by Mermoz et al. (2014) and Hamdan et al. (2015).

The mosaic covers almost all land areas and is available for the years 2007, 2008, 2009 and 2010. The dataset is divided into tiles of 1 degree latitude-longitude geographical unit. The data were obtained in HH and HV modes as normalized radar cross-section, γ^0. They are slope-corrected and ortho-rectified to a ground spacing of 25 m using the 90-m SRTM (Shimada et al., 2014). For this study, all tiles covering the area of Kalimantan were used. Acquisitions were during the dry period between June and September. Although SAR is nominally weather-independent, SAR images show distortions due to rainfall events. The 2007 mosaic will be used for the 2005 epoch and the 2009 mosaic will be used for the 2010 epoch, as the year 2010 was very wet and the backscatter is influenced by the high moisture content of the vegetation. The Kyoto & Carbon Initiative will also provide a 25 m mosaic of ALOS-2 PALSAR-2 data in HH and HV polarization which will be used for the 2015 epoch.

C-band SAR

The C-band (5.6 cm wavelength) SAR database consists of ENVISAT ASAR and Sentinel-1 data. The ASAR instrument on ENVISAT, which was launched in March 2002, provides radar data in five modes of operation with varying spatial and temporal resolution: Image mode (IM), Alternating Polarization (AP), Wide swath (WS), global monitoring (GM) and Wave mode (WM). The IM mode has a spatial resolution of 30 m and either VV or HH polarizations; AP mode has a spatial resolution of 30 m and three possible mutual exclusive polarizations (HH/VV, HH/HV or VV/VH); and WS mode has a spatial resolution of 150 m and either VV or HH polarizations. For the 2005 epoch, full coverage of the study area can be obtained by 52 IM scenes acquired in VV polarization during the dry and wet season in

2004 and 2005. Images acquired only during the dry season do not cover the whole study area but the possible influence of flooded areas will be taken into consideration. If there is likely to be an influence of moisture content or flooded areas, these images will be excluded from the analysis. The 2010 reference epoch is not fully covered by scenes acquired in one sensor mode and is therefore excluded from the analysis.

Sentinel-1 is part of the EU/ESA Copernicus Programme and consists of a constellation of two satellites providing C-band imagery. Sentinel-1a was launched in April 2014 and data acquired in the Interferometric Wide swath (IW) mode during the dry season in 2015 will be used for the 2015 reference epoch. This is the default acquisition mode over land, and gives a 250 km wide swath in VV and VH polarizations with 10 m pixel spacing, a resolution of 20x22 m with 5x1 averaging and an ENL of 4.9.

SRTM
The Digital Elevation model (DEM) from the Shuttle Radar Topography mission (SRTM) with a spatial resolution of 30 m (and vertical accuracy of ± 10 m) is used for topographic analyses: slope is used to clean up the final AGB map, since extreme overestimation of biomass occurs in steep terrain.

Water body mask
ESRI World Water Bodies provides a base map layer for lakes, seas, oceans and large rivers and was used for delineating water bodies within the regional site.

MODIS active fire data
MODIS hotspot data (MCD14DL) is used to detect thermal anomalies/active fires.

Urban areas
The ESA CCI land cover map provides 3 epoch series (200, 2005, 2010) of global land cover maps at 300 m spatial resolution. The land cover class urban areas was used to evaluate settlement areas.

3.3.2.2 *Training data*

AGB reference data were generated by relating forest inventory AGB data to LiDAR measurements, taking advantage of LiDAR's ability to create accurate biomass predictions for an area within the SAR images.

Deforestation between SAR and LiDAR acquisition dates was detected using MODIS hotspots, as fire is the main reason for deforestation in this area. Areas of 1 km x 1 km centred on the hotspots were excluded when selecting reference data.

AGB reference data were selected with a minimum distance of 75 m to the LiDAR outline as LiDAR point density and therefore AGB estimation accuracy is lower at the edge of the dataset. Areas with slope exceeding 5% and inside the ALOS mask (containing areas of layover and shadow) were also excluded from the analysis.

The AGB reference data for each epoch was randomly split into data used for training (70%), validation (15%) and accuracy assessment (15%) of the SAR-AGB models.

Field inventory data

Existing field inventories were conducted within different projects. In general, the sample plots in forests had the same circular extent consisting of nested subplots arranged in concentric circles (see Figure 3-10). Nested sample plots are recommended in highly diverse tropical forests in order to get information on both widely distributed large trees and on smaller but more densely distributed trees. In regrowth areas, rectangular plots were used and all saplings and trees within the area were recorded.

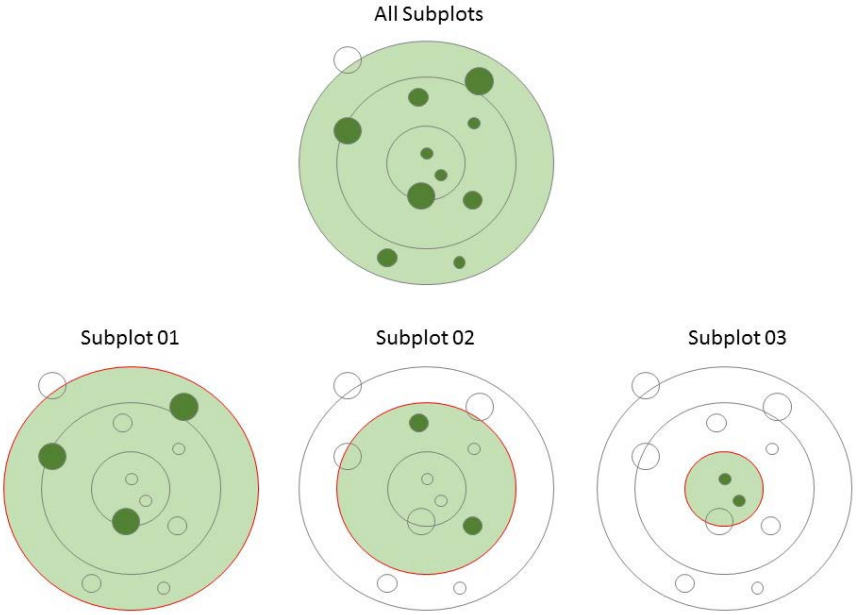


Figure 3-10. Example of a nested plot consisting of three subplots. In each of these subplots trees with different diameter at breast height (DBH) are measured.

Table 3-3 gives a detailed description of all forest inventories available for the regional AGB estimation with respect to each epoch. The methodology and applied allometric models vary slightly for the different study sites since they were processed as part of different research projects. Although the different methodology and applied allometric models might have an influence on AGB estimation, it is assumed that they do not have a substantial influence as it is not possible to re-process the plot due to lack of access to raw data.

Table 3-3. Detailed description of all available field inventory plots for the regional AGB estimation.

Site	Acquisition Date	Reference epoch	Number of plots	Plot form	Plot design	AGB allometry
Pulang Pisau & Kapuas (Central Kalimantan)	2008	2005	64	Nested circular plots	Degraded forest: Subplot 01: DBH ≥20 cm; 20 m radius Subplot 02: DBH ≥10 cm and <20 cm; 14 m radius Subplot 03: DBH ≥5 cm and <10 cm; 4 m radius Intact forest: Subplot 01: DBH ≥50 cm; 20 m radius Subplot 02: DBH ≥20 cm and <50 cm; 14 m radius Subplot 03: DBH ≥5 cm and <20 cm; 4 m radius	AGB=exp(-2.977+ln(ρ*D²*h) After Chave et al. (2005) using wood specific densities applied to degraded and intact forest plots.
Pulang Pisau & Kapuas (Central Kalimantan)	2010-2011	2010	41	Nested circular plots	Degraded forest: Subplot 01: DBH ≥20 cm; 20 m radius Subplot 02: DBH ≥10 cm and <20 cm; 14 m radius Subplot 03: DBH ≥5 cm and <10 cm; 4 m radius Intact forest: Subplot 01: DBH ≥50 cm; 20 m radius Subplot 02: DBH ≥20 cm and <50 cm; 14 m radius Subplot 03: DBH ≥5 cm and <20 cm; 4 m radius	AGB=exp(-2.977+ln(ρ*D²*h) After Chave et al. (2005) using wood specific densities a pplied to degraded and intact forest plots.

Table 3-3. Cont.

Site	Acquisition Date	Reference epoch	Number of plots	Plot form	Plot design	AGB allometry
Pulang Pisau & Kapuas (Central Kalimantan)	2010-2011	2010	46	Rectangular plots	Regrowing forest: Rectangular plots of 20 m x 50 m: record of all saplings and trees	If DBH<5 cm and h≤1.3 m after Hughes et al. (1999) $AGB=10^{-}6*\exp(4.7472+1.0915*\ln(D^2))$ If DBH<5 cm and h>1.3 m after Hughes et al. (1999) $AGB=1.14*10^{-6}*\exp(4.9375+1.0583*\ln(D^2)$ If DBH<5 cm and h>1.3 m after Hughes et al. (1999) $AGB=\exp(-2.977+\ln(\rho*D^2*h)$ All allometries were used with wood specific densities

Table 3-3. Cont.

Site	Acquisition Date	Reference epoch	Number of plots	Plot form	Plot design	AGB allometry
Berau (East Kalimantan)	2012-2013	2010	78	Nested circular plots	Subplot 01: DBH ≥ 50 cm; 35 m radius Subplot 02: DBH ≥ 20 cm and <50 cm; 25 m radius Subplot 03: DBH ≥ 10 cm and <20 cm; 10m radius Subplot 04: DBH ≥ 2 cm and <10 cm; 3 m radius	$AGB=\rho*\exp(-1.499+2.148*\ln(D)+0.207*(\ln(D))^2-0.0281*(\ln(D))^3)$ After Chave et al. (2005) using wood specific densities
Kapuas Hulu (West Kalimantan)	2009-2011	2010	82	Nested rectangular plots	Subplot 01: DBH >30 cm; 10 m x 10 m Subplot 02: DBH ≥15 cm and ≤ 30 cm; 20 m x 20 m Subplot 03: DBH ≥5 cm and ≤ 15 cm; 20 m x 50 m	$AGB=\rho*\exp(-1.499+2.148*\ln(D)+0.207*(\ln(D))^2-0.0281*(\ln(D))^3$ After Chave et al. (2005) using an average wood density for Asian tropical forests of 0.57 $mg \cdot m^{-3}$ (Brown, 1997)
Pulang Pisau & Kapuas (Central Kalimantan)	2013-2014	2015	94	Nested circular plots	Subplot 01: DBH ≥30 cm; 16 m radius Subplot 02: DBH ≥17 cm and <30 cm; 8 m radius Subplot 03: DBH ≥5 cm and <17 cm; 4 m radius	$AGB=\rho*\exp(-1.499+2.148*\ln(D)+0.207*(\ln(D))^2-0.0281*(\ln(D))^3)$ After Chave et al. (2005) using wood specific densities

Table 3-3. Cont.

Site	Acquisition Date	Reference epoch	Number of plots	Plot form	Plot design	AGB allometry
Malinau (East Kalimantan)	2015	2015	24	Nested circular plots	Subplot 01: DBH ≥50 cm; 30 m radius Subplot 02: DBH ≥20 cm and <50 cm; 20 m radius Subplot 03: DBH ≥10 cm and <20 cm; 10 m radius Subplot 04: DBH <10 cm; 3 m radius	$AGB=\rho*\exp(-1.499+2.148*\ln(D)+0.207*(\ln(D))^2-0.0281*(\ln(D))^3$ After Chave et al. (2005) using an average wood density for Asian tropical forests of 0.57 $mg \cdot m^{-3}$ (Brown, 1997)
Kapuas Hulu (West Kalimantan)	2014	2015	44	Nested circular plots	Subplot 01: DBH ≥50 cm; 30 m radius Subplot 02: DBH ≥20 cm and <50 cm; 20 m radius Subplot 03: DBH ≥10 cm and <20 cm; 10 m radius Subplot 04: DBH ≥2 cm and <10 cm; 3 m radius	$AGB=\rho*\exp(-1.499+2.148*\ln(D)+0.207*(\ln(D))^2-0.0281*(\ln(D))^3$ After Chave et al. (2005) using an average wood density for Asian tropical forests of 0.57 $mg \cdot m^{-3}$ (Brown, 1997)

Reference biomass data derived from airborne LiDAR

Airborne LiDAR (Light Detection and Ranging) measurements were acquired during the dry season (May-October) at an altitude of approximately 800 m a.s.l. This flight altitude allows a transect width with two overflights of approx. 507 – 563 m with 35% overlap of the two flight strips. **Error! Reference source not found.**-4 provides an overview of the LiDAR data.

Table 3-4. Acquisition dates and characteristics of the airborne LiDAR data.

Site	Acquisition Date	Reference epoch	Area
Pulang Pisau & Kapuas (Central Kalimantan)	2007	2005	300 km²
Pulang Pisau & Kapuas (Central Kalimantan)	2011	2010	7000 km²
Malinau (East Kalimantan)	2012	2010	240 km²
Berau (East Kalimantan)	2012	2010	340 km²
Kapuas Hulu (West Kalimantan)	2012	2010	420 km²

Previous studies revealed that LiDAR point cloud height metrics like Quadratic Mean Canopy Height (QMCH) or Centroid Height (CH) are appropriate parameters to estimate AGB in tropical forests by taking also the point distribution over the different vegetation layers into account (Ballhorn et al., 2011; Kronseder et al., 2012; Jubanski et al., 2013). LiDAR height histograms were calculated by normalising all points within a grid of 30 m (similar to the size of the largest nest of the field inventory plots) to the ground using the DTM as reference, i.e. the height of each LiDAR return was calculated relative to the DTM. The number of points within each 0.5 m interval was stored as a histogram. The first (lowest) interval was considered as the ground return and excluded from further processing. The QMCH and the CH of the height histogram were calculated by weighting each 0.5 m height interval by the fraction of points stored within this interval. The QMCH and CH were related to AGB estimated by field inventory using regression models. Jubanski et al. (2013) showed that the accuracy of AGB estimations derived from LiDAR height histograms increased with higher point densities, so point density was also included in the regression as an input variable.

The LiDAR AGB regression model combined a power function in the lower biomass range, up to certain threshold $QMCH_0$ (the example here uses QMCH but the same would be done with CH) and a linear function in the higher biomass range. The $QMCH_0$ threshold was determined by increasing the value of $QMCH_0$ in steps of 0.001 m and identifying the lowest RMSE. The linear function is the tangent through $QMCH_0$ and was calculated using the first derivative of the power function (Eq. 3-12):

$$AGB = \begin{cases} a \cdot QMCH^b & if\ QMCH \leq QMCH_0 \\ \left(a \cdot b \cdot QMCH_0^{(b-1)}\right)(QMCH - QMCH_0) + a \cdot QMCH_0^b & if\ QMCH > QMCH_0 \end{cases} \quad \text{(Eq.3-12)}$$

where *a* and *b* are coefficients (the same can be done using CH). More details are provided in Englhart et al. (2013). The CH was used for AGB estimation in Central Kalimantan and Malinau while the QMCH was used in Berau and Kapuas Hulu. The different LiDAR AGB regression models are depicted in Figure 3-11.

The AGB reference data were generated by relating forest inventory AGB data to LiDAR measurements, taking advantage of LiDAR's ability to create accurate biomass predictions for an area within the SAR images. This upscaling from point data (field inventory) to transects (LiDAR) was chosen to provide a more powerful basis for AGB model calibration and validation from SAR backscatter data.

Deforestation between SAR and LiDAR acquisitions was detected using MODIS hotspots, as fire is the main reason for deforestation in this area. Areas of 1 km x 1 km centred on the hotspots were excluded when selecting reference data.

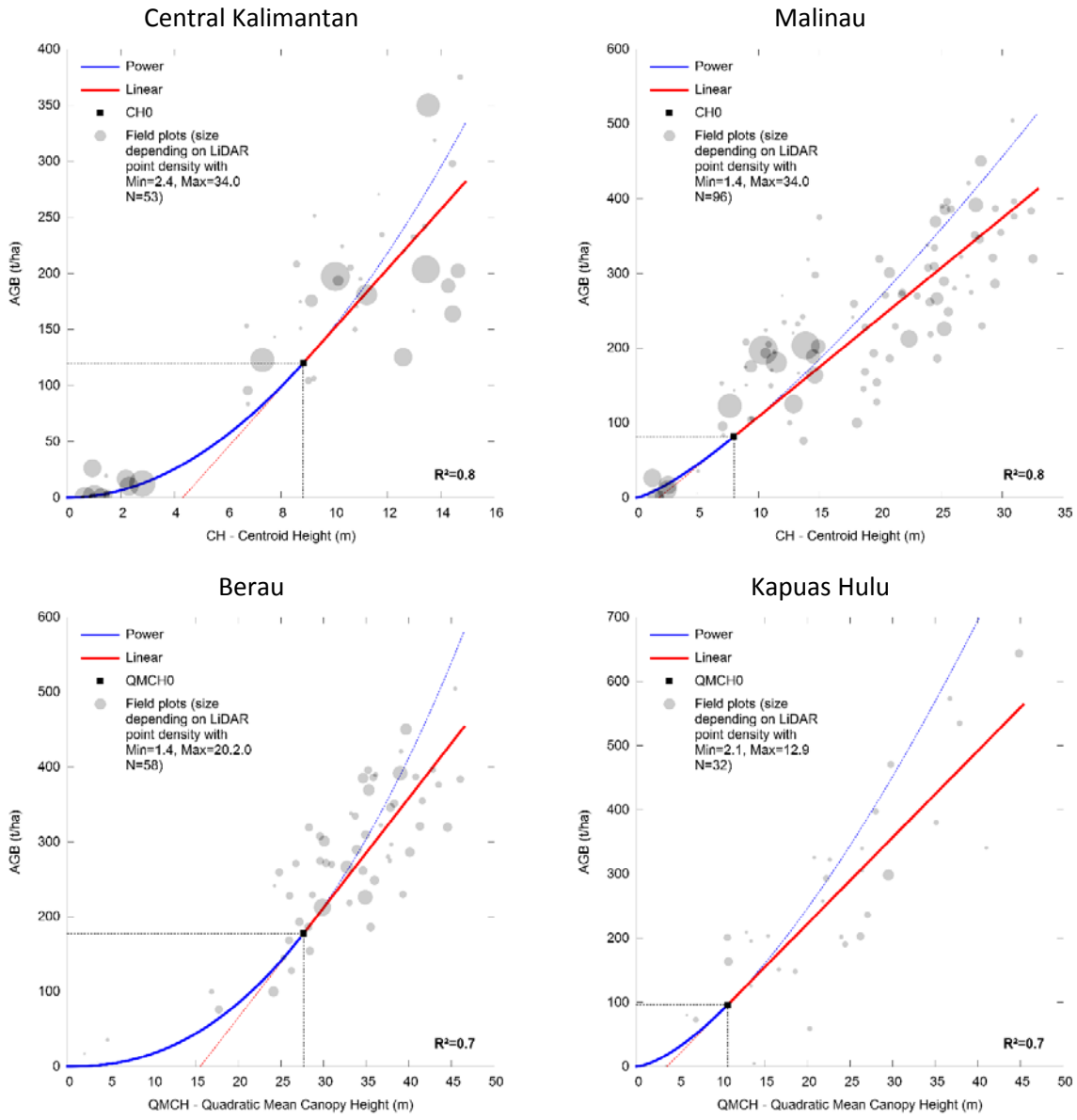


Figure 3-11. LiDAR AGB regression models based on field inventory data and the Quadratic Mean Canopy Height and Centroid point cloud height metrics.

Spaceborne LiDAR data (ICESAT GLAS)

ICESAT GLAS data could not be used for AGB estimation in this study as intended in the proposal since there was not enough overlap between ICESAT GLAS footprints and AGB reference data (field plots or airborne LiDAR).

3.3.2.3 *Validation data*

The AGB reference data for each epoch were randomly split into data used for training (70%), validation (15%) and accuracy assessment (15%) of the SAR-AGB models.

3.3.2.4 *Data used for accuracy assessment*

The AGB reference data for each epoch were randomly split into data used for training (70%), validation (15%), and accuracy assessment of the SAR-AGB models (15%).

3.3.3 Methods

3.3.3.1 *Pre-processing of SAR data*

The preprocessing of SAR imagery included co-registration, radiometric calibration, geometric correction and speckle filtering.

The multi-temporal SAR imagery was co-registered using the ALOS PALSAR mosaic of 2010 as reference and was processed using γ^0 backscatter coefficients. A workflow showing the steps in the pre-processing based on the example of ALOS PALSAR 25 m mosaic is shown in Figure 3-12.

To reduce speckle, a multi-channel speckle filter with a 7x7 moving window was applied (Quegan et al., 2000). This generates new images with reduced speckle from multi-temporal and multi-polarised images. It is based on the following relation:

$$J_k(x,y) = \frac{\langle I_k(x,y)\rangle}{N}\sum_{i=1}^{N}\frac{I_i(x,y)}{\langle I_i(x,y)\rangle} \quad \text{with} \quad k = 1,\ldots,N \qquad \text{(Eq. 3-13)}$$

where, at position (x,y), $J_k(x,y)$ is the radar intensity of output image *k*, $I_i(x,y)$ is the radar intensity of input image *i*, $<I_i(x,y)>$ is the local average intensity of input image *i* and *N* is the number of images.

In addition to the backscatter images, ratio images were prepared after speckle filtering using the following equations (depending on available polarizations) in order to examine the potential for AGB estimation (Thapa et al. 2015; Hamdan et al., 2014):

- $R_{hvhh} = \frac{HV}{HH}$
- $R_{vhvv} = \frac{VH}{VV}$
- $SQRT_{hhhv} = \sqrt{HH \cdot HV}$
- $SQRT_{vhvv} = \sqrt{VH \cdot VV}$

where HH, HV, VV and VH indicate the polarizations of the γ^0 backscattering coefficients.

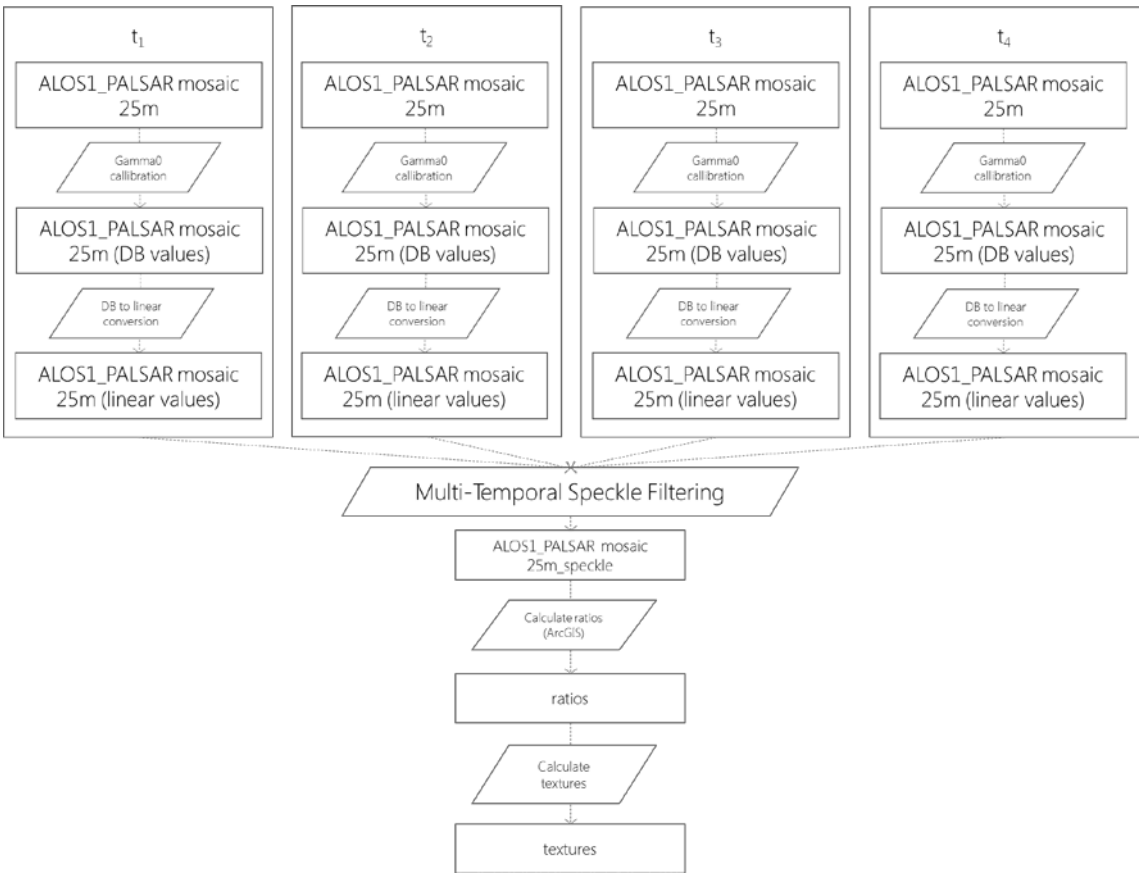


Figure 3-12. Workflow of SAR pre-processing using the example of ALOS PALSAR 25 m mosaic.

In order to yield more information on forest structure and improve the model accuracy, texture measures and their relationship to AGB were evaluated. Texture describes properties of objects, such as smoothness, regularity, and tonal variation in a SAR image (Thapa et al., 2015). Using the freely available Orfeo toolbox (https://www.orfeo-toolbox.org), textures were calculated based on different ratios and on the single polarized images, including simple and higher order Haralick textures over a sliding window with user defined radius:

(Eq. 3-14 to 3-20)

- $SRE = \frac{1}{n_r}\sum_{i,j}\frac{p(i,j)}{j^2}$
- $LRE = \frac{1}{n_r}\sum_{i,j} p(i,j) * j^2$
- $RLN = \frac{1}{n_r}\sum_{j}(\sum_{i} p(i,j))^2$
- $RP = \frac{n_r}{n_p}$
- $HGRE = \frac{1}{n_r}\sum_{i,j} p(i,j) * i^2$
- $SRLGE = \frac{1}{n_r}\sum_{i,j}\frac{p(i,j)}{i^2 j^2}$
- $ST = \sum_{i,j}\frac{(i-\mu)(j-\mu)g(i,j)}{\sigma^2}$

where $p(i,j)$ is the element in cell *(i, j)* of a normalized Run Length Matrix, n_r is the total number of runs and n_p is the total number of pixels (Haralick et al. 1973).

Due to varying resolution, all input data (filtered backscatter, ratios and textures) were averaged to a pixel spacing of 100 m, which is consistent with the pixel spacing of the final AGB map.

3.3.3.2 *Biomass estimation algorithms*

The regional AGB estimation map is based on full coverage C- and L-band SAR data or only L-band data if full coverage by C-band data is not available. Regression analysis is a common way to develop AGB estimation models using SAR data (Lu 2006) and this method was chosen to develop this regional AGB estimation model.

In previous work, we have compared different methodologies to estimate AGB at this regional site and found that multivariate linear regression models are superior to support vector regression and artificial neural network models in terms of AGB variability and saturation (Englhart et al. 2012). Therefore, we decided to use regression modelling as it is a very common approach in the field of AGB estimation in the tropics using SAR data (Thapa et al. 2015, Mermoz et al. 2015, Mitchard et al. 2009.

As a first step, the plots showing the relationship between input parameters such as backscatter, ratios or textures and AGB were assessed. After analysing the relationship between AGB and input variables, a multiple linear regression model was applied using the linearized input variables.

$$AGB = a_1 \cdot \mathrm{var}_1 + a_2 \cdot \mathrm{var}_2 + \ldots + a_n \cdot \mathrm{var}_n + c \qquad \text{(Eq. 3-21)}$$

where $a_1, a_2, \ldots, a_n$ are coefficients and $\mathrm{var}_1, \mathrm{var}_2, \ldots, \mathrm{var}_n$ are linearized input variables (e.g. exponential values of input data), such as γ^0, ratios or textures.

3.3.3.3 *Training methods*

To establish the relationship between AGB and the independent variables (SAR derived parameters), a multiple linear regression approach is adopted. Although complexity increases, combining multiple variables often provides better AGB estimates than using single variables. However, multiple regressions are often affected by multi-colinearity and overfitting among the independent variables. In order to identify a model with the greatest explanatory power, stepwise multiple linear regression was performed to automate the selection of the best explanatory variables.

In addition, the p-value of each variable was used to filter out those variables that do not contribute to the model; this approach is often exploited in regression-based AGB modelling (Asner et al., 2010; Mitchard et al., 2011; Thapa et al., 2015). The Variable Inflation Factor (VIF) was calculated for the selected variables and models with higher VIF variables were avoided. Higher r^2, lower root mean square error (RMSE), the significance of p-value < 0.05 for the independent variables and the VIF were incorporated into the model selection and development process.

3.3.3.4 *Strengths, limitations and weaknesses of the developed approach*

The strength of the developed approach for epoch 2010 is that biomass variability due to different degradation stages in the forest is clearly visibly as indicated in Figure 3-13. AGB variability due to different disturbances are clearly visible in contrast to non-disturbed area or clearcuts.

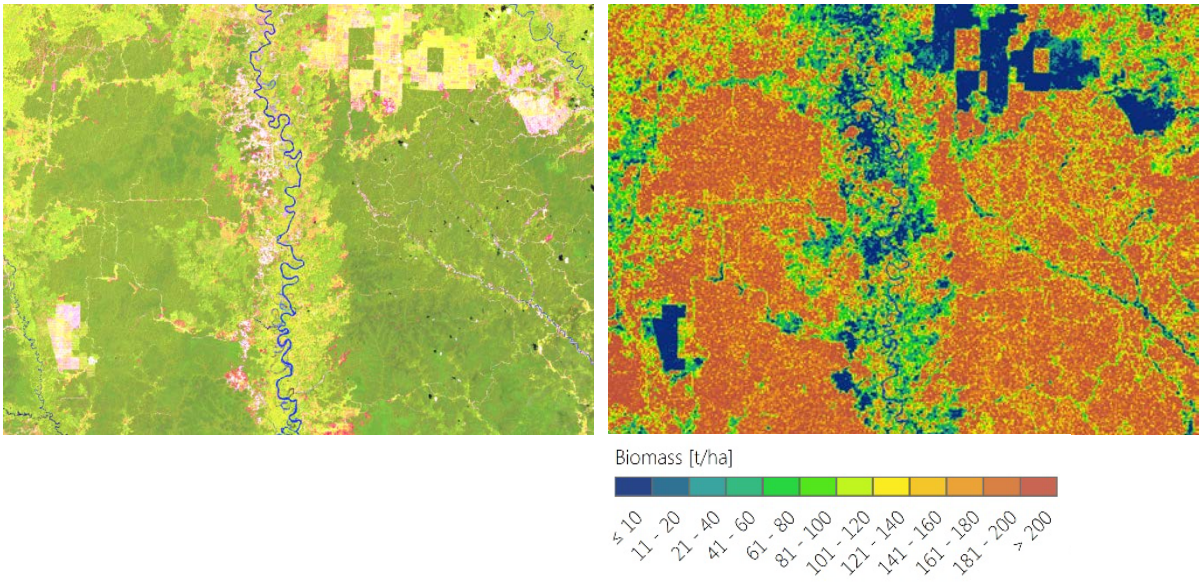


Figure 3-13. Comparison of Landsat image acquired on 10.02.2010 (left) and regional AGB map (right) showing AGB loss due to different degradation stages.

The comparison of the developed regional AGB map, other pan-tropical AGB maps and LiDAR-derived AGB estimation depicted in Figure 3-14 show that the developed model correctly estimates the variability in the lower and higher biomass ranges.

One of the limitations of this approach is that in general, SAR-based AGB retrieval suffers from saturation of the backscatter signal in the higher biomass range. The saturation level of L-band SAR was found in previous studies between 50 t/ha and 250 t/ha, and HV data gave higher saturation levels than HH data (Hamdan et al. 2011; Hamdan et al. 2015; Mitchard et al. 2012; Mitchard et al. 2009; Saatchi et al. 2007; Saatchi et al. 2011b, Englhart et al. 2011). Due to the use of texture, the saturation level of the 2010 model, which is only based on ALOS PALSAR 25 m mosaic data, is approximately at 250 t/ha. In addition, AGB estimations in steep terrain (with slope > 10°) and settlement areas are not reliable and have been flagged in the quality assurance (QA) layer.

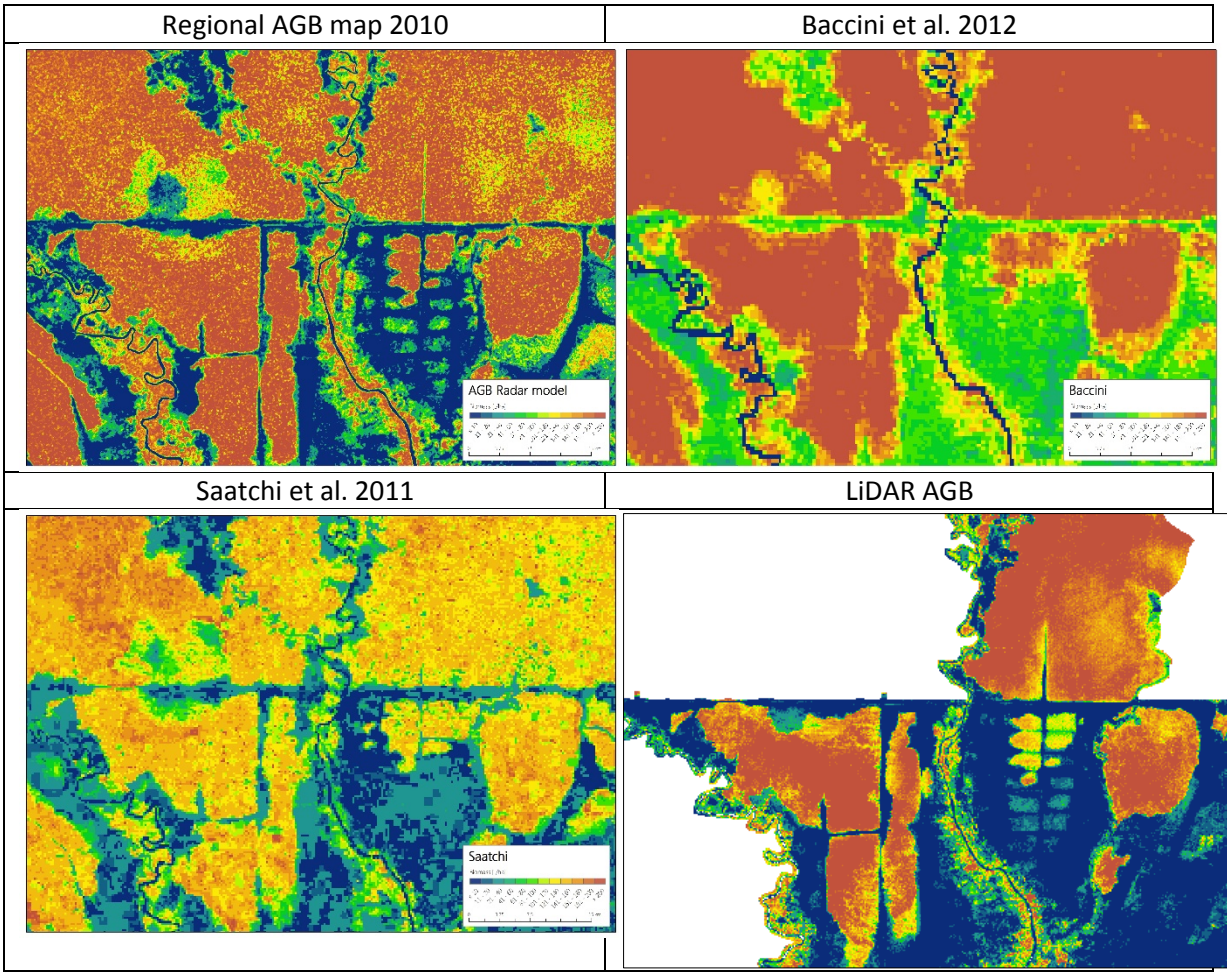


Figure 3-14. Comparison of the developed regional AGB map, other pan-tropical AGB maps from Baccini et al. 2012 and Saatchi et al. 2011 as well as airborne LiDAR AGB estimations developed after the description provided in 3.3.2.2.

3.3.3.5 *Methods to assign accuracies in the uncertainty map*

The total uncertainty at pixel level is composed of different sources of error which are assumed to be random and independent. These are propagated using the following equation proposed by Saatchi et al. (2011):

$$\varepsilon_{AGB} = (\varepsilon^2_{measurement} + \varepsilon^2_{allometry} + \varepsilon^2_{sampling} + \varepsilon^2_{prediction})^{1/2} \quad \text{(Eq. 3-22)}$$

$\varepsilon_{measurement}$: The measurement error of tree level parameters such as diameter and tree height averaged at plot level (Chave *et al.*, 2004). By re-measuring trees several times, the measurement error can be assessed. Similar to Mitchard et al. (2011), a measurement error of 10% is assumed.

$\varepsilon_{allometry}$: The error in estimating AGB using allometric equations. Chave et al. (2005) found an error on the estimation of a tree's biomass was approximately ±5%. As we mainly used this allometry, an error of 5% is assumed here.

$\varepsilon_{sampling\ size}$: This originates from the variability of AGB within the pixel area and depends on the size of the plots used to upscale the AGB measurements to the pixel level. According to Saatchi et al. (2011) and Chave et al. (2003), a sampling error of 20% is assumed.

$\varepsilon_{prediction}$: The prediction error includes both the sampling error associated with the representativeness of the training data of the actual spatial distribution of AGB and the model predictions. The uncertainty of the predicted AGB is esimated per Biomass ranges by dividing the RMSE by the mean of the AGB per range (Saatchi *et al*.,2011).

$$\varepsilon_{prediction} = \sigma_{\overline{AGB}}/\overline{AGB}$$

3.3.3.6 *Methods to test the accuracy of the measurements*

The validation dataset is used for accuracy assessment based on the recommendations from the GlobBiomass Validation protocol (D5 document).

3.3.4 Products

The final products for each epoch (2005, 2010 and 2015) include an AGB map, a quality assurance layer and an uncertainty map with a spatial resolution of 100 m. Current pan-tropical maps (Saatchi et al., 2011, Baccini et al., 2012, Avitable et al. 2016) have a coarser spatial resolution (500 m and 1000 m, respectively). Besides we created a binary change/no-change layer, a layer with the absolute values of change and a layer containing the probability of change with a spatial resolution of 100 m for the periods 2005-2010, 2010-2015 and 2005-2015.

3.3.5 Modifications for the 2005 and 2015 epochs

Methods

There are no changes in the methods for 2005 and 2015 AGB estimation when compared to the 2010 epoch.

Data

As described above, the following datasets will be used for the different epochs.

Input data

	2005	2010	2015
L-band SAR			
ALOS PALSAR mosaic 25m	X	X	
ALOS-2 PALSAR-2 mosaic 25m			X
C-band SAR			
Envisat ASAR	X		
Sentinel-1			X

Training data

	2005	2010	2015
Field inventory data	X	X	X
Airborne LiDAR data	X	X	

3.4 Mexico

3.4.1 General description of the region

Mexico's total land area is estimated at approximately 1.97 million km^2, of which 43% is covered by forest and other wooded land. The country borders with United States, Belize and Guatemala. The country's geology, topography, and climate are extremely varied, presenting different kinds of forest ecosystems. It presents broad mountains ranges, coastal lowlands, and large plateaus. The large climate gradient from North to South generates a wide spectrum of complex climate conditions. The country is divided by the Tropic of Cancer into sub-tropical north and tropical south. Mexico is also one of the countries with the highest biodiversity in the world. This biodiversity is mostly concentrated in the tropical forest areas of the country.

The north of the country is dominated by deserts, coniferous and broadleaved forests, while the centre and south are dominated by a diverse mix of coniferous forest, broadleaved forest, mixed forest, cloud forest in high elevations, savannahs, wetlands, and evergreen, deciduous and semi-deciduous tropical forests (Figure 3-15. Therefore the seasonality of the remote sensing observations will be a critical issue.

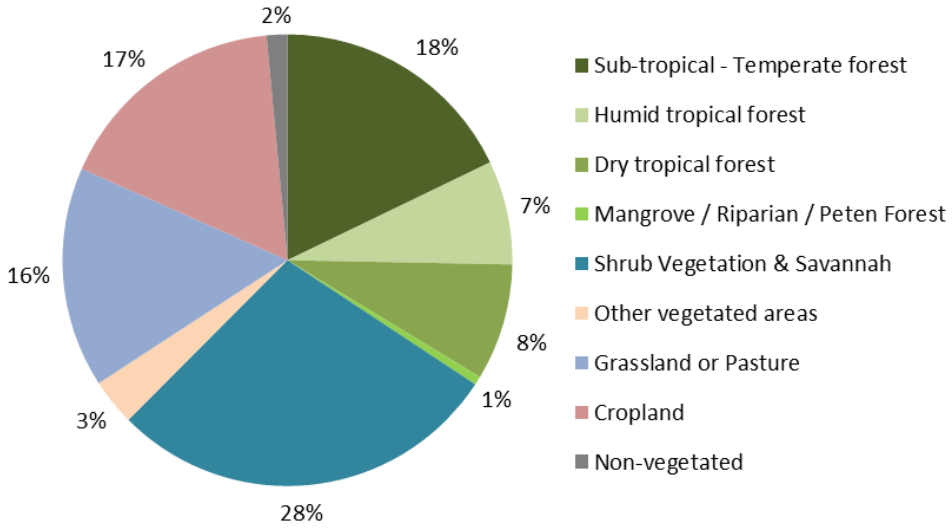


Figure 3-15. Land Cover and Vegetation types in Mexico (% total area)

Two sites covering the most representative biomes in Mexico were selected (Figure 3-16). The first is the Yucatan peninsula, which comprises a mix of tropical moist and tropical dry forest biomes, as well as the mangrove biome. The second proposed site is central Mexico which covers subtropical coniferous forest, tropical dry forest, tropical moist forest and xeric shrubland biomes, and includes forest with some of the highest biomass per ha in Mexico (i.e. Oyamel forest). The total area covered by both sites exceeds 300.000 km^2.

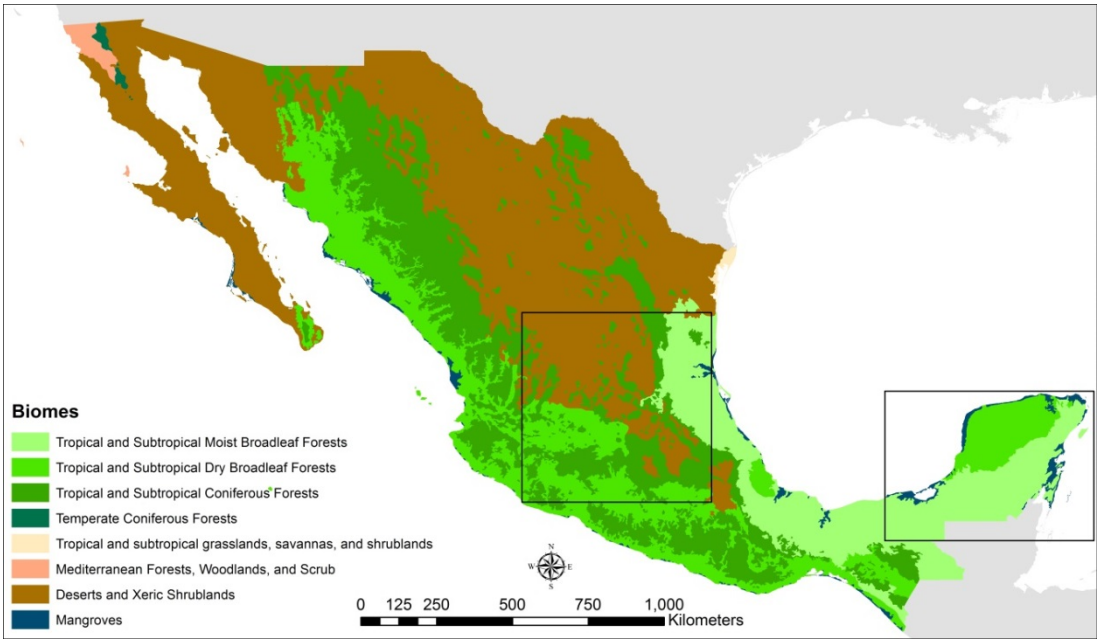


Figure 3-16. Biomes in Mexico. Squares correspond to study sites.

Few assessments have been carried out to estimate the carbon stock in Mexico. The main assessments, based on forest inventory data, were carried out by Comision Nacional Forestal (CONAFOR) to provide information for the FAO Forest Resource Assessments. The most recent assessments covering the whole country are the tropical carbon maps for 2000 and 2005 by Saatchi, Harris et al. (2011) and Baccini, Goetz et al. (2012) respectively. Cartus, Kellndorfer et al. (2014) also produced a map of forest above-ground carbon stocks for 2005 for Mexico. These studies followed different approaches and display substantial differences in the amount and distribution of AGB stocks in Mexico (Table 3-5).

3.4.2 Datasets

3.4.2.1 *Input data (used to drive the biomass algorithm)*

Remote sensing imagery sensitive to forest AGB from different satellite sensors (optical and SAR) will be used (Table 3-6). The core datasets will be ALOS PALSAR and Landsat. Freely available ALOS PALSAR (and ALOS-2 PALSAR-2) mosaics of γ^0 at 25 m pixel spacing in HH and HV polarizations were acquired from JAXA (http://www.eorc.jaxa.jp/ ALOS/en/palsar_fnf/fnf_index.htm) for all available years (2007-2011 & 2015-2016). Landsat Surface Reflectance (SR) imagery computed by the LEDAPS method (http://ledaps.nascom.nasa.gov/) (Masek, Vermote et al. 2006) was used to generate multi-temporal composites. Google Earth Engine (GEE) cloud computing (https://earthengine.google.org/) is used for this task. Landsat 7 ETM+ and Landsat 8 OLI imagery are used to generate these composites. Landsat Percent Tree Cover (PTC) and forest cover loss layers (Hansen, Potapov et al. 2013) were acquired from University of Maryland repository (https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.3.html) This dataset will be used as additional input. Additionally, freely available 30 m spatial resolution elevation data from the Shuttle Radar Topography Mission (void-filled SRTM Plus NASA V3) will be obtained from the USGS Earth Explorer repository (http://earthexplorer.usgs.gov/). Sentinel-1 dual polarisation Interferometric Wide (IW) swath mode imagery (5 m x 20 m pixel spacing) acquired from the Sentinel-1 scientific data hub (https://scihub.esa.int/) will be evaluated for the 2015 epoch map, and used if it contributes to improve the estimations.

Table 3-5. Comparison between previous assessments and proposed map for Mexico.

Assessment	Period	Method	Datasets	Spatial Resolution	Spatial Outputs	Total Carbon (Pg C)	Mean Carbon (t C ha^{-1})	RMSE pixel level (t C ha^{-1})
FAO (2010)	2000 2005 2010	Forest Inventory	INFyS, LUV	N/A	N/A	1.75 1.72 1.68	26.24 26.22 25.93	N/A
Saatchi, Harris et al. (2011)	2000	MaxEnt	GLAS, MODIS (1000m), QSCAT (2.25km), SRTM (90m)	1000 m	AGB map & Uncertainty map	2.24	32.94	±27.3%(*)
Cartus, Kellndorfer et al. (2014)	2005	Random Forest	INFyS, LUV, Landsat PTC (30m), ALOS PALSAR (30m), SRTM (90m)	30 m	AGB map	1.53	23.61	±14.4
Baccini, Goetz et al. (2012)	2005	Random Forest	GLAS, MODIS, SRTM (500m)	500 m	AGB map	1.95(**)	46.35	±25.0
Proposed Mexico GlobBiomass Regional map	2005 2010 2015	MaxEnt	INFyS, Landsat composites (SR) (30m), ALOS PALSAR (25m), SRTM (30m) Optional: Landsat PTC (30m), Sentinel-1 (5 m x 20 m)	25 m	AGB & Uncertainty maps, AGB-change maps			(***)

(*) Relative error; (**) Country half covered; (***) Expected better than existing maps

Several spatial products classifying Mexico are available free of cost at medium and coarse resolutions (WWF biomes and eco-regions, GlobCover, MODIS Land Cover, etc.). The different series (III, IV, & V) from the Land Cover and Vegetation map from the Mexican National Institute for Statistics and Geography (INEGI), developed using Landsat and Spot imagery, were also acquired (INEGI 2009). The map provides land use and vegetation types at a scale of 1:250000 and will be used to mask out urban areas from the study sites.

Table 3-6. Remote sensing imagery for the regional biomass maps in Mexico.

DATASET	SPATIAL RESOLUTION	Number of Looks	2005	2010	2015
ALOS PALSAR	25 m	16	x[(1)]	x	
ALOS-2 PALSAR-2	25 m	16[(2)]			x
Landsat PTC[(3)]	30 m	N/A	x	x	x
Landsat 7 ETM+ SR	30 m	N/A	x	x	
Landsat 8 OLI SR	30 m	N/A			x
SRTM Plus (NASA V3)[(4)]	30 m	N/A	x	x	x
Sentinel-1 Dual Polarisation IW Mode	5 m x 20 m	5x1 (4.9 ENL)			x

[(1)] Data from 2007. [(2)] Assuming the same processing chain used in the ALOS PALSAR mosaics. [(3)] PTC updated using the loss layer. [(4)] Shuttle Radar Topography Mission void-filled elevation data for the year 2000. As it is assumed that topography remains constant, the same dataset is used for all epochs.

3.4.2.2 *Training data*

The Mexican National Forest and Soil Inventory (INFyS) of the Comision Nacional Forestal (CONAFOR) is a rigorously designed and extensive ground-based national forest inventory that provides accurate and current information on the size, spatial distribution and condition of forest resources (SEMARNAT 2004). This information is used to support the development of national policies for sustainable development and to promote forestry sector activities. The average forest AGB for the country is estimated, based on the INFyS dataset, as approximately 53 t ha^{-1}. Maximum AGB values recorded in this dataset are as high as 400 t ha^{-1}, but very few plots have values above 200 t ha^{-1}. (Figure 3-17):

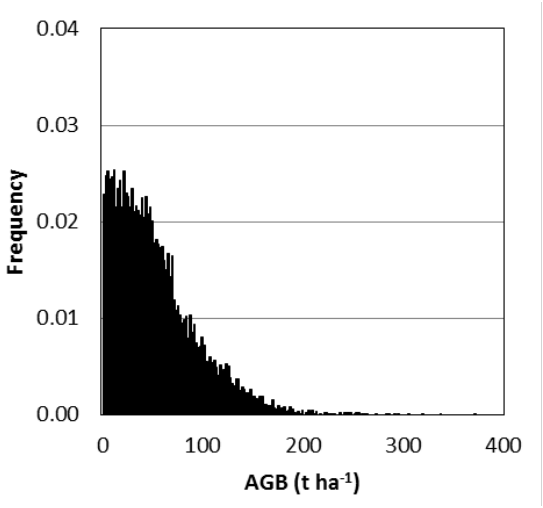


Figure 3-17. Histogram of AGB distribution across the whole country based on the Mexican national forest Inventory dataset (INFyS).

The ground plot dataset from INFyS contains data from 17,171 cluster plots each comprising four 400 m^2 (0.04 ha) rectangular (tropical forest) or circular (other forest) sub-plots representing a circular area of 56.42 m radius (1 ha) systematically located across forested areas in Mexico for the period 2004-2012 (Figure 3-18). A total area of 0.16 ha is sampled at each 1 ha plot. The algorithm is trained using the average value of the pixels within the boundaries of the INFyS ground plot. Edge pixels that intersect the outer boundaries of the plots will be excluded. Only plots with 4 sub-plots are used. Exponential models are fitted between AGB INFyS values and the EO predictors, removing plots with residuals exceeding twice the residual standard deviation (Cartus, Kellndorfer et al. 2014). In our case around 4% of the plots were excluded: 225 plots in total out of more than 5000. By class: 144 (class 0-50 t/ha), 46 (class 50-100), and 35 (class > 100 t/ha). Plots in areas with slopes > 25% are also excluded. The maps are trained using the plot data for the whole region of interest.

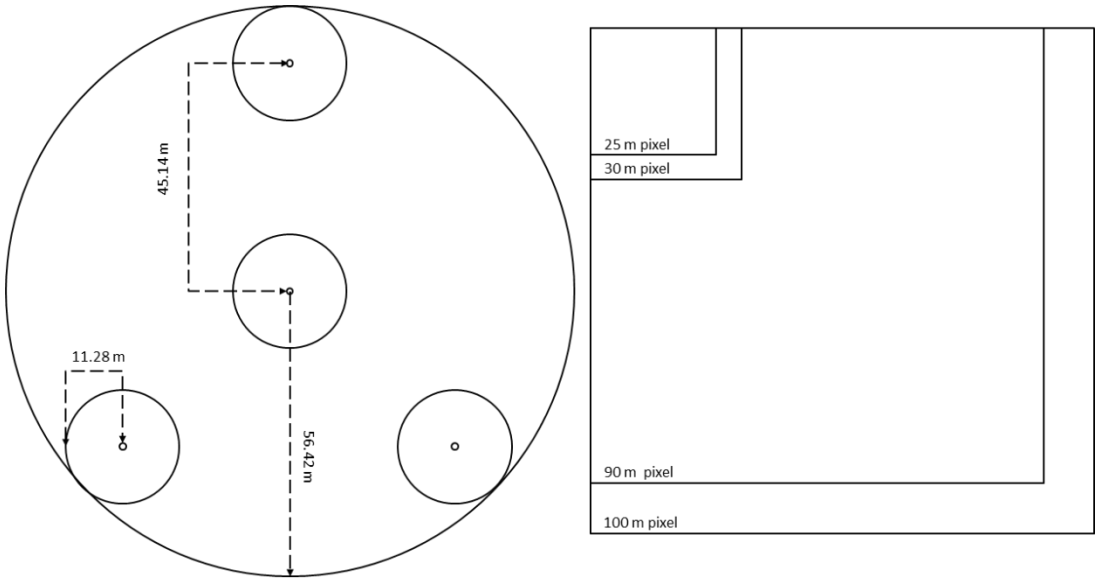


Figure 3-18. left) Mexican INFyS sample plot consisting of 1 ha primary unit and four 400 m^2 sub-units (0.16 ha sampled area); right) Different pixels sizes at the same scale as the INFyS plot (25 m, 30 m, 90 m, and 100 m).

The sampling distance between centres of the plots is 5 x 5 km in forest areas and mangroves, and 10 x 10 km in dry tropical forest and semi-arid vegetation.

3.4.2.3 *Validation data*

An independent validation dataset was generated by setting aside 10-30% of the INFyS dataset after stratifying the data into major biomass classes. This dataset is not used for training purposes.

3.4.2.4 *Data used to quantify accuracy*

The AGB probability distribution generated from the MaxEnt algorithm and ancillary data are used to quantify the uncertainty based on the error propagation approach (see Methods section).

3.4.2.5 *Data used for accuracy assessment*

The independent INFyS validation data are used for accuracy assessment.

3.4.3 Methods

3.4.3.1 *Pre-processing of data*

A collection of USGS Landsat Surface Reflectance imagery computed by the LEDAPS method (Masek, Vermote et al. 2006) is accessible via GEE. Multi-temporal 50 percentile composites are generated for the study areas for each epoch. For the 2010 epoch, scenes from the period 2010±1 are used. The Landsat QA layers are used to exclude pixels with cloud cover, snow and shadow.

K&C ALOS 25m mosaics include a standard SAR pre-processing as explained in Shimada, Itoh et al. (2014). This involves calibration, multi-looking (output of 16 looks), projection, ortho-rectification and slope correction using SRTM DEM. A destriping process (Shimada and Isoguchi 2002) is also applied to equalize the intensity differences between neighbouring strips. The differences between strips were attributed largely to seasonal and daily differences in surface moisture conditions. If significant effects are still encountered in a strip of the study area, substitution of the strip for one from another period and histogram matching of new strip with neighbouring strips will be attempted. A multichannel filter (Quegan and Yu 2001) is applied using multi-temporal data from 2007, 2008, 2009, and 2010 to reduce speckle. After testing several moving window sizes (i.e. 3x3, 5x5, 7x7, 9x9, 11x11) a 7x7 window was selected as optimal. At this point, the level of speckle was considered acceptable, as no increase in window size or additional spatial filtering showed further improvement.

Landsat PTC product (Hansen, Potapov et al. 2013) corresponds to the year 2000. The forest cover loss layer included in the product indicating the annual loss of tree cover from 2000-2015 (pixels that become 0% tree cover) was used to genera PTC products for 2005, 2010, and 2015. This approach however assumes no forest cover gain over the study area.

3.4.3.2 *Biomass estimation algorithms*

A total of 339 biomass allometric equations (http://www.mrv.mx/index.php/en/mrv-m-3/work-areas/allometric-modells) and 214 species-specific wood densities were used by CONAFOR (CONAFOR, 2012) to estimate tree-level AGB following a protocol for allometric model selection which prioritises the use of species-specific models and wood densities within their diameter range of applicability. If more than one equation is available the equation with highest R^2 or regional (closest spatial location) is selected. If no species-specific model is available, the same procedure is followed at higher levels (genus and forest type). Due to the lack of species- or genus-specific allometries for all tree species,

generalized models are used (Chave et al., 2005, Brown, 1997) for approximately half of the plots in the INFyS database (Cartus et al., 2014). Most plots have been measured twice during the period 2004-2012, having 2 estimations of AGB per plot. The dataset includes information on forest parameters such as AGB, tree density, crown cover, basal area, mean canopy height and forest types.

3.4.3.3 *Training methods*

The MaxEnt algorithm is a flexible and general purpose non-parametric algorithm that estimates the probability distribution with the maximum entropy subject to the constraints established by the input information (Phillips, Anderson et al. 2006). The MaxEnt algorithm is widely used for estimating species distribution models (SDM), and has been recently used for classifying remote sensing data (Li and Guo 2010, Saatchi, Harris et al. 2011).

Jaynes (1957) postulated that a distribution which agrees with everything that is known and avoids any assumptions not supported by a priori information should have maximum entropy. The maximum entropy distribution is the most widespread or the closest to the uniform distribution. The unknown probability distribution π is defined over the finite space *X* (here the values of the pixels in our study area). The probability distribution π gives a non-negative probability $\pi(x)$ to each individual element of the space *X*, and the sum of these probabilities equals 1. The approximation of π is the probability distribution $\hat{\pi}$ whose entropy is defined as follows (Phillips, Anderson et al. 2006):

$$H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x)\, ln\, \hat{\pi}(x)$$, where *ln* is the natural algorithm *Eq. 3-23*

The constraints on π are represented by a set of known real-valued functions or features ($f_1, \ldots, f_n$) on the space *X.* The information assumed about π is the expectations (approximated by averages) of each feature f_j under π. This feature expectation is defined as:

$$\pi[f_j] = \sum_{x \in X} \pi(x)\, f_i(x)$$ *Eq. 3-24*

This can be approximated using a set of localities ($x_1, \ldots, x_n$) independently drawn from *X* according to π, so that the empirical average is:

$$\tilde{\pi}[f_j] = \frac{1}{m} \sum_{i=1}^{m} f_i(x)$$ *Eq. 3-25*

$\tilde{\pi}[f_j]$ is used as an estimate of $\pi[f_j]$, and the objective is to find the approximate distribution of maximum entropy which satisfies the constraint that each f_i match the same empirical average under $\hat{\pi}$. Based on convex duality (Della Pietra, Della Pietra et al. 1997), this MaxEnt distribution is equal to the maximum likelihood Gibbs distribution.

This regional study uses MaxEnt software 3.3.3k (Phillips, Dud et al. 2004, Phillips, Anderson et al. 2006). The features used to constrain the prior distribution represent the environmental variables (here the EO datasets) or transformations of them. The types of functions that can be used by this version of the algorithm are linear (the variable itself), product (pair-wise product combinations between EO datasets), quadratic (square values of the EO datasets), threshold (functions that allow a step in the fitted function), hinge (allowing a change in the gradient of the response), and categorical

(for discrete EO datasets) (Phillips, Anderson et al. 2006, Elith, Phillips et al. 2011). The algorithm will restrict by default the type of features used according to the number of occurrences available; if this is greater than or equal to 80 all types of features can be used.

The algorithm can produce three different output formats for the model: raw, cumulative and logistic. The logistic output format, which is a post-transformation of the raw output, was used in this regional study. It aims to provide the closest estimate to the probability of the species presence (probability of the class) for each pixel for the given EO datasets. This is scaled from 0 to 1 for the species, 0 being the least suitable and 1 the most suitable. The main purpose of this transformation is to convert the exponential model output of the algorithm into a logistic model, which will prevent the probabilities exceeding the value 1, and makes the model suitable for use with new data or for extrapolation to new areas.

MaxEnt presents similarities with other approaches such as generalized linear models (GLM), generalized additive models (GAM), Bayesian approaches and neural networks (Phillips, Anderson et al. 2006). It has been extensively used in biogeography, conservation biology and ecology in recent years (e.g. Wollan, Bakkestuen et al. 2008, Cordellier and Pfenninger 2009, Kharouba, Algar et al. 2009, Murray-Smith, Brummitt et al. 2009, Saatchi, Harris et al. 2011). The algorithm has been shown to outperform well-established modelling methods such as GLM, GAM, Genetic Algorithm for Rule Set Production (GARP), and BIOCLIM (Elith, Graham et al. 2006, Guisan, Zimmermann et al. 2007), and to have similar performance to other machine learning algorithms (MLA) such as Random Forest (Williams, Seo et al. 2009), One-Class Support Vector Machine (OC-SVM) (Li and Guo 2010), and Boosted Decision Trees (BDT) (Elith, Graham et al. 2006, Guisan, Zimmermann et al. 2007).

The MaxEnt algorithm requires presence-only data as input, as it uses background environmental data for the whole study area, which makes it advantageous in cases of limited training data, such as in AGB studies. This means that the algorithm only needs the locations of the presence of the class (where the species occurs and is observed) for calibration. MaxEnt outperforms other algorithms for small sample sizes (Hernandez, Graham et al. 2006, Pearson, Raxworthy et al. 2007). Continuous and categorical EO datasets, as well as interactions between them (features and functions), can be used as input. As the MaxEnt probability distribution is mathematically well-defined, the relative importance of the EO datasets can be easily analysed. The algorithm also includes a regularization feature to avoid model over-fitting. This refers to the smoothing of the model by making it more regular, so the fitting of a too complex model is avoided. L1-regularization (Tibshirani 1996) is a common approach in model selection (though not used here without further assessment), and trades model fit against model complexity (Elith, Phillips et al. 2011). Several studies (Hastie, Tibshirani et al. 2005, Wollan, Bakkestuen et al. 2008, Elith, Phillips et al. 2011) have found that MaxEnt is reliable and performs well in comparison to other machine learning algorithm, being more stable with correlated variables than, for example, stepwise regression, so there is less need to remove correlated variables, or pre-process covariates by using PCA and selecting dominant axes, which are more likely to degrade the results.

A recent study (Saatchi, Harris et al. 2011) demonstrated the possibility of modelling a continuous biophysical parameter (AGB) by combining the probabilistic outputs generated from a MaxEnt algorithm, and estimating the uncertainty of the estimation on a pixel-by-pixel basis. The study used 3 continental allometric models derived from ground data to relate GLAS-derived canopy height to AGB. Saatchi, Harris et al. (2011) used the AGB estimated from GLAS footprints to calibrate a MaxEnt algorithm, estimating for the first time AGB over the whole tropical region using EO datasets.

Species distribution modelling usually confronts the problem of incomplete input information due to the lack of data. AGB modelling faces the same challenge. Research studies and forest inventories collect data in forested areas, but there are insufficient samples in inaccessible areas such as tropical forests, especially for the highest AGB ranges. This represents a significant challenge for AGB mapping. The Maximum Entropy algorithm is designed to work with incomplete sets of information which make it very suitable for mapping forest parameters. The theoretical framework is shown in Figure 3-19.

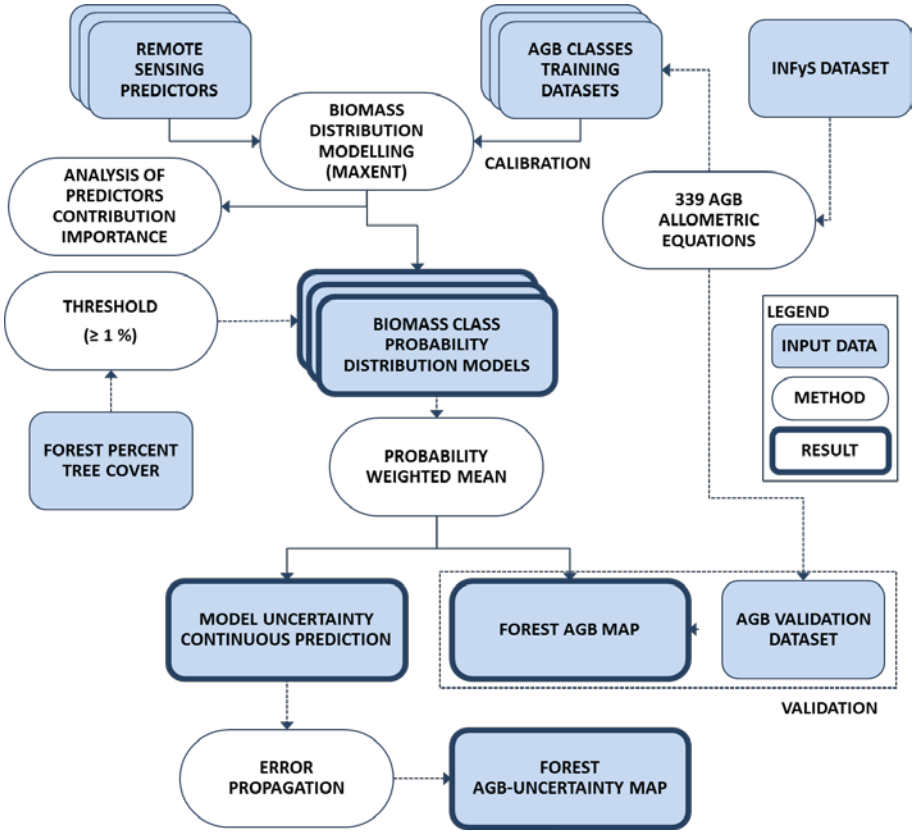


Figure 3-19. Theoretical framework and components of the MaxEnt AGB distribution model approach. EO datasets used as predictors include optical, SAR and topographic datasets (Based on Rodriguez-Veiga 2015).

AGB classes are assumed to behave as different "species" whose distributions are to be constrained by the EO datasets. Henceforth the term "biomass distribution model" (BDM) is used instead of species distribution model. Classifications of remote sensing imagery, as seen in Li and Guo (2010), will be carried out for each AGB class to generate BDMs. As explained before, the probability calculated by the MaxEnt algorithm is equal to the Gibbs probability which is proportional to the conditional probability of the class (here AGB class) (Li and Guo 2010). Over numerous iterations for each AGB class, the weights for combining the EO datasets are adjusted to maximise the average sample likelihood (training gain), and to estimate the distribution over the whole extent of the region. The higher the probability for the pixel, the more suitable the pixel is for representing the same characteristics as the training pixels. In this study, EO datasets commonly used to map vegetation are used to produce several BDMs covering the whole AGB range in the region, instead of the climatic information commonly used in species distribution modelling. Therefore, the occurrences for each AGB class can be represented as localities (x and y map coordinates) sharing the "geographical space" with the EO datasets. The values of the EO datasets are extracted at the localities and used in the

multidimensional "environmental space" to generate the BDMs (Figure 3-20). Therefore, the models are pixel-based and do not include interaction between neighbouring pixels in calibration (Elith and Leathwick 2009, Pearson 2010), but this is already inherent in the models through the spatial autocorrelation of AGB reflected in the remote sensing imagery.

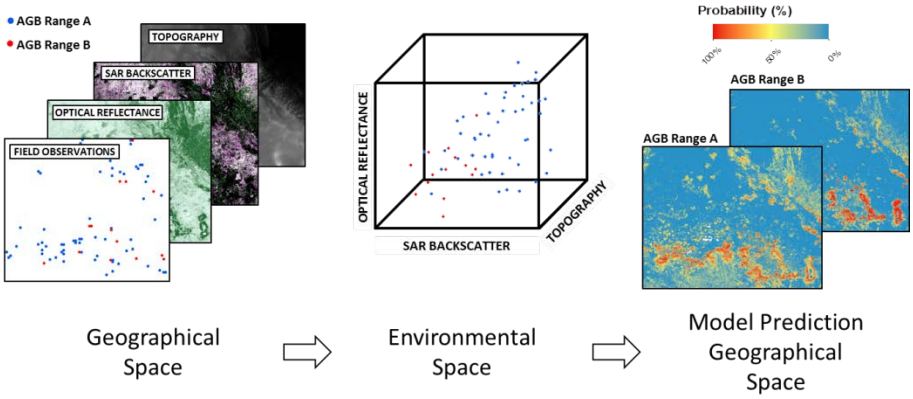


Figure 3-20. Connection between field observations (localities) from two different AGB ranges and earth observation datasets in the geographical space (left), their relationship in the environmental space (centre), and the model predictions (probabilities) in the geographical space (right) (Rodriguez-Veiga 2015).

This approach allows the synergistic use of different types of sensors to scale up the AGB ground measurements. It adapts the method of Saatchi, Harris et al. (2011), which uses the probability logistic outputs from the MaxEnt algorithm to produce AGB and uncertainty maps. AGB occurrence data is obtained from a rigorously designed and extensive ground-based forest inventory over the study area (Section 3.4.3.4).

The approach assumes a finite geographic space *X* formed by a set of discrete grid-cells. A set of points representing recorded values of AGB (occurrences) are the localities of the model (training dataset). The implementation of this probabilistic method requires the set of AGB localities to be classified into AGB classes. The minimum size of the training set to be used for an AGB class is set to be 100, following the suggestion given in Saatchi, Harris et al. (2011), so as to make use of all the features (linear, product, quadratic, threshold, and hinge) available for this version of the algorithm. The training dataset is then divided into AGB classes covering specified intervals (e.g. 0-20 t ha^{-1}, 21-40 t ha^{-1}, …, >200 t ha^{-1}), and each AGB class is assumed to behave as a different "species". The EO layers contain information correlated to AGB, such as optical reflectance, SAR backscatter and elevation (DEM). The AGB classes are used in combination with the set of earth observation datasets defined in the space *X* as inputs to MaxEnt. The aim is then to estimate the probabilistic distribution of each AGB class (BDM) (Figure 3-21).

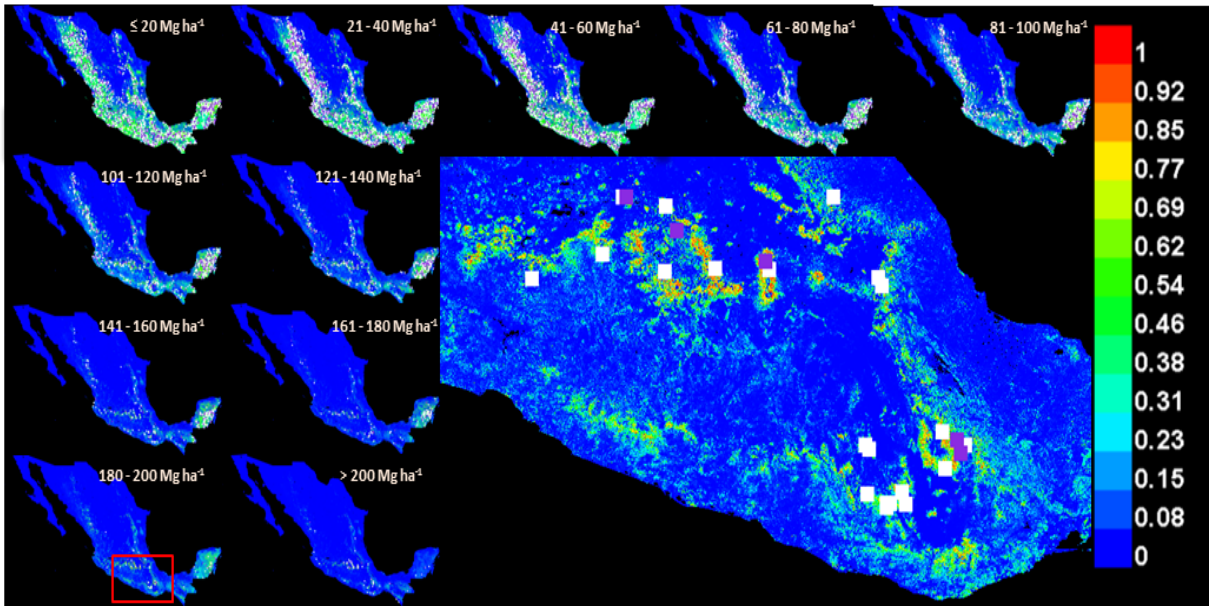


Figure 3-21. Example of BDMs generated by the MaxEnt algorithm with pixels ranging from 0 (least suitable) to 1 (most suitable) for each biomass range. White squares represent training data and violet squares represent test data (Rodriguez-Veiga 2015).

Once the BDMs are generated for each AGB class, the continuous values of AGB for each pixel are calculated as the weighted average AGB per pixel with the probabilities as weights:

$$\widehat{AGB} = \frac{\sum_{i=1}^{N} P_i^n AGB_i}{\sum_{i=1}^{N} P_i^n} \qquad \textit{Eq.}\ 3\text{-}26$$

where $\widehat{AGB}$ is the AGB prediction per pixel and P_i is the probability estimated by MaxEnt for each AGB range AGB_{i} (average value within class *i*). The power of the probability n is used to weight the predicted value towards the maximum probability closest to the true value when other probabilities are small. This acts a bias reduction method. Previous studies such as Saatchi, Harris et al. (2011) used $n = 3$. This study however observed better results with different n values. An optimization of the n value to preserve the skewness in distributions for each pixel and produce the less biased estimates was performed by cross-validation tests.

The AGB density in Mexico can reach above 200 t ha^{-1} (in a very few areas) which exceeds the theoretical saturation level of each individual EO dataset. The combination of different datasets (SAR, optical and topographical information) allows estimation of AGB beyond the individual theoretical saturation level of each sensor as seen in previous studies (e.g. Saatchi, Harris et al. 2011, Mitchard, Saatchi et al. 2012). The different types of data reflect different characteristics of the vegetation (e.g. volume scattering, reflectance, percent tree cover, elevation, slope).

The Landsat PTC layer is used to limit the region which the MaxEnt algorithm uses for background samples and generation of the probabilistic outputs (Elith, Phillips et al. 2011). Therefore, any land area with a forest percent tree cover below 1% has a value of 0 by default. Pixels with permanent and seasonal water are assigned a NoData value based on the JRC Global Surface Water (GSW) product (Pekel, Cottam et al. 2016).

The MaxEnt algorithm performance assessment is made by bootstrapping 25% of the training data. Additionally, Jackknife analyses are performed to select the most suitable input variables to predict AGB and forest distribution. The analyses are based on the Area Under the Receiver Operator Curve (AUC) and Model Gain (Phillips, Dud et al. 2004, Phillips, Anderson et al. 2006), for both training and test data.

The importance of each remote sensing layer is assessed by its relative contribution to the MaxEnt model gain by sensor and per different AGB class. This is done using jackknife analyses with models that omit one EO dataset each time and by models based on single EO datasets. For each remote sensing layer the training data are also randomly permuted and the model is then re-evaluated. Any drop in AUC is normalised to percentage values between 0 and 100%.

Weaknesses and Strengths of MaxEnt

MaxEnt has several possible strengths and weaknesses:

Strengths:

i. The MaxEnt algorithm outperforms or has similar performance to well-established methods such as GLM, GAP, GARP, BIOCLIM, Random Forest, OC-SVM, and BDT (Elith, Graham et al. 2006, Guisan, Zimmermann et al. 2007, Williams, Seo et al. 2009, Li and Guo 2010), especially with limited training data (Hernandez, Graham et al. 2006, Pearson, Raxworthy et al. 2007).
ii. The relative importance of the input variables can be easily calculated for the overall AGB estimation or per biomass range.
iii. One of the main strengths of the approach is that it can generate pixel uncertainty based on the probabilities estimated from the model fitting.

Weaknesses:

i. Phillips, Anderson et al. (2006) noted that it is not a mature statistical method such as GLM. However, other authors argued that MaxEnt is equivalent to logistic regression (Fithian and Hastie 2013, Renner and Warton 2013), whose implementation is well understood.
ii. The training data has to be converted into discrete AGB ranges to generate several BDMs, which are then combined to obtain continuous AGB estimations, so increasing the demand for training data. This step is needed to allow calculation of pixel-level uncertainty. Saatchi, Harris et al. (2011) suggested that at least 100 plots should be used as training data for each AGB range. This restricts the use of Maxent in areas with lack of data. The model can be trained with data from different areas but this should be done with caution, as the correlation of AGB to different EO datasets might exhibit regional variations due to factors such as forest structure, species composition and wood density, allometry, atmospheric effects, and vegetation moisture.
iii. If needed, MaxEnt can deal with overfitting using a regularization parameter (L1-regularization) to reduce model complexity. However, the amount of regularization to smooth the model is not clear. Smaller values than the default value of 1 might overfit the model and result in a closer fit to the training data, while higher values will result in distributions with higher dispersion. The default regularization value (no regularization) will initially be used and then optimized as required.

An exponential model is used to estimate the MaxEnt probabilities. As this model is not bounded above, using the model to extrapolate to regions out of the range of the EO layers in the study area can be problematic. This can be averted by using MaxEnt's logistic output which transforms the exponential model to a logistic model.

Mangroves and wetlands

Due to the effect that soil moisture and water level occurring on mangrove and other wetland vegetation have on the signal of SAR backscatter and optical reflectance, as well as the limited amount of ground data for those vegetation types for calibration purposes, a different method was used on those areas. Empirical linear regression models with AGB as dependant variable (Ln-transformed) and SAR backscatter intensity (DN values) of the HV polarization as the independent variable were fitted for each epoch on those areas (Figure 3-21b). As the SAR backscatter sometimes suffer from double-bouncing effects in this kind of vegetation deriving in very high intensity response, the models were also capped from above limiting the AGB estimation to a maximum of 200 t ha^{-1}. The mangroves and wetland vegetation are delimited using the classes "Manglar", "Popal", and "Tular" (in Spanish) from the Land Use and Vegetation (LUV) layer from INEGI.

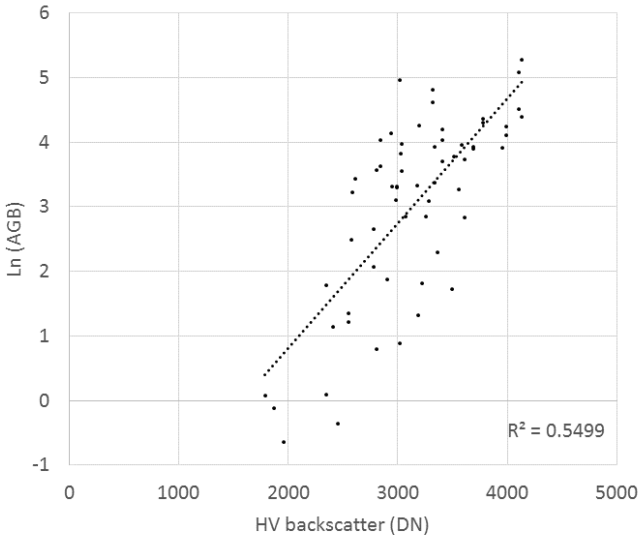


Figure 3-21b. Linear regression between AGB (Ln-transformed) and HV backscatter in mangroves for the epoch 2015.

3.4.3.4 *Methods to assign accuracies in the uncertainty map*

The uncertainty of the AGB prediction ($\varepsilon_{prediction}$) is calculated from the RMS error ($\sigma_{\widehat{AGB}}$) per pixel. The following equations are used (Saatchi, Harris et al. 2011):

$$\varepsilon_{prediction} = \sigma_{\widehat{AGB}} / \widehat{AGB}\, x100 \qquad \textit{Eq. 3-27}$$

$$\sigma_{\widehat{AGB}} = \sqrt{\frac{\sum_{i=1}^{N} (AGB_i - \widehat{AGB})^2\, P_i}{\sum_{i=1}^{N} P_i}} \qquad \textit{Eq. 3-28}$$

The total uncertainty at pixel level is composed of different sources of error which are assumed to be random and independent. These are propagated using the following equation proposed by Saatchi, Harris et al. (2011):

$$\varepsilon_{AGB} = (\varepsilon^2_{measurement} + \varepsilon^2_{allometry} + \varepsilon^2_{sampling} + \varepsilon^2_{prediction})^{1/2}, \qquad \textit{Eq. 3-29}$$

The values of these errors are estimated from the prediction probabilities of the MaxEnt and from current literature (Chave, Condit et al. 2004, Mitchard, Saatchi et al. 2011, Saatchi, Harris et al. 2011, Weisbin, Lincoln et al. 2014). The error sources used are the following:

$\varepsilon_{measurement}$: The measurement error of tree level parameters such as diameter and tree height averaged at plot level (Chave, Condit et al. 2004). This was assumed to be 10% in this study (Mitchard, Saatchi et al. 2011).

$\varepsilon_{allometry}$: The error in estimating AGB using allometric equations. Depending on the allometry used this varies from 11% to 22% (Chave, Condit et al. (2004).

$\varepsilon_{sampling\ size}$: This originates from the variability of AGB within the pixel area and depends on the size of the plots used to upscale the AGB measurements to the pixel level. It is approximated using data from Chave, Condit et al. (2003) on the AGB variability of a 50 ha plot. Using the sampling size equation for a 95% confidence interval, Chave, Condit et al. (2003) found that a minimum of 160 plots of 0.04 ha is needed to estimate the biomass of a 50 ha plot with ±10% uncertainty. This means that a sampling intensity of 12.8% (0.04x160/50) is needed. By assuming the same variations of AGB in the 1 ha INFyS primary unit (see Training data section), the number of 0.04 ha subplots needed to reach the same sampling intensity will be 3.2. Each 1 ha primary unit used for pixel calibration uses four 0.04 ha subplots. Thus, the uncertainty of the AGB estimation will decrease to approximately 8.4% ($10 \times \sqrt{3.2/4}$).

$\varepsilon_{prediction}$: This is calculated for each pixel from the prediction probabilities of the MaxEnt model. It also accounts for the representativeness of the sampling sites of the true distribution of AGB in the region (Saatchi, Harris et al. 2011). In the case of mangroves and wetland vegetation, the prediction error is based on errors estimated from the regression model.

3.4.3.5 *Methods to test the accuracy of the estimates*

The validation data are used for accuracy assessment based on the recommendation from the GlobBiomass Validation protocol (D5 document).

3.4.4 Products

The proposed AGB map for Mexico aims to improve on previous studies using the most recent EO datasets and state-of-the-art methods. Current pan-tropical maps (Saatchi, Harris et al. 2011, Baccini, Goetz et al. 2012) have pixel spacing of 500m and 1000m respectively, but the proposed AGB map has pixel spacing of 25 m, while the AGB map of Mexico due to Cartus, Kellndorfer et al. (2014) has a pixel spacing of 30 m (see Table 3-5). Recent studies suggest that optical imagery (Landsat, MODIS) is correlated with AGB beyond the theoretical saturation due to canopy closure, especially in the infrared bands (Kellndorfer, Walker et al. 2011, Baccini, Goetz et al. 2012) which are sensitive to shadowing and moisture differences. These bands were not used by Cartus, Kellndorfer et al. (2014) but are used in the current map. The 30 m posting SRTM DEM is used instead of the 90 m used by Cartus, Kellndorfer et al. (2014), which may lead to better predictions of AGB in mountainous regions. We produce not only an AGB map, but also an Uncertainty map.

3.4.5 Modifications for the 2005 and 2015 epochs

Methods

The method used is the same for all epochs

Data

2005: Landsat 7 SR 2004-2006, ALOS PALSAR mosaics for 2007-08, Hansen's Landsat percent tree cover, and 30m SRTM.

2010: Landsat 7 SR 2009-2011, ALOS PALSAR mosaics for 2009-10, Hansen's Landsat percent tree cover and 30m SRTM.

2015: Landsat-8 SR 2014-16, ALOS-2 PALSAR-2 mosaics for 2015-16, Sentinel-1 2015-16, Hansen's Landsat percent tree cover and 30m SRTM.

3.5 South Africa

3.5.1 General description of the region

The South African regional site (Figure 3-22) covers 333 500 km^2 and is situated along a 1300 km North-South transect with a large precipitation gradient and various types of soil, such as granitic and basaltic soils. It encompasses the eastern forest belt of the country, including the eastern part of the Mpumalanga province, the Limpopo Provinces to the north-east, KwaZulu-Natal and a large part of the Eastern Cape Province to the south. It contains various forest types: savannas with AGB mostly below 100 t ha^{-1}, which forms the main ecosystem (about 68% of the area), forest plantations stretching over hundreds of km with various biomass classes of pine and eucalyptus, indigenous dense forests, and coastal forests such as thickets. It also includes the Kruger National Park, which is very well documented in terms of disturbances, such as fire and herbivory (in particular elephant browsing).

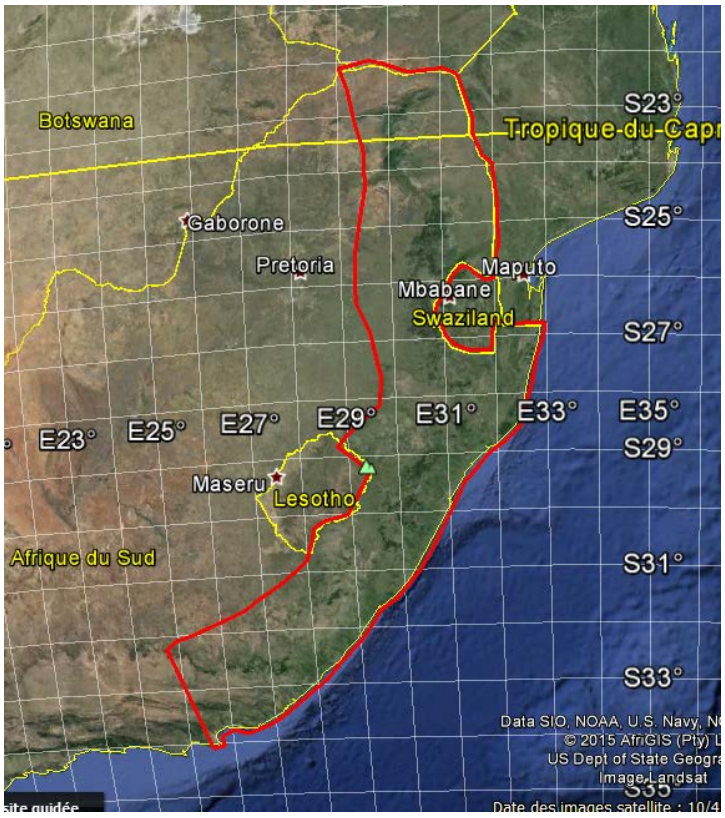


Figure 3-22. The South African 333 500 km^2 regional site in red.

3.5.2 Datasets

3.5.2.1 *Input data (used to drive the biomass algorithm)*

Remote sensing imagery sensitive to forest AGB from different satellite sensors (optical and SAR) will be used (Table 3-7). The core datasets will be L-band ALOS PALSAR and ALOS-2 PALSAR-2. The reason for using ALOS data instead of ASAR or Sentinel-1 data is that the sensitivity of L-band SAR data to biomass is better than C-band SAR data. However, the use of C-band data is not excluded for research purposes, especially for the 2015 epoch when Sentinel-1 time series are available. Freely available ALOS PALSAR mosaics of HH and HV γ^0 with 25 m pixel size were produced by JAXA in 2007, 2008, 2009 and 2010 (http://www.eorc.jaxa.jp/ ALOS/en/palsar_fnf/fnf_index.htm). PALSAR-2 data from 2014 will first be supplied by JAXA through the K&C initiative, before the free distribution of PALSAR-2 mosaics (HH and HV). Mosaic data are provided as tiles of 112 km × 112 km, or 1° of latitude and 1° of

longitude. Approximately 40 tiles are needed to cover the whole regional site. Mosaics are pre-processed by JAXA, including ortho-rectification, slope correction and radiometric calibration between neighbouring strips. Additional processing applied on the mosaic data by CESBIO ensures the suitability of mosaics data for biomass estimation. Using FBD data would require far more processing work from CESBIO and is not expected to provide better results as the mosaics already consist of dry-season images in this area, which are optimal for AGB retrieval.

Freely available Landsat tree cover continuous field with 30 m resolution (a combination of Landsat TM and Landsat ETM+) will also be acquired (from http://landcover.org/data/landsatTreecover/) Additionally, freely available 30 m posting elevation data from the Shuttle Radar Topography Mission will be obtained from EarthExplorer (earthexplorer.usgs.gov).

Table 3-7. Available remote sensing products and associated acquisition dates

DATASET	SPATIAL RESOLUTION	2000	2005	2010	2015
ALOS PALSAR archived mosaics (HH, HV)	25 m		x	x	
ALOS-2 PALSAR-2 FBD and mosaics (HH, HV)	25 m				x
Landsat products	30 m	x	x		
Shuttle Radar Topography Mission-SRTM(1)	30 m	x			

(1)Elevation data from SRTM corresponds to the year 2000. As it is assumed that topography remains constant, the same dataset will be used for all epochs

3.5.2.2 *Training data*

In situ AGB for 37 1-ha plots measured by CSIR in 2012 in the Kruger National Park's savannas and woodlands will be used for AGB mapping for the 2010 epoch. For the 2005 epoch, a high resolution airborne Lidar transect acquired in 2004, which contains about 38000 samples that have been transformed to AGB, were used. For the 2015 epoch, CSIR has measured 56 new 1-ha biomass plots between 2014 and 2016 in different biomes: 31 plots in savannas, 17 plots in pine and eucalyptus plantations, and 7 plots in indigenous forests. 50% of the *in situ* data for each epoch will be selected using stratified random sampling to train the algorithm.

3.5.2.3 *Validation data*

The remaining 50% of the *in situ* data will be made available for validation.

3.5.2.4 *Data used for accuracy assessment*

The validation data are used for accuracy assessment by CESBIO. In addition, Lidar data available from CSIR will be used to validate the biomass maps. These Lidar data are an independent data source and can provide reliable biomass estimation locally.

3.5.3 Methods

A method has been developed to map AGB using L-band SAR data in woody ecosystems with AGB less than 100 t ha^{-1}. It was first developed using PALSAR data in Fine Beam Dual Polarisation mode (FBD) at a few test sites, then extended to country scale using PALSAR mosaic data to map AGB in the Cameroon savanna (Mermoz et al., 2014). The method was derived from analysis of PALSAR data as a function of AGB, using a reliable dataset of *in situ* AGB estimates from 41 plots, each of about 1 ha.

The method consists of a pixel-based inversion of the pre-processed SAR data into AGB. The development of the inverse model was based on experimental results interpreted using the theoretical MIPERS model (Villard, 2009). The theoretical model was used to simulate the impacts of forest structure, environmental conditions and SAR parameters on the relationship between L-band SAR data and the AGB of the forest under study. Model simulations indicated that for given SAR specifications (frequency, polarisation and incidence), and for a given forest structure and soil and vegetation moisture conditions, the relationship between L-band backscatter and AGB can be approximated by a regression equation, which is found to closely follow the simplified water cloud model formulation. whose parameters are determined using experimental data.

The parameters of this equation are determined using experimental and /or ancillary data. In the present study, the SAR parameters are restricted in terms of frequency (L-band), polarization (HH and HV), and range of incidence angle (34-40°). The effect of canopy structure is expected to be reduced in the woody savanna, the main forest ecosystem under study. The other forest type in the study site, which is forest plantation, will be taken into account. For the other types, which form approximately 30% in area, they will in a later phase of the project, when *in situ* plots in the plantations will be available for the analysis of the relationship between backscatter and AGB. The PALSAR data forming the mosaic to be used for 2007 (epoch 1 with a two-year shift) and 2010 (epoch 2) were acquired during the dry season (May to July), minimizing the effect of varying soil and vegetation moisture.

Figure 3-23 shows the AGB mapping methodology comprising the following steps: 1) pre-processing of SAR data to reduce uncertainties due to speckle while preserving the SAR resolution,; 2) establishing the direct and inverse statistical regressions relating radar backscatter to AGB; 3) pixel-based mapping of AGB and its uncertainty using a Bayes inversion approach.

3.5.4 Pre-processing of data

Pre-processing of SAR data includes slope correction, multi-image filtering, temporal image inter-calibration and geocoding. Mosaic data are slope-corrected by JAXA as described in Shimada et al. (2010). Multi-image filtering (Bruniquel and Lopes, 1997; Quegan and Yu, 2001) is applied to reduce speckle while preserving the fine structure in the image, which is essential to detect small areas that change between epochs. The filter requires the estimation of local mean intensities by averaging intensity values in a local window around each pixel in each image. The window size is commonly 7x7 but is likely to change, depending on the final uncertainty to be obtained. For the biomass change estimation, it is necessary to further minimise temporal changes caused by environmental conditions or the SAR radiometric stability. To do so, the multi-temporal intensity data were normalised to have approximately the same temporal mean and variance over the 'forest pixels' using the methodology of Du et al. (2001). In this method, pixels at the same location in two different scenes form corresponding pixel pairs. A selection of dense forest pixel pairs is performed, avoiding the obviously changed (i.e., deforested) ones. These pixels are used to compute an inter-annual backscatter.

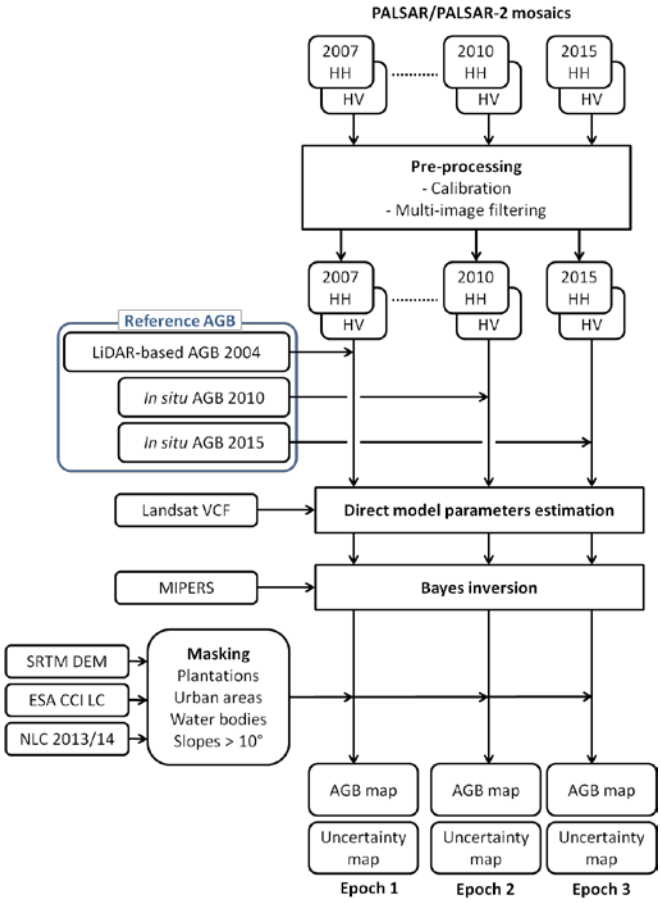


Figure 3-23: Flowchart of the methodology to map biomass using L-band SAR data.

3.5.4.1 *Biomass estimation algorithms*

For both HH and HV, the model for L-band indicates that the backscatter at low AGB (e.g. < 100 t ha^{-1}) is dominated by ground scattering, and at AGB > 100 t ha^{-1}, by volume scattering. In between, the backscatter is a sum of volume scattering and ground scattering, attenuated by the canopy.

This can be expressed by a simplified formulation of the well-known Water Cloud Model (WCM) (note that this treats the vegetation as a uniform slab of material, so has limitations in representing the considerable amounts of bare or sparsely vegetated ground which occur in savanna):

$$\gamma^0 = a\, e^{-c.AGB} + b(1-e^{-c.AGB}) \qquad \text{[Eq. 3-30]}$$

where:

- $a\, e^{-c.AGB}$ is the underlying ground scattering, attenuated by the forest layer;
- $b\,(1-e^{-c.AGB})$ is the vegetation scattering;
- c is the attenuation coefficient of the vegetation layer, which depends on vegetation structure and water content.

a is the unattenuated backscatter from bare ground (and gaps in the forest) and varies as a function of soil moisture, surface roughness, topography, and to a lesser extent at L-band, herbaceous vegetation cover; b is the backscatter value at saturation. The backscatter lying between a and b is a sum of the attenuated ground contribution and the contribution from the forest canopy.

The biomass mapping method relies on inverting regression models based on data at a single polarisation (HV) or dual polarisation (HH, HV). HV backscatter, which is from volume scattering, is less sensitive to topography, and for low AGB (roughly below 20 t ha^{-1}), more sensitive to variations of AGB than HH, which is dominated by soil scattering. However, very low backscatter close to the noise equivalent sigma-zero (e.g. approximately -32 dB for HH and -34 dB for HV) can be seriously affected by system noise. On the other hand, HH can have higher AGB saturation level than HV, because the backscatter comes from both volume and ground scattering. However, the relationship is site- and time-dependent. The single HV polarisation approach is expected to be more robust because HV is less sensitive to ground effects, but the single HH polarisation approach could extend the AGB saturation region. Combining HH and HV polarisations has been shown to increase the retrieval performance in this study. This could be achieved by a multivariate regression, but the results would be highly dependent on the *in situ* plot dataset. We therefore use a method that takes into account the joint probability density function of HH and HV, by formulating the inversion problem in a Bayesian form (Tarantola et al., 2005).

In order to minimize the error propagation which would result from the direct use of Equation 3-29 to convert γ^0 into AGB, a Bayesian approach is preferred, as proposed in previous studies (Notarnicola and Posa, 2004; Tarantola, 2005). Note that in the following, we use simplified expressions for the probability density functions, where $\gamma^0_{HH} = \gamma^0_{HH\ obs}$ is replaced by $\gamma^0_{HH\ obs}$, $\gamma^0_{HV} = \gamma^0_{HV\ obs}$ is replaced by $\gamma^0_{HV\ obs}$, and $\mathrm{AGB} = \mathrm{B}$ is replaced by B.

The Minimum Mean Square Error (MMSE) estimator of AGB is given by the conditional expectancy (Couhert, 2009):

$$\mathrm{AGB}_{estim} = E[\mathrm{AGB}|\gamma^0_{HH\ obs}\ ,\gamma^0_{HV\ obs}] = \int_0^{AGB_{max}} \mathrm{B}.p(\mathrm{B}|\gamma^0_{HH\ obs}\ ,\gamma^0_{HV\ obs}).dB \qquad \text{[Eq. 3-31]}$$

The inversion therefore requires the calculation of the posterior probability $p\left(\mathrm{B}|\gamma^0_{HH\ obs}\ ,\ \gamma^0_{HV\ obs}\right)$ of AGB given the observation $\left(\gamma^0_{HH\ obs}, \gamma^0_{HV\ obs}\right)$, which is obtained using Bayes theorem:

$$p\left(B|\gamma^0_{HH\ obs}\ ,\ \gamma^0_{HV\ obs}\right) = \frac{p(\gamma^0_{HH\ obs}\ ,\gamma^0_{HV\ obs}|B)p(B)}{p(\gamma^0_{HH\ obs}\ ,\gamma^0_{HV\ obs})} \qquad \text{[Eq. 3-32]}$$

Three terms need to be quantified. The marginal likelihood $p(\gamma^0_{HH\ obs}\ ,\ \gamma^0_{HV\ obs})$ is constant in the scene and is therefore neglected, since the expression will be normalized to ensure the integrated probability equals 1. The prior probability, $p(\mathrm{B})$, corresponds to the distribution of AGB in the scene. This distribution is unknown in most cases, and its probability density function is therefore considered uniform over the [0 AGB_{max}] range, where AGB_{max} represents the assumed highest biomass found in the savannahs and woodlands of South Africa. By analysing the distribution of AGB obtained from existing AGB maps across the areas of savannahs and woodlands identified in South Africa from the ESA Climate Change Initiative (CCI) Land Cover 2010 map (http://www.esa-landcover-cci.org/), we found that a reasonable estimate of AGB_{max} ranges between 100 and 150 $Mg.ha^{-1}$.

The likelihood function $p\left(\gamma^0_{HH\ obs}\ ,\ \gamma^0_{HV\ obs}|\ \mathrm{B}\right)$ accounts for the dispersions of HH and HV backscatter caused by environmental conditions and forest structure, for a given AGB value. These dispersions are considered independent in HH and HV (simulations with MIPERS indicate that the inter-dependence can be significant if we constrain some of the forest descriptors for a given AGB, but becomes negligible if we consider the whole set of possible realizations (forest structure, ground topography, soil

moisture). The joint conditional probability density function can thus be expressed as the product of the individual conditional probability density functions:

$$p\left(\gamma^0_{HH\ obs}\ ,\ \gamma^0_{HV\ obs}|\mathrm{B}\right) = p\left(\gamma^0_{HH\ obs}\,|\mathrm{B}\right)\cdot p\left(\gamma^0_{HV\ obs}\,|\mathrm{B}\right) \qquad \text{[Eq. 3-33]}$$

For a given biomass value B, the estimation of the probability density function $p\left(\gamma^0_{obs}\,|\mathrm{B}\right)$ for each polarization requires two steps.

First, we estimate the theoretical backscatter γ^0_{theo} through inversion of the water cloud model. This represents the expected backscatter values for the environmental conditions (soil moisture, vegetation structure) prevailing for the reference data. We then express the probability that the observed backscatter γ^o_{obs} deviates from this theoretical backscatter γ^0_{theo} because of local environmental effects. We model the probability of γ^o_{obs} using a Gaussian distribution characterised by a mean value equal to γ^o_{theo}, and a standard deviation σ_{simu} estimated by simulation using the MIPERS (Multistatic Interferometric Polarimetric Electro-magnetic model for Remote-Sensing) model (Villard, 2009). Like other EM models, MIPERS has a limited capacity to predict absolute values, and is much more relevant for relative trends (temporal dynamics, polarisation ratios, etc.). Hence it is not used here to predict the backscattering coefficient, only its variability.

This σ_{simu} term models the variability of the backscatter due to environmental conditions or forest structure. The forest growth model used to feed the geometrical parameters to the MIPERS model is fully described in Mermoz et al. (2015) and is calibrated using the available *in situ* data. Assumptions are made to define the range of values of other required MIPERS inputs, such as soil and vegetation moisture, soil roughness etc. We vary the MIPERS inputs and propagate these variations by Monte Carlo simulations to obtain the standard deviation of the backscatter, denoted σ_{simu} . Note that σ_{simu} varies with AGB: higher values are associated with low AGB (the backscatter being very dependent on the environmental conditions) and lower values are associated with high AGB (the backscatter is more stable over dense forests than over bare soils). For a given γ^o_{obs}, the likelihood function $p\left(\gamma^0_{obs}|\mathrm{B}\right)$ is therefore estimated for each polarisation as follows:

$$p\left(\gamma^0_{obs}|B\right) = \frac{1}{\sqrt{2\pi}\,\sigma_{simu}(B)}\,e^{-\frac{1}{2}\left(\frac{\gamma^0_{obs}-\gamma^0_{theo}(B)}{\sigma_{simu}(B)}\right)^2} \qquad \text{[Eq. 3-34]}$$

Combining Equations 3-31 to 3-33 leads to the following expression for the posterior probability:

$$p\left(\mathrm{B}|\gamma^0_{HH\ obs}\ ,\gamma^0_{HV\ obs}\right) = K\ \frac{e^{-\frac{1}{2}\left(\frac{\gamma^0_{HH\ obs}-\gamma^0_{HH\ theo(B)}}{\sigma_{HH\ simu}(B)}\right)^2}}{\sigma_{HH\ simu}(B)}\ \frac{e^{-\frac{1}{2}\left(\frac{\gamma^0_{HV\ obs}-\gamma^0_{HV\ theo(B)}}{\sigma_{HV\ simu}(B)}\right)^2}}{\sigma_{HV\ simu}(B)} \qquad \text{[Eq. 3-35]}$$

where K is a normalization factor.

The mode of the posterior probability density function described in Equation Eq. 3-34 gives the maximum likelihood estimate of AGB, i.e. the most likely value of AGB for a given observation and for a uniform distribution of AGB, but the MMSE estimator given in Equation Eq. 3-31 is preferred to minimize the error.

3.5.4.2 *Training methods*

Three parameters, *a*, *b*, and *c*, are required for equation [3-30]. Different approaches can be used to estimate them:

i) Use of statistical regressions between the SAR data and *in situ* plot data, provided that the latter are of sufficient number and distributed over the relevant range of AGB and plot conditions.

ii) When the *in situ* plot data do not meet these requirements, *a* and *b* can be estimated from the SAR data, or using ancillary data.. Parameter *a is e*xtracted from pixels corresponding to bare *g*round and *b* from pixels corresponding to closed forests:

- Using ancillary data such as Landsat tree cover continuous field from Sexton et al., (2013). Mean values of backscatter from pixels corresponding to 0 and 100% cover are used as proxy for *a* and *b* by Santoro et al. (2011).
- Using a histogram of the backscatter of the SAR scene over forest land (where water, steep slope and manmade targets are masked out). Parameters *a* and *b* could be derived from the low and high end of the distribution (e.g. the 5th and 95th percentiles, to be defined after testing).

Parameter *c*, which represents the vegetation attenuation coefficient, varies as a function of vegetation water content and vegetation structure (vertical and horizontal distribution of scatterers, number of stems per hectare, etc.). Hence *c* changes with forest type, and must be derived using *in situ* AGB plot data.

In this study, the reference AGB data (from in situ measurements or from Lidar, depending on the epoch) are not distributed sufficiently well to use approach i). We will therefore use approach ii), using ancillary data for estimating *b*. For the 2015 epoch, the *in situ* dataset may be representative enough to use approach i). We will also test approach ii) to compare the performance of both approaches. The approaches will be evaluated in terms of AGB uncertainty assessment.

3.5.4.3 *Methods to assign accuracies in the uncertainty map*

The uncertainties in *in situ* AGB and the SAR measurements, together with errors due to retrieval methods, are taken into account in quantifying the overall accuracy of the retrieved AGB.

In situ AGB estimates are affected by a number of error sources, including most importantly field measurement errors (diameter at breast height, tree height), allometric models, and discrepancy between the sampled area and the pixel coverage (Chave et al., 2004). An error of 10% is associated with the wood density and an error of 5% with the allometric equations (Chave et al., 2004), while the errors related to diameter at breast height and tree height measurements are estimated to be 2.25% and 4.47% respectively in a previous study (Mermoz et al., 2014). Similarly to what was done in this study, we propagate these errors by Monte Carlo simulations to yield a standard deviation associated with plot-based AGB estimates. This approach yields a mean field measurement error of σ_M = 9.7% (standard deviation of the in situ AGB linked to measurement) for the 37 1-ha plots measured in 2012 in South Africa. The sampling errors σ_S (standard deviation of the in situ AGB linked to sampling) associated with the mismatch in spatial scales between the field plots and the pixel size is estimated for each plot using Figure S10 from Réjou-Méchain et al. (2014), with a 9% mean error. The field measurement error σ_M and sampling error σ_S are then combined by adding the associated variances, to obtain the overall field data SD at the plot level:

$$\sigma_{AGB\ Field} = \sqrt{\sigma_M^2 + \sigma_S^2} \qquad \text{[Eq. 3-36]}$$

The estimation of radar backscatter at the pixel level is affected by two main error sources: the radiometric accuracy and speckle. Speckle is modelled as a multiplicative noise and depends on the ENL. The ENL is defined as μ^2/σ^2, where μ and σ are the mean value and standard deviation of the backscatter. With an ENL of 112, the standard deviation caused by speckle σ_{SP} is almost constant in a dB scale and equal to approximately 0.45 dB. The radiometric accuracy can be split into the absolute radiometric bias and the radiometric stability. Shimada et al. (2009) estimated the overall radiometric accuracy σ_{RA} to be 0.76 dB using corner reflectors and 0.22 dB from analysis of Amazon forest data; the latter could be referred to as radiometric stability. In our study, we use a radiometric accuracy of 0.5 dB. Under the assumption of independence between the aforementioned sources of errors, the overall SAR backscatter SD can again be estimated using a quadratic sum of variances:

$$\sigma_{SAR} = \sqrt{\sigma_{SP}^2 + \sigma_{RA}^2} \qquad \text{(Eq. 1.37)}$$

The resulting σ_{SAR} value is 0.64 dB.

Assuming that uncertainties $\sigma_{AGB\ Field}$ and σ_{SAR} are independent, their variances can be added, leading to the resulting total error of the AGB estimate uncertainty.

3.5.5 Products

The final products are AGB and AGB accuracy maps at 25m resolution in 2007, 2010 and 2015, and change maps between 2007, 2010 and 2015.

3.5.6 Modifications for the 2005 and 2015 epochs

Methods

The same method will be used for all epochs.

Data

Input data

Table 3-8. Available remote sensing products and associated acquisition dates

DATASET	SPATIAL RESOLUTION	2000	2005	2010	2015
ALOS PALSAR archived mosaics (HH, HV)	25 m		x	x	
ALOS-2 PALSAR-2 mosaics (HH, HV)	25 m				x
Landsat products	30 m	x	x		
Shuttle Radar Topography Mission-SRTM(1)	30 m	x			

Reference Data

The reference data for the 3 epochs are described in Section 3.5.2.2.

4 Commonalities and differences between the regional approaches

All the regional approaches are based on PALSAR data, although Sweden already has an operational approach to national biomass mapping based on optical and lidar data, and associated national maps for 2005 and 2010. The methods are essentially of two types, data-driven (Poland, Indonesia, Mexico) and model-driven (Sweden with PALSAR data, South Africa). The data-driven approaches rely on substantial amounts of *in situ* data to allow regression methods or, in the case of Mexico, Maxent methods to be developed. The two model-driven approaches both use a simplified version of the Water Cloud model, for which auxiliary datasets of forest cover and forest density are essential in order to estimate parameters. The main methodological and data aspects of the different approaches are given in Table 4-1.

Table 4-1. The primary methods and data requirements of the regional approaches, together with comments on issues concerning high biomass forests

Region	Primary Data	Primary Method	High Biomass	Ground Data
Poland	PALSAR 25m mosaics HH+HV (2009+2010), additionally ERS-2 + ASAR	Random Forest; the whole country as an analysis unit	An implicit assumption is that Random Forest can overcome the saturation issues around 100 - 150 m^3/ha shown in their measurements, even though the average stem volume is up to 270 m^3/ha	Plots, radius = 12.6 m, screened to represent homogenous areas and aggregated
Sweden	Optical Spot 4/5, PALSAR strips, 8+4 observations for 2014	kNN estimation with Spot 4/5 and NFI data + water cloud model and BIOMASAR approach with multi-temporal PALSAR data. New Lidar data used to select the best alternative map	Saturation effects not commented on but the model relies on stem volume of "dense forest", which must be estimated for 30 km x 30km processing tiles from external sources	National Forest Inventory plots with 7m or 10m radius
Indonesia	PALSAR 25m mosaics HH+HV (2007+2009), additionally one ASAR/IM coverage 2004+2005, Sentinel-1 IW VV+VH (2015), PALSAR-2 mosaics HH+HV (2015)	Regression models, using Kalimantan as an analysis unit. Ratio and texture values at pixel spacing of 100 m: $AGB = a_1 e^{v1} + a_2 e^{v2} + c$ With so many input variables, over-fitting identified as a challenge	No saturation assumed, based on earlier work by the same researchers (and others but with lower biomass forest) but reference is made to Mitchard et al (2011) where L-band saturation limit is stated to be around 200 t/ha.	Plots, rectangular (20m x 50m) and circular (radius varying between 16 m and 35 m
Mexico	PALSAR 25m mosaics HH+HV, Landsat data and DEM data	MaxEnt to estimate probabilities of discrete biomass classes and the probability weighted average over the biomass classes; whole of Mexico planned as an analysis unit	AGB above 200 t/ha is assumed to be insignificant in Mexico	Plots from the Mexican National Forest Inventory, 1 ha sampled with 4 plots with a radius of 11.28 m
South Africa	PALSAR 25m HH+HV mosaics for 2007 and 2010, PALSAR-1 dual-pol	Bayesian approach using the water cloud model	No saturation assumed in savanna areas. It is unclear whether other forests are mapped	1 ha plots (37 - 120 of these)

	(strip?) data for 2015			

5 Properties of the regional biomass maps

5.1 Introduction

In this section we present analysis of the above-ground biomass (AGB) estimates for all five regional products and for the 2005, 2010 and 2015 epochs in terms of the distributions of the biomass estimates within different biomass ranges. The purpose of this analysis is to give a more complete description of the properties of the regional products, to assess whether they satisfy what was requested in the Statement of Work, and to quantify how they differ from the products requested by the users in Task 1. The analysis is based on histograms of the estimated biomass values within specified biomass ranges, together with summary statistics. In particular, for each region the following information is provided for the three epochs:

1. A Table showing for each biomass range:
    a. the number of reference values, n;
    b. the average estimated AGB;
    c. the average reference AGB;
    d. the root mean square error (RMSE);
    e. the R^2 value;
    f. the standard deviation (SD) of the errors;
    g. the bias.

    The same information is also provided for the overall dataset for each epoch.

2. Histograms of the errors (residuals) defined as estimated value – reference value in each AGB range, together with summary statistics on the errors (number of points, and their average value and standard deviation (SD)).

3. Histograms of the AGB values in the reference data and the overall estimated data.

4. A scatterplot of the estimated AGB against the reference values.

Note that for South Africa only information for the 2005 epoch is provided. This is because for 2005 the reference data is an AGB map obtained from a Lidar transect, which contains about 38000 samples that are divided equally into training and validation data, while in 2010 only 44 reference plots are available; this is insufficient for meaningful statistical analysis.

This information is discussed for each region separately, but in Section 5.7 we provide an overview of the overall findings and discuss their implications in the context of GlobBiomass objectives.

To clarify how some of the quantities are defined, denote the set of reference values in a given range of AGB as x_i and the associated estimated AGB as y_i. Then write the average values as:

$$\bar{x} = \frac{1}{n}\sum_{1}^{n} x_i \quad \text{and} \quad \bar{y} = \frac{1}{n}\sum_{1}^{n} y_i$$

The bias in the given range is defined as $b = \bar{y} - \bar{x}$. This (non-standard) definition of the bias in the estimator means that the residual errors must have mean equal to *b* since the error in estimating point *i* is given by $e_i = y_i - x_i$, and $\bar{e} = \bar{y} - \bar{x} = \frac{1}{n}\sum_1^n y_i - x_i = b$.

So we can write the error as $e_i = b + \varepsilon_i$ where ε_i is mean-zero, and $y_i = x_i + b + \varepsilon_i$

The Mean Square Error (MSE) is given by:

$$\frac{1}{n}\sum_1^n (y_i - x_i)^2 = \frac{1}{n}\sum_1^n (b - \varepsilon_i)^2 = b^2 + \frac{1}{n}\sum_1^n \varepsilon_i^2 = b^2 + \mathrm{var}(\varepsilon).$$

where *var* denotes variance = SD^2 and *SD* is standard deviation. Hence the RMSE is given by:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\left(b^2 + SD^2(\varepsilon)\right)}$$

This relation helps us to see whether the dominant contributor to the RMSE is the bias or the variance.

5.2 Poland

For both all epochs, the estimated AGB has its lowest bias in the middle of the AGB range. AGB tends to be increasingly overestimated for smaller values of AGB (by up to ~150 t/ha) and increasingly underestimated for larger values of AGB (Table 5-1 and Figures 5-1a-c). This is what would be expected for many forms of data fitting (e.g., in regression, the regression curve passes through the point $(\overline{B}_{ref}, \overline{B}_{est})$, where $\overline{B}_{ref}$ and $\overline{B}_{est}$ are the overall means of the reference and estimated datasets respectively). This compression of the estimates around the mean value is clearly seen in the overall histogram of estimated AGB, which is roughly Gaussian (though skewed), while the reference data exhibit a uniform distribution. For all epochs, the relative bias (i.e., bias/(reference mean)) exceeds 100% for the lowest biomass range and exceeds 25% for AGB > 250 t/ha; it is consistently less than 20% only in the AGB ranges from 150-250 t/ha. The RMSE values are larger for large and small AGB and smaller near the middle of the AGB range. Both these effects are again simply a result of the fitting procedure. By comparing the biases with the SDs shown on the histograms it can be seen that the RMSE is dominated by the bias except for the 150-200 t/ha range for 2005 and 2015 and the 100-150 and 150-200 t/ha ranges for 2010 . The overall relative RMSE is 33% in 2005, 39% in 2010 and 35% in 2015, but Table 5-1 shows that this does not reflect the way the error is divided across biomass ranges. Figures 5-1a-b indicates that the retrievals show little sensitivity to AGB above about 175 t/ha, which is the familiar saturation effect expected for L-band data, though this saturation value is higher than is often reported. The AGB retrievals in 2015 (Figure 5-1c) were based on Sentinel-1 (C-band) and Sentinel-2 and show a similar saturation effect to that observed in 2005 and 2015.

Table 5-1. Summary statistics for the above-ground biomass (AGB) maps of Poland for the 2005, 2010 and 2015 epochs. The values are based on an independent sample of reference AGB values and reported by AGB class across the entire range of reference AGBs. NA – not available; RMSE – root mean square error; SD – standard deviation.

2005 epoch							
AGB classes (t/ha)	**n**	**Average estimated AGB (t/ha)**	**Average reference AGB (t/ha)**	**RMSE (t/ha)**	**R^2**	**SD(error) (t/ha)**	**Bias (t/ha)**
0-50	13	79	26	58	0.3	25	53
50-100	14	132	77	62	0.2	26	55
100-150	16	166	130	41	0.1	20	36
150-200	16	178	176	21	0.0	22	2
200-250	14	192	222	40	0.0	27	-30
>250	16	210	280	73	0.5	22	-70
Overall	**89**	**162**	**156**	**52**	**0.7**	**NA**	**6**
2010 epoch							
AGB classes (t/ha)	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**R^2**	**SD(error) (t/ha)**	**Bias (t/ha)**
0-50	13	89	26	67	0.4	23	63
50-100	19	130	81	54	0.0	24	49
100-150	17	136	128	26	0.2	25	8
150-200	16	156	177	32	0.0	25	-21
200-250	11	172	242	62	0.0	20	-70
>250	8	184	269	86	0.5	16	-85
Overall	**84**	**140**	**137**	**54**	**0.6**	**54**	**3**
2015 epoch							
AGB classes (t/ha)	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**R^2**	**SD(error) (t/ha)**	**Bias (t/ha)**
0-50	25	114	22	99	0.17	37	92
50-100	64	152	80	78	0.14	29	72
100-150	89	174	127	54	0.15	26	47
150-200	98	181	174	28	0.02	27	7
200-250	95	187	224	43	0.07	23	-37
> 250	80	202	273	77	0.02	30	-71
Overall	**451**	**177**	**171**	**60**	**0.40**	**60**	**6**

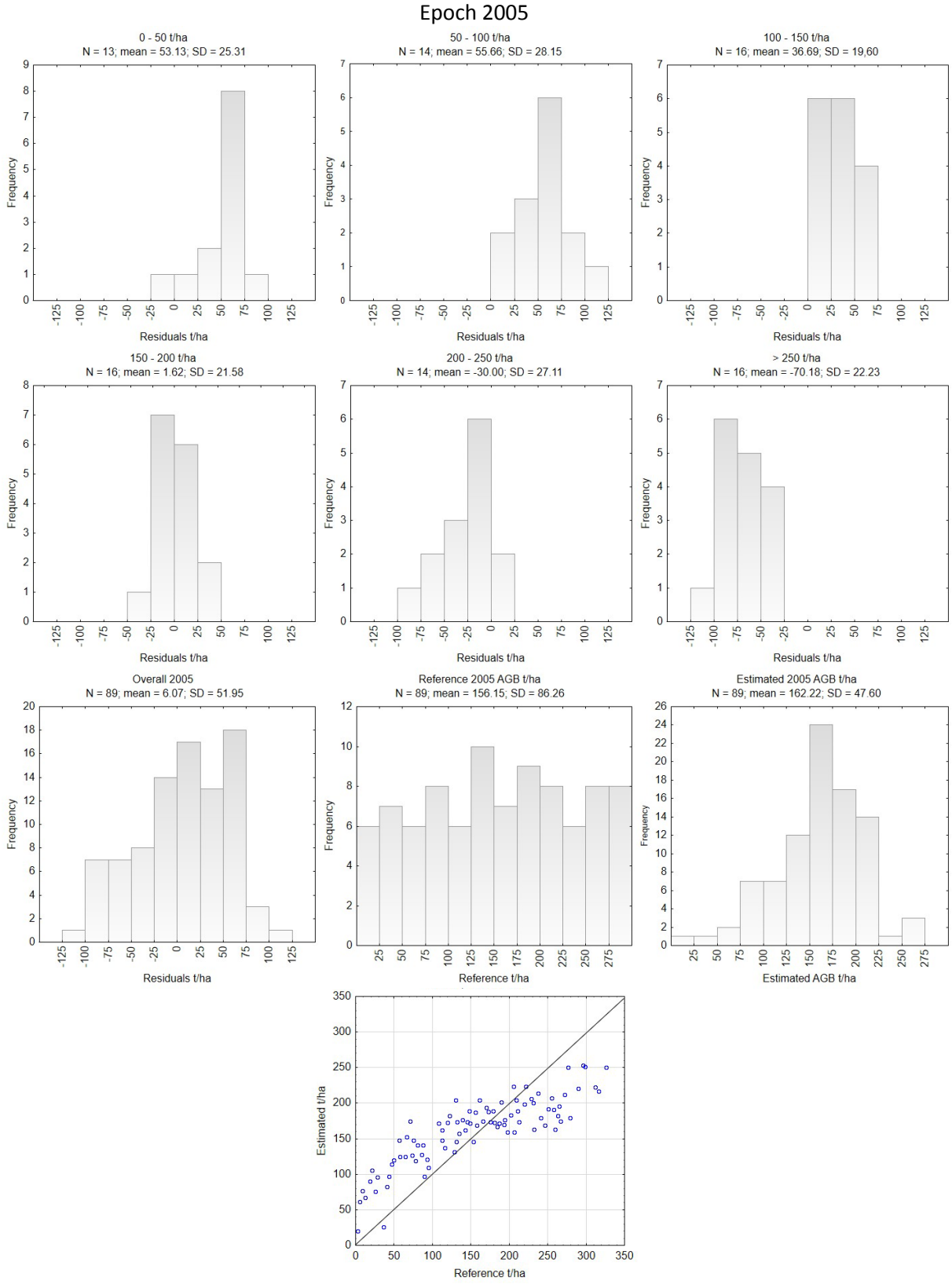


Figure 5-1a. Histograms of above-ground biomass (AGB) residual errors (estimated minus reference) by AGB class (0-50 t/ha, 50-100 t/ha, 100-150 t/ha, 150-200 t/ha, 200-250 t/ha, >200 t/ha) and overall, histograms of reference and estimated AGB data, and scatterplot of estimated against reference AGB values for epoch 2005 in Poland.

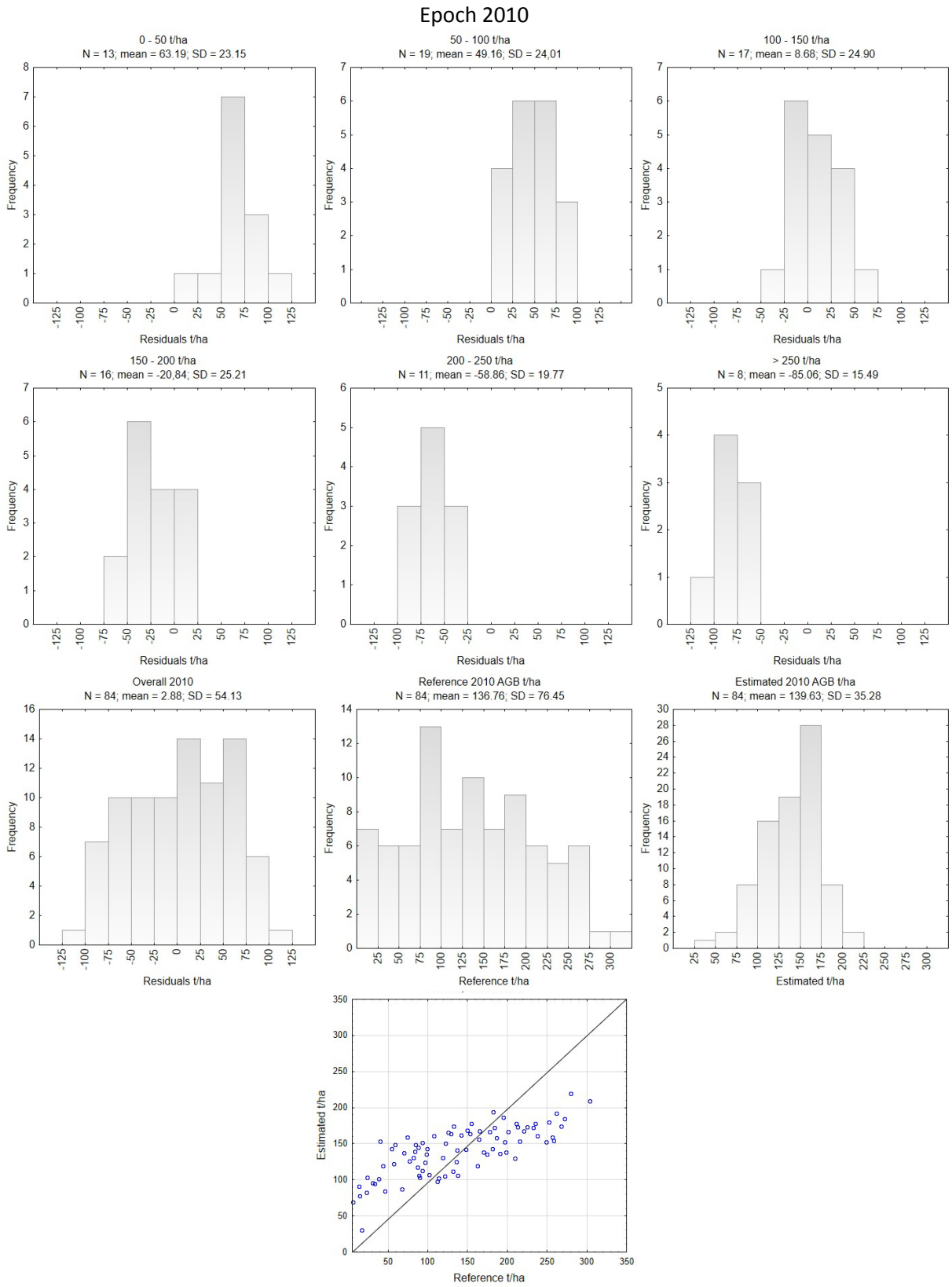


Figure 5-1b. Histograms of above-ground biomass (AGB) residual errors (estimated minus reference) by AGB class (0-50 t/ha, 50-100 t/ha, 100-150 t/ha, 150-200 t/ha, 200-250 t/ha and >200 t/ha) and overall, histograms of reference and estimated AGB data, and scatterplot of estimated against reference AGB values for epoch 2010 in Poland.

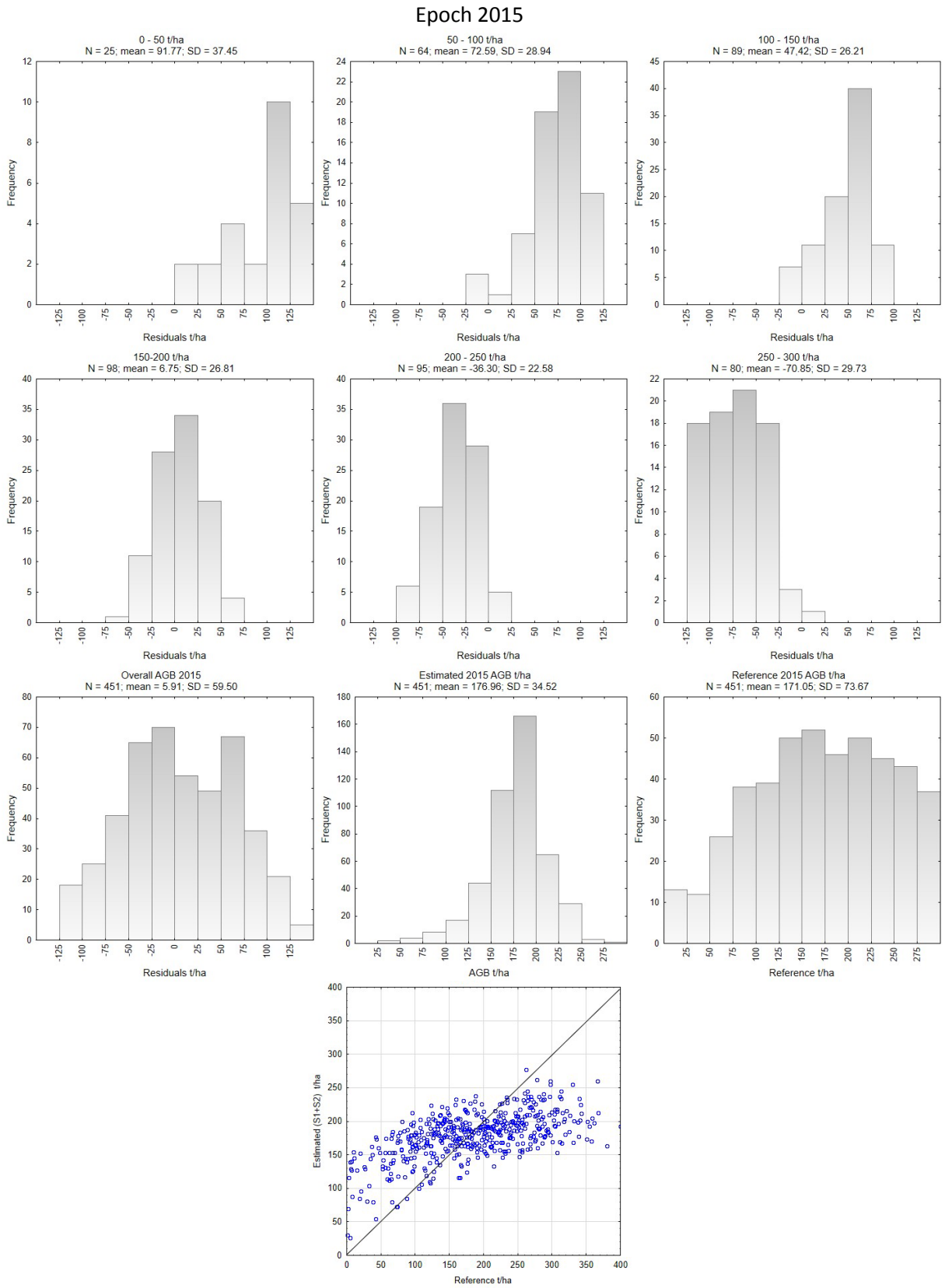


Figure 5-1c. Histograms of above-ground biomass (AGB) residual errors (estimated minus reference) by AGB class (0-50 t/ha, 50-100 t/ha, 100-150 t/ha, 150-200 t/ha, 200-250 t/ha and >200 t/ha) and overall, histograms of reference and estimated AGB data, and scatterplot of estimated against reference AGB values for epoch 2015 in Poland.

5.3 Sweden

For Sweden, the chosen approach overestimates the lowest values of AGB by up to 34 t/ha and increasingly underestimates AGB for higher values for all epochs (Table 5-2); for the 2005 epoch this underestimate is 82 t/ha for the highest AGB class. The bias is least near the mid-range of AGB. This is exactly the same pattern as is seen for Poland, probably for similar reasons, even though the methodology is different. The relative bias at the lower end of the AGB range is greater than 150% at all years and at the upper end is around 40% (all epochs). As for Poland, the RMSE values are smaller near the middle of the AGB range and except in the middle of the range the main contributor to the RMSE is bias. The overall relative RMSE is 59% in 2005, 57% in 2010 and 62% in 2015, and exceeds 29% in all the AGB ranges. Comparison between the histograms for the NFI reference plots and the retrieved values shows that the retrievals to some extent capture the actual AGB structure of the Swedish forests for the higher AGB levels, but do not represent the lower AGB levels well, in particular failing to pick up the large proportion of forest stands in the 0-50 t/ha range (Figure 5-2a-c). In 2010 and 2015, the inclusion of L-band (BIOMASAR-L PALSAR-2) does not improve the retrievals in comparison with those obtained in 2005 and 2010, confirmed by the high overall relative RMSE (57% and 62% respectively).

Table 5-2. Validation statistics of the above-ground biomass (AGB) maps of Sweden, derived using the kNN method (2005), combined kNN and BIOMASAR-L (2010) and the BIOMASAR-L method (2015 epoch). The values are based on an independent sample of reference AGB and reported by AGB class and across the entire range of reference AGB. NA – not available; RMSE – root mean square error; SD – standard deviation.

2005 (kNN, best map)						
AGB classes (t/ha)	**# of validation data**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD(error) (t/ha)**	**Bias (t/ha)**
0-30	878	48	14	40	21	34
30-60	822	71	45	37	26	26
60-90	808	84	74	28	26	10
90-120	538	92	104	31	29	-11
120-150	361	101	134	45	31	-33
150-180	265	106	164	67	34	-57
180-210	137	112	193	888	33	-82
Overall	**3925**	**80**	**80**	**47**	**47**	**0**
2010 (merged kNN+BIOMASAR-L)						
AGB classes (t/ha)	**# of validation data**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD(error) (t/ha)**	**Bias (t/ha)**
0-30	901	41	14	36	24	27
30-60	871	69	46	35	26	24
60-90	850	86	75	30	28	12
90-120	606	100	104	30	29	-5
120-150	361	106	133	42	31	-27
150-180	245	115	164	60	35	-49
180-210	155	122	194	80	36	-72
Overall	**4147**	**81**	**81**	**46**	**46**	**0**

Table 5-2. Cont.

2015 (BIOMASAR-L)						
AGB classes (t/ha)	**# of validation data**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD(error) (t/ha)**	**Bias (t/ha)**
0-30	883	36	14	38	31	23
30-60	967	63	46	39	35	18
60-90	926	81	75	39	38	6
90-120	718	95	104	42	40	-10
120-150	446	104	134	52	43	-30
150-180	285	111	163	67	42	-52
180-210	174	118	194	87	42	-76
Overall	**4607**	**78**	**85**	**53**	**53**	**7**

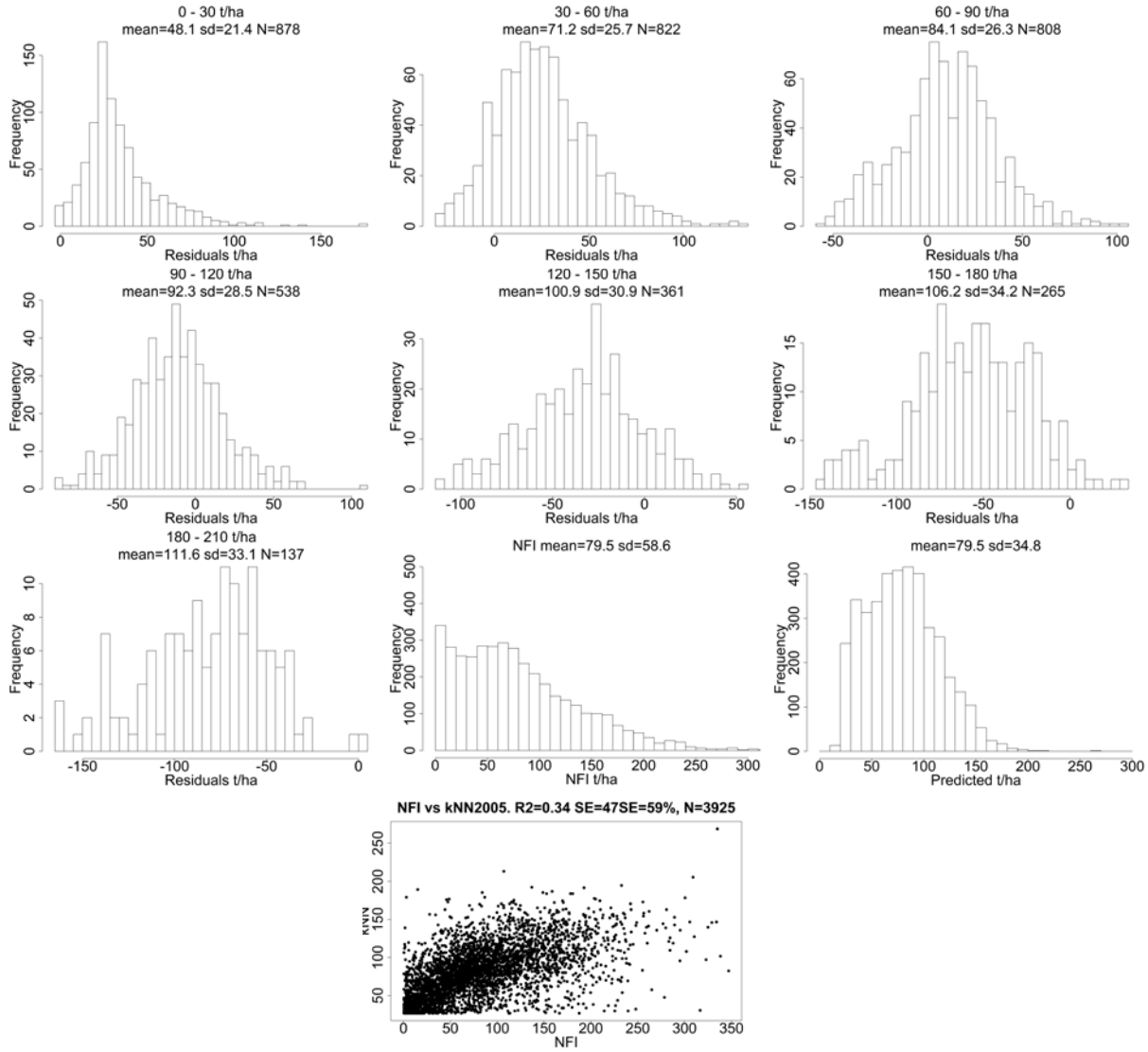


Figure 5-2a. Histograms of above-ground biomass (AGB) residuals (estimated minus reference) by AGB class (0-30 t/ha, 30-60 t/ha, 60-90 t/ha, 90-120 t/ha, 120-150 t/ha, 150-180 t/ha and 180-210 t/ha), overall histograms of estimated and reference AGB data, and scatterplots of estimated against reference AGB values using the KNN method for the 2005 epoch in Sweden.

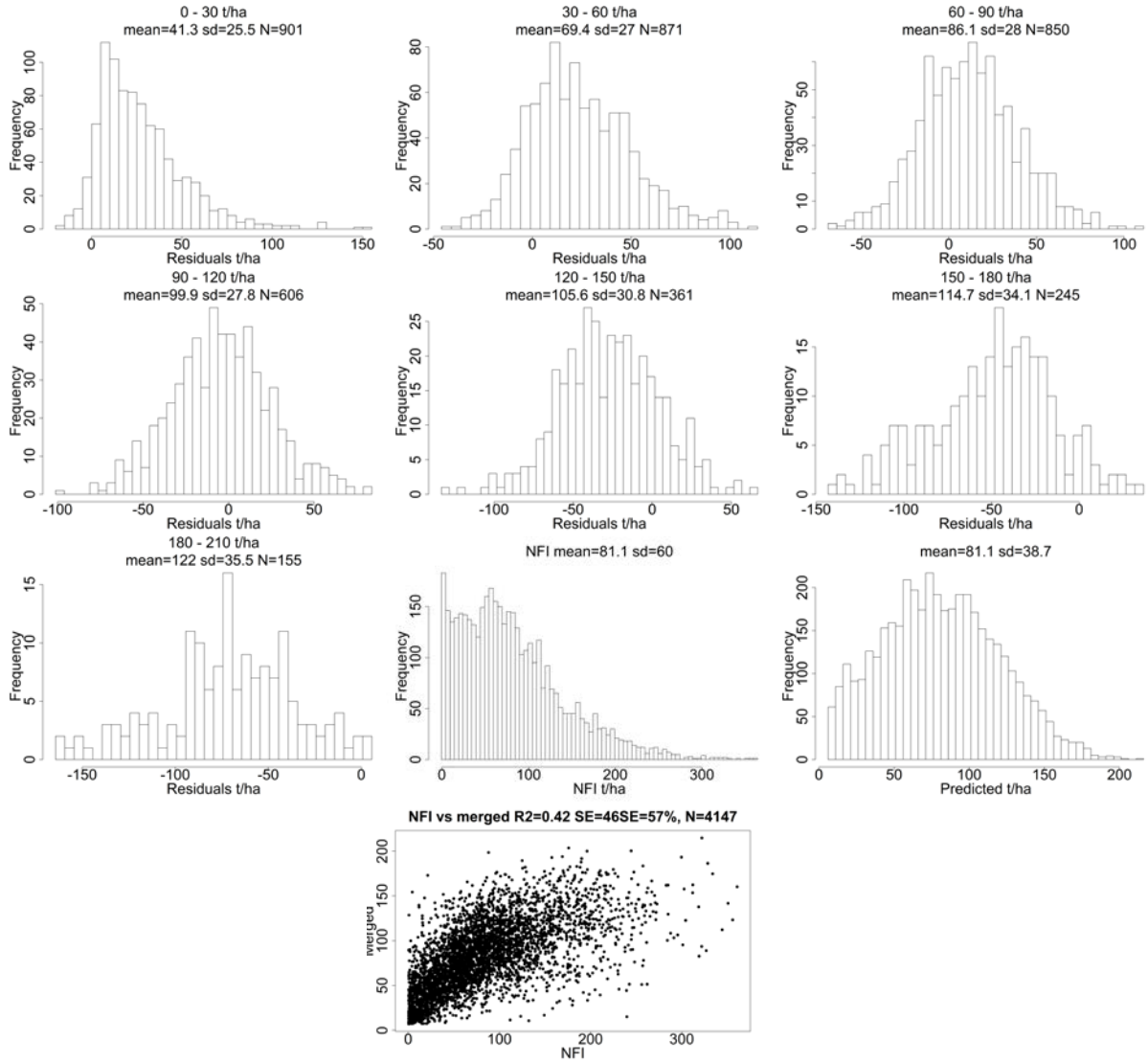


Figure 5-2b. Histograms of above-ground biomass (AGB) residuals (estimated minus reference) by AGB class (0-30 t/ha, 30-60 t/ha, 60-90 t/ha, 90-120 t/ha, 120-150 t/ha, 150-180 t/ha and 180-210 t/ha), overall histograms of estimated and reference AGB data, and scatterplots of estimated against reference AGB values using the combined KNN+BIOMASAR-C (ASAR) method for the 2010 epoch in Sweden.

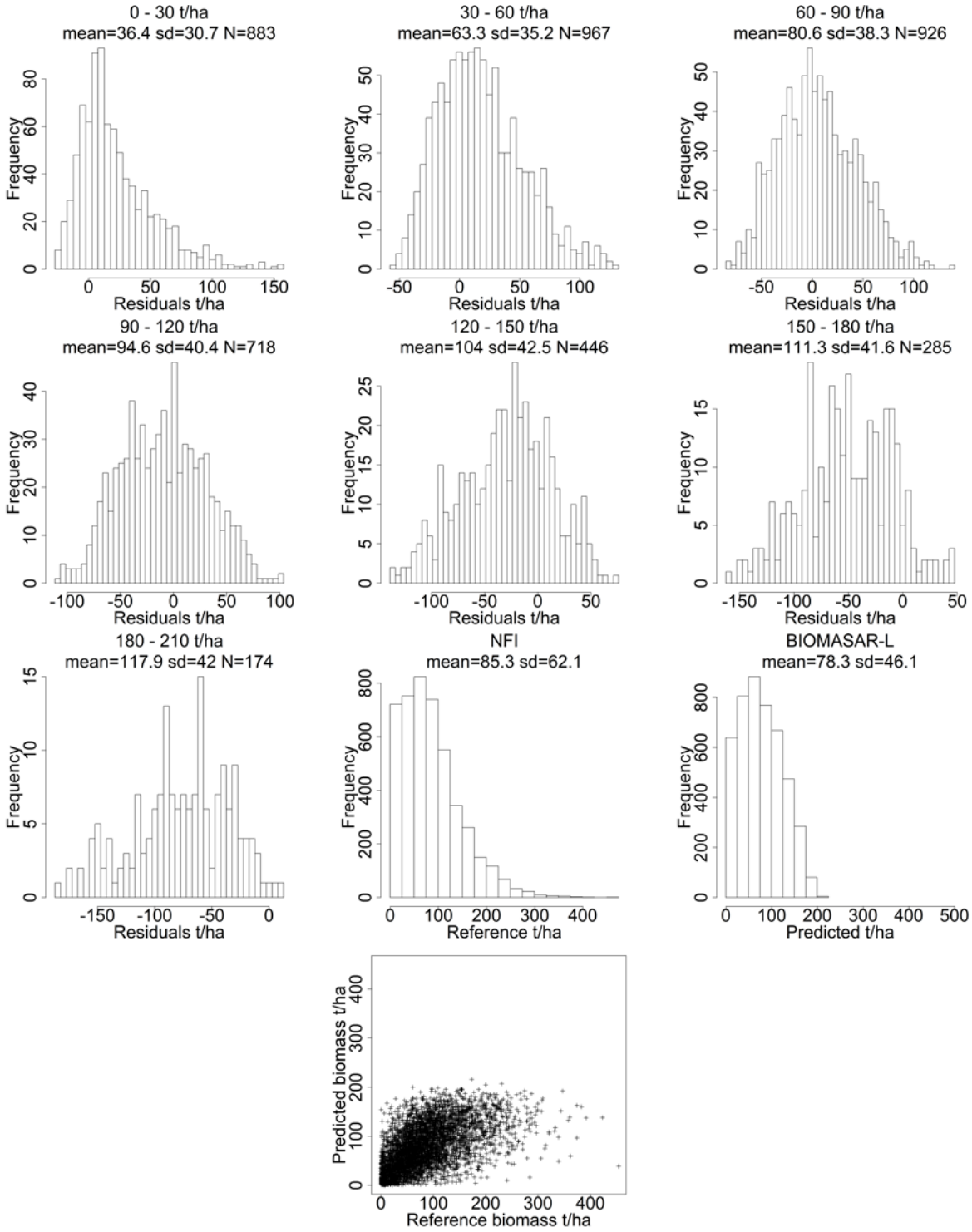


Figure 5-2c. Histograms of above-ground biomass (AGB) residuals (reference minus estimated) by AGB class (0-30 t/ha, 30-60 t/ha, 60-90 t/ha, 90-120 t/ha, 120-150 t/ha, 150-180 t/ha and 180-210 t/ha), overall histograms of estimated and reference AGB data, and scatterplots of estimated against reference AGB values using the BIOMASAR-L (ALOS-2 PALSAR-2) method for the 2015 epoch in Sweden.

5.4 Kalimantan (Indonesia)

Table 5-3a shows the validation statistics of AGB mapped in Kalimantan during the 2005, 2010 and 2015 epochs. Average biomass estimates are consistently higher than those obtained from reference data, indicated by a positive bias, and the bias is somewhat higher for the 2005 and 2015 epochs. The magnitude of bias is different across AGB classes, with higher values in the 50-100 t/ha and 100-150 t/ha classes (which is the opposite of what is seen for Poland and Sweden). Figures 5.3a-c show the histograms of AGB residuals (estimated minus reference) by AGB class (and overall), histograms of estimated and reference AGB values and the scatterplot of estimated against reference AGB for the 2005, 2010 and 2015 epochs respectively. The histogram of the AGB estimates and the reference data set depicts a similar distribution in all epochs. However, AGB values up to ~ 200 t/ha are strongly overestimated in the three epochs, but more markedly in 2005 and 2010. Only for the AGB range exceeding 200 t/ha are the residuals evenly distributed around 0 t/ha. In 2005 the relative biases in the 50-100 t/ha and 100-150 t/ha ranges are 124% and 41% respectively, in 2010 the corresponding values are 67% and 30%, whereas in 2015 the values are 71% and 32% respectively. This reflects the structure seen in the histograms of the residuals: in 2005 the residuals in both cases are sharply peaked about 75 t/ha, whereas in 2010 and 2015 they are more widely spread, particularly in the 100-150 t/ha range. The RMSE for all ranges is similar in all epochs; the relative RMSE exceeds 23% in all ranges, but for the lower ranges (0-50 t/ha and 50-100 t/ha) in many cases it exceeds 100%. Hence the relative overall RMSE, ranging between 33% (2015) and 38% (2005), fails to capture the way the error is distributed across the biomass ranges. In most cases the scatter in the estimates dominates the RMSE; only for the 100-150 t/ha and 150-200 t/ha ranges for 2005 and 0-50 t/ha and 50-100 t/ha ranges for 2010 does the contribution from bias become comparable. The highest biomass values in the AGB estimates are around 300 t/ha (2005 and 2010) and 250 t/ha (2015), whereas the reference data show many data points with higher values in these regions. It should be noted that more than 55% (70% in 2015) of the points lie in the range above 150 t/ha and the scatterplots indicate that this is close to the saturation value of AGB. This suggests that the regression will be dominated by these high values (which will tend to make the mean values agree) and the scatter around these values will be roughly random (which is seen in the histograms of the residuals). This would tend to lead to low bias in the upper biomass ranges, as is observed.

Table 5-3a. Validation statistics of the above-ground biomass (AGB) maps of Kalimantan for the 2005, 2010 and 2015 epochs. The values are based on an independent sample of reference AGB and reported by AGB class and across the entire range of reference AGB. NA – not available; RMSE – root mean square error; SD – standard deviation.

	2005 epoch						
AGB classes (t/ha)	**n**	**Average estimated AGB (t/ha)**	**Average reference AGB (t/ha)**	**RMSE (t/ha)**	**R^2**	**SD (error) (t/ha)**	**Bias (t/ha)**
0-50	134	27	7	41	0.56	36	19
50-100	19	151	68	90	0.12	35	84
100-150	34	179	128	68	0.00	45	52
150-200	78	219	176	55	0.02	35	43
>200	154	239	273	62	0.11	52	-34
Overall	**419**	**159**	**149**	**57**	**0.77**	**56**	**10**
	2010 epoch						
AGB classes (t/ha)	**n**	**Average estimated AGB (t/ha)**	**Average reference AGB (t/ha)**	**RMSE (t/ha)**	**R^2**	**SD (error) (t/ha)**	**Bias (t/ha)**
0-50	141	50	13	54	0.23	39	37
50-100	21	130	78	71	0.07	50	52
100-150	38	171	132	66	0.08	54	39
150-200	117	194	178	42	0.07	39	16
>200	184	207	247	60	0.01	44	-40
Overall	**501**	**154**	**149**	**55**	**0.69**	**55**	**4**
	2015 epoch						
AGB classes (t/ha)	**n**	**Average estimated AGB (t/ha)**	**Average reference AGB (t/ha)**	**RMSE (t/ha)**	**R^2**	**SD (error) (t/ha)**	**Bias (t/ha)**
0-50	83	19	10	27	0.08	26	9
50-100	27	124	72	82	0.10	57	51
100-150	42	165	125	62	0.10	48	40
150-200	117	205	179	42	0.04	33	26
>200	244	208	257	70	0.01	50	-49
Overall	**513**	**169**	**179**	**59**	**0.64**	**58**	**-10**

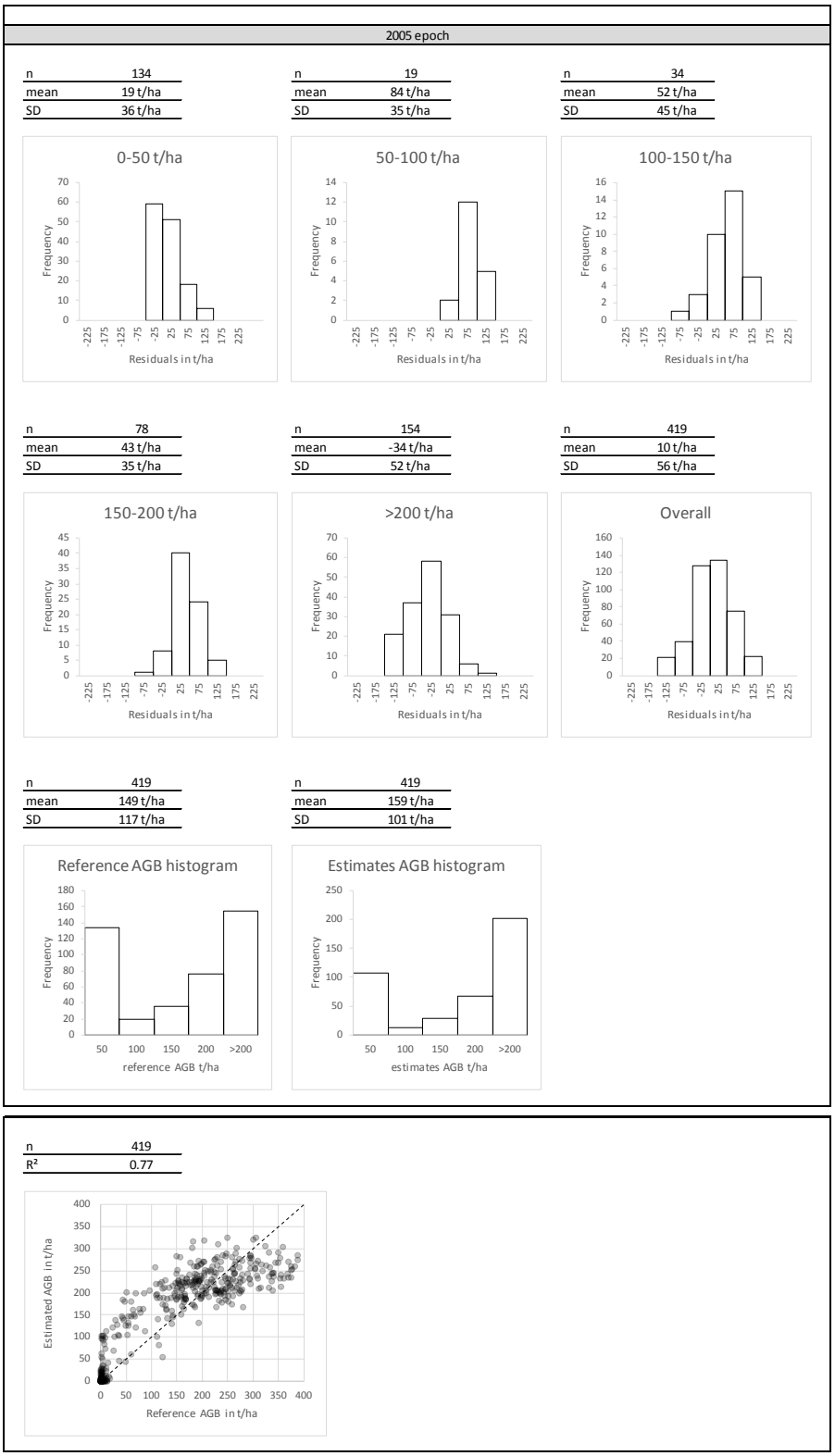


Figure 5-3a. Histograms of above-ground biomass (AGB) residuals (estimated minus reference) by AGB class (0-50 t/ha, 50-100 t/ha, 100-150 t/ha, 150-200 t/ha, >200 t/ha) and overall, histograms of reference and estimated AGB data, and scatterplot of estimated against reference AGB values for epoch 2005 in Kalimantan.

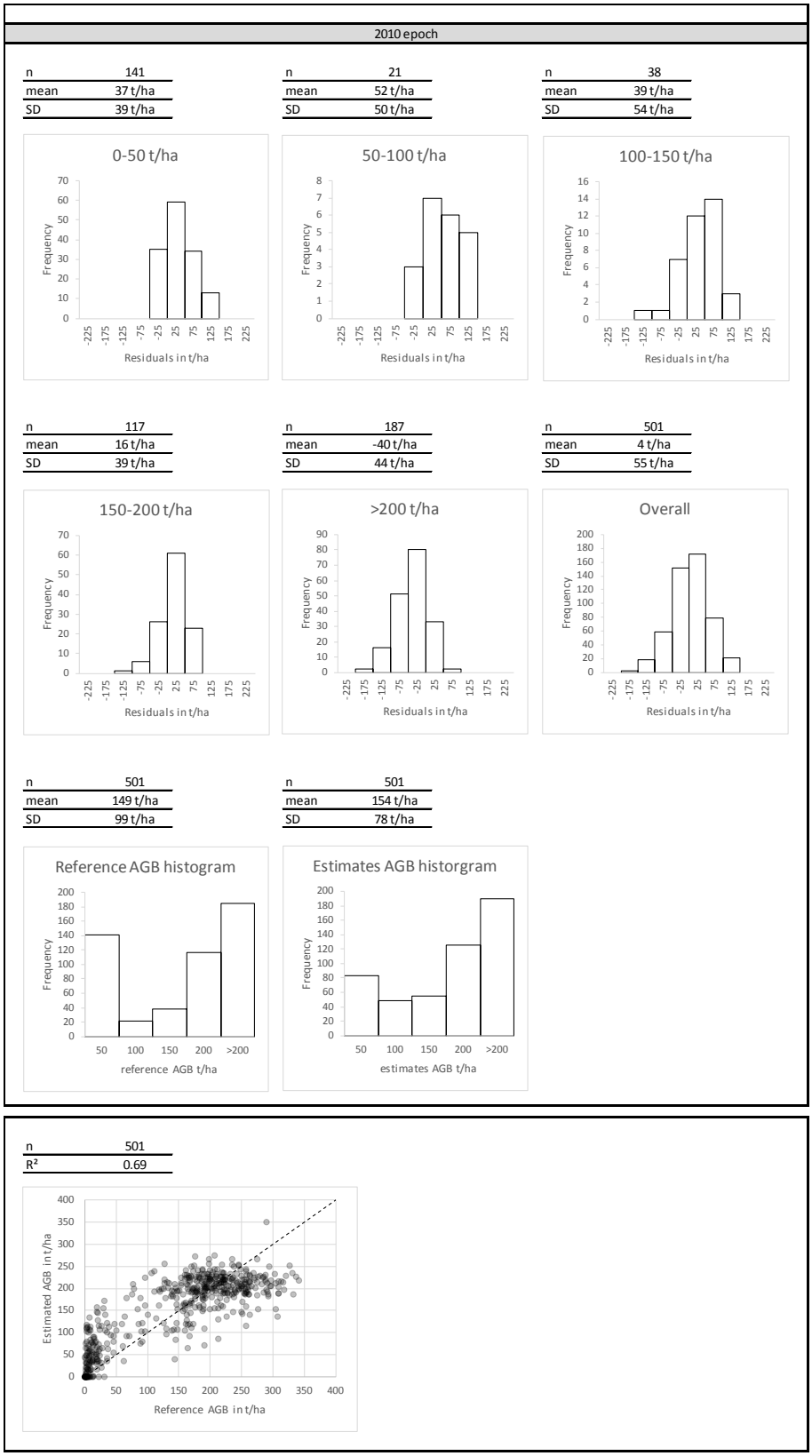


Figure 5-3b. Histograms of above-ground biomass (AGB) residuals (estimated minus reference) by AGB class (0-50 t/ha, 50-100 t/ha, 100-150 t/ha, 150-200 t/ha, >200 t/ha) and overall, histograms of reference and estimated AGB data, and scatterplot of estimated against reference AGB values for epoch 2010 in Kalimantan.

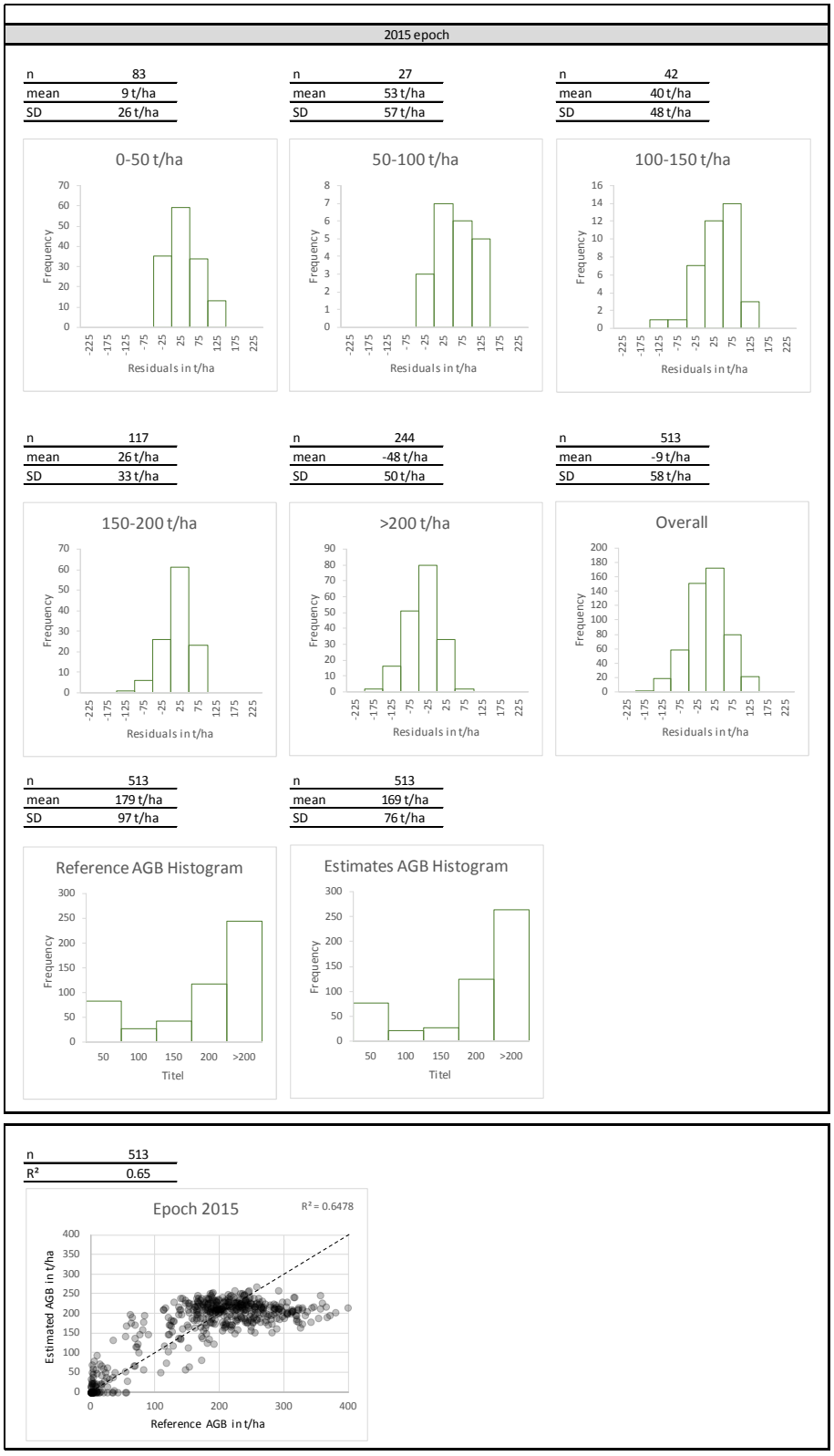


Figure 5-3c. Histograms of above-ground biomass (AGB) residuals (estimated minus reference) by AGB class (0-50 t/ha, 50-100 t/ha, 100-150 t/ha, 150-200 t/ha, >200 t/ha) and overall, histograms of reference and estimated AGB data, and scatterplot of estimated against reference AGB values for epoch 2015 in Kalimantan.

The uncertainty tables per epoch (Table 5-3b) show the errors of measurement, allometry, sampling, and prediction as well as the total propagated error (Ɛ AGB) (Eq. 3-22), expressed as proportion (i.e., 1.0 ≡ 100%).

Table 5-3b. Uncertainty values of the above-ground biomass maps of Kalimantan for the 2005, 2010 and 2015 epochs.

Epoch 2005					
AGB Classes (observed)	**Ɛ measurement**	**Ɛ allometry**	**Ɛ sampling**	**Ɛ prediction**	**Ɛ AGB**
0-50	0.100	0.050	0.200	1.519	**1.179**
50-100	0.100	0.050	0.200	0.596	**0.204**
100-150	0.100	0.050	0.200	0.380	**0.098**
150-200	0.100	0.050	0.200	0.251	**0.058**
>200	0.100	0.050	0.200	0.259	**0.060**
Overall	0.100	0.050	0.200	0.358	**0.091**
Epoch 2010					
AGB Ranges (observed)	**Ɛ measurement**	**Ɛ allometry**	**Ɛ sampling**	**Ɛ prediction**	**Ɛ AGB**
0-50	0.100	0.050	0.200	1.080	**0.609**
50-100	0.100	0.050	0.200	0.546	**0.175**
100-150	0.100	0.050	0.200	0.386	**0.101**
150-200	0.100	0.050	0.200	0.216	**0.050**
>200	0.100	0.050	0.200	0.290	**0.068**
Overall	0.100	0.050	0.200	0.357	**0.090**
Epoch 2015					
AGB Ranges (observed)	**Ɛ measurement**	**Ɛ allometry**	**Ɛ sampling**	**Ɛ prediction**	**Ɛ AGB**
0-50	0.100	0.050	0.200	1.421	**1.036**
50-100	0.100	0.050	0.200	0.661	**0.245**
100-150	0.100	0.050	0.200	0.376	**0.097**
150-200	0.100	0.050	0.200	0.205	**0.047**
>200	0.100	0.050	0.200	0.337	**0.083**
Overall	0.100	0.050	0.200	0.349	**0.087**

5.5 Mexico

An independent validation by AGB range was performed over Mexico using measurements from 696 NFI plots for the Epoch 2005-07 (533 Yucatan and 163 Central); and 700 NFI plots for Epoch 2010 and 2015 (524 Yucatan and 176 Central). The epochs and regions are similar in terms of bias and RMSE across the set of biomass ranges. The pattern of bias by biomass range (Tables 5-4, 5-5 and 5-6) is similar to that seen for Poland and Sweden, with overestimation in the lower biomass ranges and underestimation in the upper ranges, even though the retrieval approach is quite different and the range of biomass values encountered is much less in Mexico (the upper biomass class is >150 t/ha in Mexico, but >250 t/ha in Poland and 180-210 t/ha in Sweden). This pattern of bias is not at all reflected in the overall bias, which is small for all epochs and regions; this occurs because the underestimation in the upper ranges is compensated by the overestimation in the lower ranges. Notable is that the standard deviation of the residuals is roughly the same (around 20-30 t/ha) in all biomass classes and epochs and the biases indicate the central tendencies in these histograms. The relative RMSE considering both sites together is rarely less than 30% in any biomass range and in most cases greatly exceeds this. Hence the relative overall RMSEs of 51% in the 2005/07 epoch, 54% in the 2010 epoch and 56% in the 2015 epoch do not in any way capture the way error is distributed across the biomass ranges. The main contributor to the RMSE is the scatter (although it is not as dominant as, for example, in Sweden) except for the highest biomass range, where the dominant contribution comes from the bias. It is worth noting that the overall AGB histograms for the retrievals and the NFI are quite different (Figures 5-4a-c): the NFI data gives decreasing frequencies up to 150t/ha, whereas the inversions are roughly uniform up to a clear mode at around 100 t/ha, cutting off fairly sharply above this. Finally, although the combination of variables used in this method was sensitive to AGB up to ~150 t/ha, the scatterplots show clear signs of saturation towards the upper end of the biomass range being observed (beyond around 120 t/ha). Higher saturation levels can be observed in the map of Central Mexico for all epochs.

Table 5-4. Validation statistics of the above-ground biomass (AGB) maps of Mexico produced for the 2005/07 epoch. The values are based on an independent sample of reference AGB and reported by AGB class and across the entire range of reference AGB. RMSE – root mean square error; SD – standard deviation.

	2005-07						
Site	**AGB Ranges**	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD (error) (t/ha)**	**Bias (t/ha)**
Central	**0-30**	67	19	13	21	21	6
	30-60	44	51	45	29	28	6
	60-90	29	71	75	35	35	-4
	90-120	12	99	105	49	49	-6
	120-150	9	94	133	55	39	-39
	> 150	2	154	173	29	22	-19
	Overall	**163**	**49**	**48**	**31**	**31**	**1**
	2005-07						
Site	**AGB Ranges**	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD (error) (t/ha)**	**Bias (t/ha)**
Yucatan	**0-30**	133	34	14	29	21	20
	30-60	112	62	45	33	28	17
	60-90	110	83	74	24	23	9
	90-120	77	87	104	29	23	-17
	120-150	77	98	134	41	20	-36
	> 150	24	97	165	71	21	-68
	Overall	**533**	**70**	**70**	**34**	**34**	**0**
	2005-07						
Site	**AGB Ranges**	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD (error) (t/ha)**	**Bias (t/ha)**
Both sites	**0-30**	200	29	14	26	22	15
	30-60	156	59	45	32	29	14
	60-90	139	81	74	27	26	7
	90-120	89	89	104	32	28	-15
	120-150	86	98	134	43	23	-36
	> 150	26	101	166	69	25	-65
	Overall	**696**	**65**	**65**	**33**	**33**	**0**

Table 5-5. Validation statistics of the above-ground biomass (AGB) maps of Mexico produced for the 2010 epoch. The values are based on an independent sample of reference AGB and reported by AGB class and across the entire range of reference AGB. RMSE – root mean square error; SD – standard deviation.

	2010						
Site	**AGB Ranges**	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD (error) (t/ha)**	**Bias (t/ha)**
Central	**0-30**	83	20	13	21	19	7
	30-60	48	51	43	30	28	8
	60-90	25	70	73	34	34	-3
	90-120	12	106	104	42	42	2
	120-150	6	134	130	24	24	4
	> 150	2	160	200	64	49	-40
	Overall	**176**	**47**	**42**	**28**	**28**	**5**
	2010						
Site	**AGB Ranges**	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD (error) (t/ha)**	**Bias (t/ha)**
Yucatan	**0-30**	130	35	14	33	26	21
	30-60	109	60	44	33	30	16
	60-90	111	81	75	23	22	6
	90-120	85	91	104	25	21	-13
	120-150	54	91	133	46	17	-42
	> 150	35	102	168	67	16	-66
	Overall	**524**	**69**	**70**	**35**	**35**	**-1**
	2010						
Site	**AGB Ranges**	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD (error) (t/ha)**	**Bias (t/ha)**
Both sites	**0-30**	213	29	14	29	25	15
	30-60	157	57	44	32	30	13
	60-90	136	79	75	26	25	4
	90-120	97	93	104	28	25	-11
	120-150	60	95	133	44	23	-38
	> 150	37	106	170	67	20	-64
	Overall	**700**	**64**	**63**	**34**	**34**	**1**

Table 5-6. Validation statistics of the above-ground biomass (AGB) maps of Mexico produced for the 2015 epoch. The values are based on an independent sample of reference AGB and reported by AGB class and across the entire range of reference AGB. RMSE – root mean square error; SD – standard deviation.

	2015						
Site	**AGB Ranges**	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD (error) (t/ha)**	**Bias (t/ha)**
Central	**0-30**	83	20	13	20	19	7
	30-60	48	44	43	28	28	1
	60-90	25	68	73	37	37	-5
	90-120	12	82	104	36	28	-22
	120-150	6	118	130	26	23	-12
	> 150	2	149	200	65	40	-51
	Overall	**176**	**43**	**42**	**27**	**27**	**1**
	2015						
Site	**AGB Ranges**	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD (error) (t/ha)**	**Bias (t/ha)**
Yucatan	**0-30**	130	36	14	35	27	22
	30-60	109	56	45	31	29	11
	60-90	111	75	75	21	21	0
	90-120	85	84	105	32	24	-21
	120-150	54	87	133	52	23	-46
	> 150	35	96	168	75	21	-72
	Overall	**524**	**65**	**71**	**38**	**37**	**-6**
	2015						
Site	**AGB Ranges**	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**SD (error) (t/ha)**	**Bias (t/ha)**
Both sites	**0-30**	213	30	14	30	25	16
	30-60	157	53	44	30	29	9
	60-90	136	74	75	25	25	-1
	90-120	97	83	104	32	24	-21
	120-150	60	90	133	50	25	-43
	> 150	37	99	170	74	23	-71
	Overall	**700**	**60**	**63**	**35**	**35**	**-3**

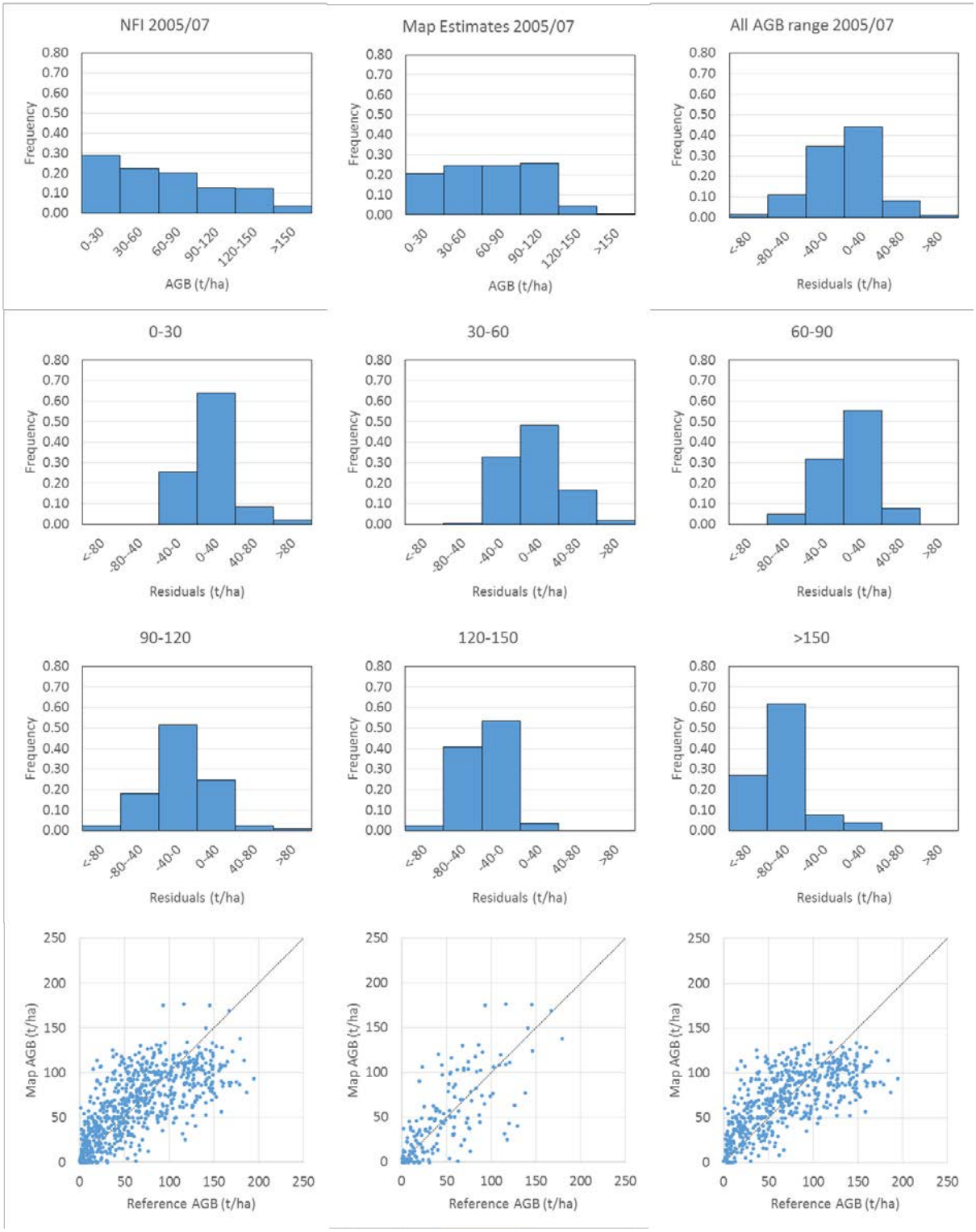


Figure 5-4a. Histograms of reference (NFI) and estimated above-ground biomass (AGB) for 2005/7. Histograms of residual errors (estimated minus reference) by AGB class (0-30 t/ha, 30-60 t/ha, 60-90 t/ha, 90-120 t/ha, 120-150 t/ha & >150 t/ha) and overall. Scatterplots of estimated against reference AGB values for the 2005 epoch for Yucatan (bottom right), Central Mexico (bottom central) and for both maps (Yucatan & Central) combined (bottom left).

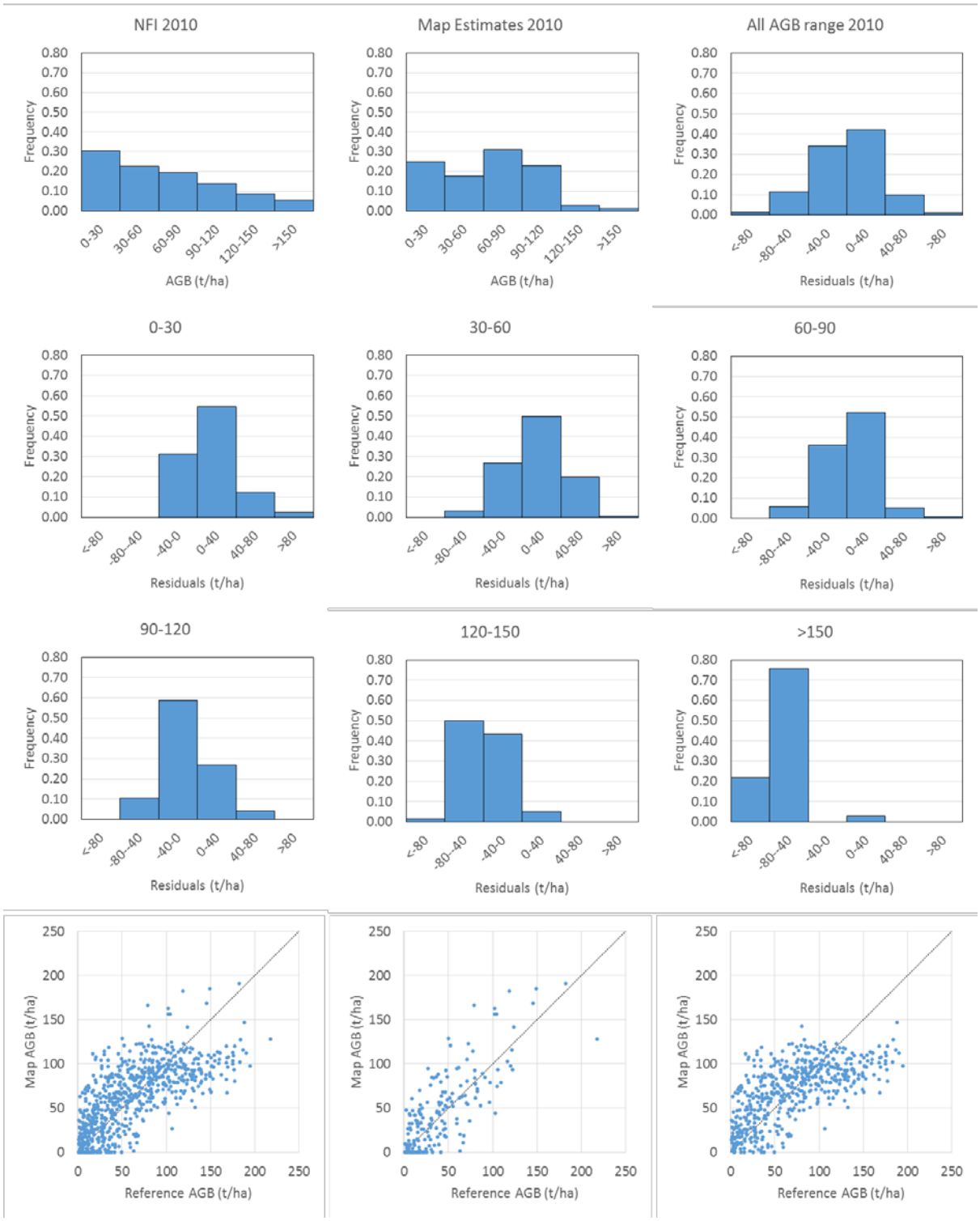


Figure 5-4b. Histograms of reference (NFI) and estimated above-ground biomass (AGB) for 2010. Histograms of residual errors (estimated minus reference) by AGB class (0-30 t/ha, 30-60 t/ha, 60-90 t/ha, 90-120 t/ha, 120-150 t/ha & >150 t/ha) and overall. Scatterplots of estimated against reference AGB values for the 2010 epoch for Yucatan (bottom right), Central Mexico (bottom central) and for both maps (Yucatan & Central) combined (bottom left).

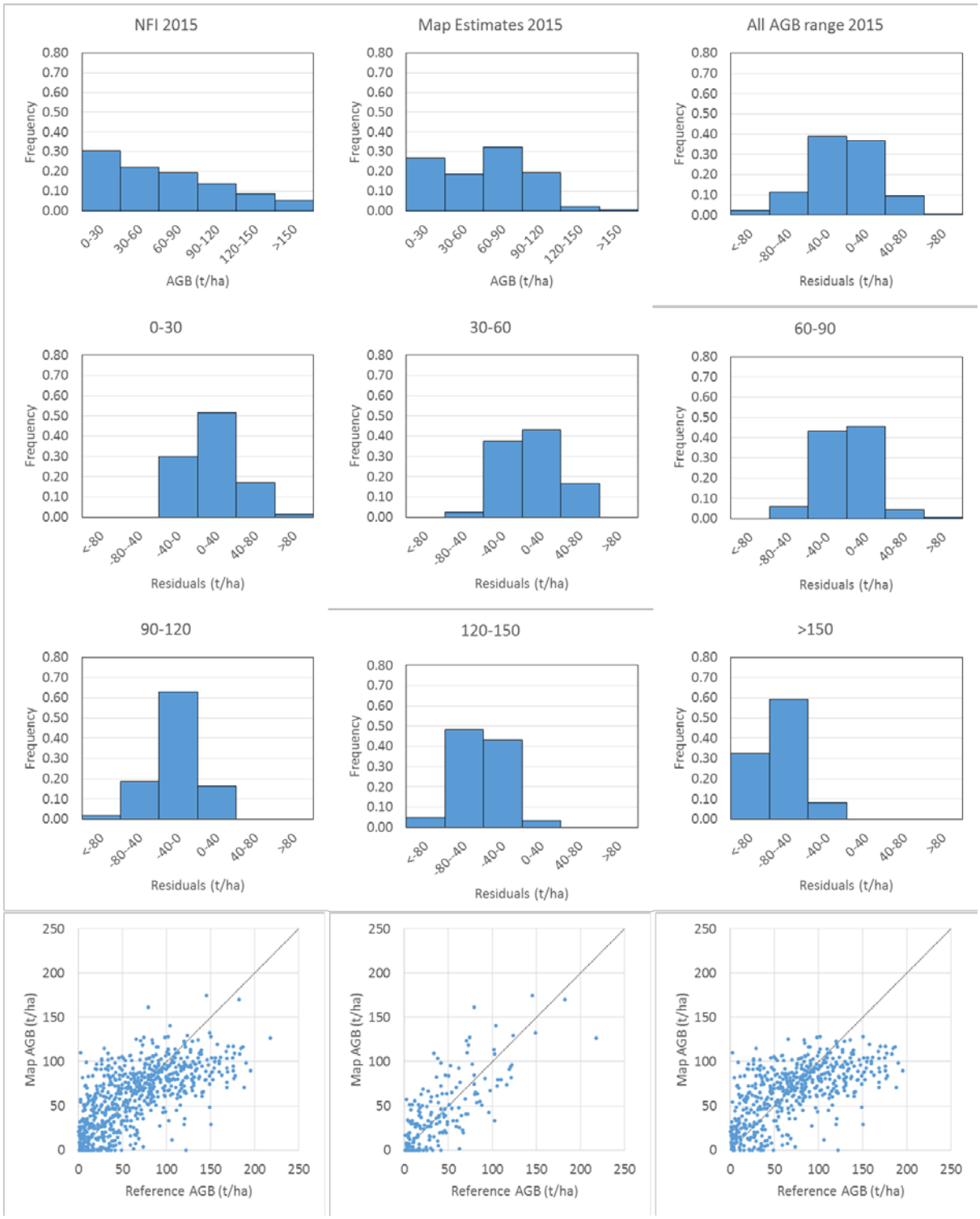


Figure 5-4c. Histograms of reference (NFI) and estimated above-ground biomass (AGB) for 2015. Histograms of residual errors (estimated minus reference) by AGB class (0-30 t/ha, 30-60 t/ha, 60-90 t/ha, 90-120 t/ha, 120-150 t/ha & >150 t/ha) and overall. Scatterplots of estimated against reference AGB values for the 2015 epoch for Yucatan (bottom right), Central Mexico (bottom central) and for both maps (Yucatan & Central) combined (bottom left).

5.6 South Africa

The range of biomass covered by the South African sites is only up to 100 t/ha, and as for all other regions the AGB is overestimated over this range and the bias is larger for the lower biomass range. The RMSE is dominated by the scatter in the errors, with relatively little contribution from the bias, except in the lower AGB class. For the upper AGB range the relative bias is between 0-6%, which probably reflects the fact that the larger AGB values play a key part in controlling the fitting procedure, so that one would expect the mean values of the reference and estimated AGB to be similar for larger AGB. However, Figure 5-5a and Figure 5-5c shows that the points where AGB exceeds 80 t/ha are under-estimated (although no histogram is given for this range). Hence the model appears to suffer the same problems of over-estimation of low biomass and under-estimation of high biomass seen in the other regions. The histograms indicate the compensation of low values by high values in a widely scattered distribution. The relative RMSE for the upper range is between 11% (2010) and 21% (2005). Note that the overall relative RMSE between 62% (2010) and 83% (2015) is caused by the large number of points in the lower biomass range which leads to the overall mean AGB being only in the range 20-37 t/ha. Comparing the histograms for the overall reference and estimated AGBs reveals that the estimated data does not capture the overall structure of biomass, showing a much more even distribution of frequencies across the biomass range than in the reference data, and not picking out the prominent spike in the lowest biomass range, with this clearly marked in 2005 and 2015.

Table 5-7. Validation statistics of above-ground biomass (AGB) estimates of South Africa for epochs 2005, 2010 and 2015 using an independent AGB reference dataset. RMSE – root mean square error; SD – standard deviation.

	Epoch 2005						
AGB ranges	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**R^2**	**SD(error) (t/ha)**	**Bias (t/ha)**
0-20	7,441	22	8	26	0.01	21	14
20-40	5,924	36	30	19	0.07	18	6
40-60	3,982	53	48	17	0.08	16	5
60-80	1,163	68	68	14	0.06	14	0
All	**18,788**	**37**	**28**	**21**	**0.40**	**19**	**9**
	Epoch 2010						
AGB ranges	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**R^2**	**SD(error) (t/ha)**	**Bias (t/ha)**
0-20	2,216	12	8	8	0.35	8	4
20-40	734	36	29	14	0.26	11	7
40-60	233	53	46	15	0.06	11	7
60-80	5	66	63	7	0.45	5	3
All	**3,188**	**20**	**16**	**10**	**0.73**	**16**	**5**
	Epoch 2015						
AGB ranges	**n**	**Average AGB estimates (t/ha)**	**Average AGB reference (t/ha)**	**RMSE (t/ha)**	**R^2**	**SD(error) (t/ha)**	**Bias (t/ha)**
0-20	12,732	13	10	11	0.23	10	3
20-40	7,170	37	28	19	0.23	14	9
40-60	1,183	64	47	22	0.10	14	17
60-80	112	71	67	13	0.00	9	4
All	**21,209**	**24**	**18**	**15**	**0.61**	**18**	**6**

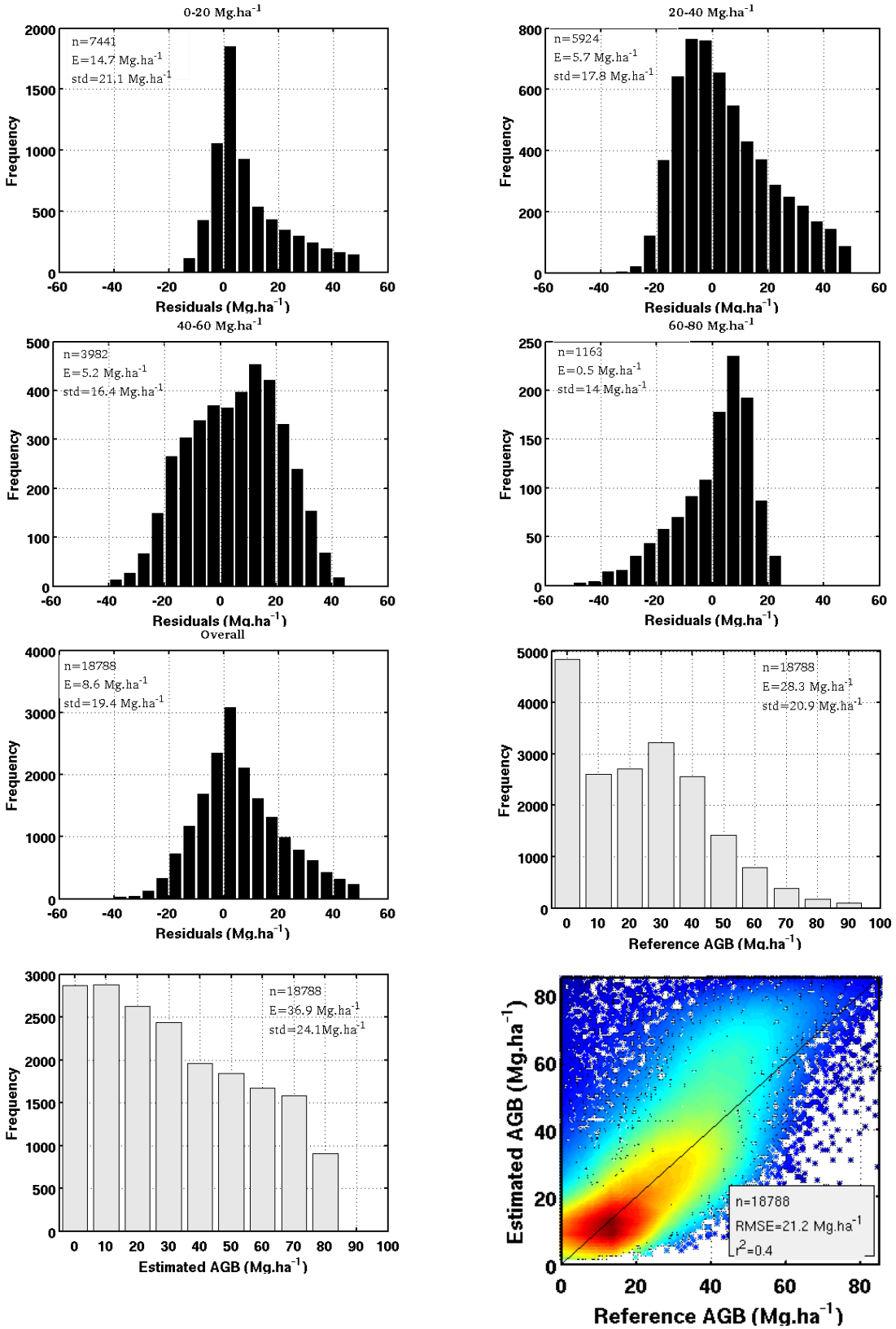


Figure 5-5a. Histograms of above-ground biomass (AGB) residuals (reference minus estimated) by AGB class (0-20 t/ha, 20-40 t/ha, 40-60 t/ha, and 60-80 t/ha) and overall, histograms of reference and estimated AGB data, and scatterplot of estimated against reference AGB values for epoch 2005 in South Africa.

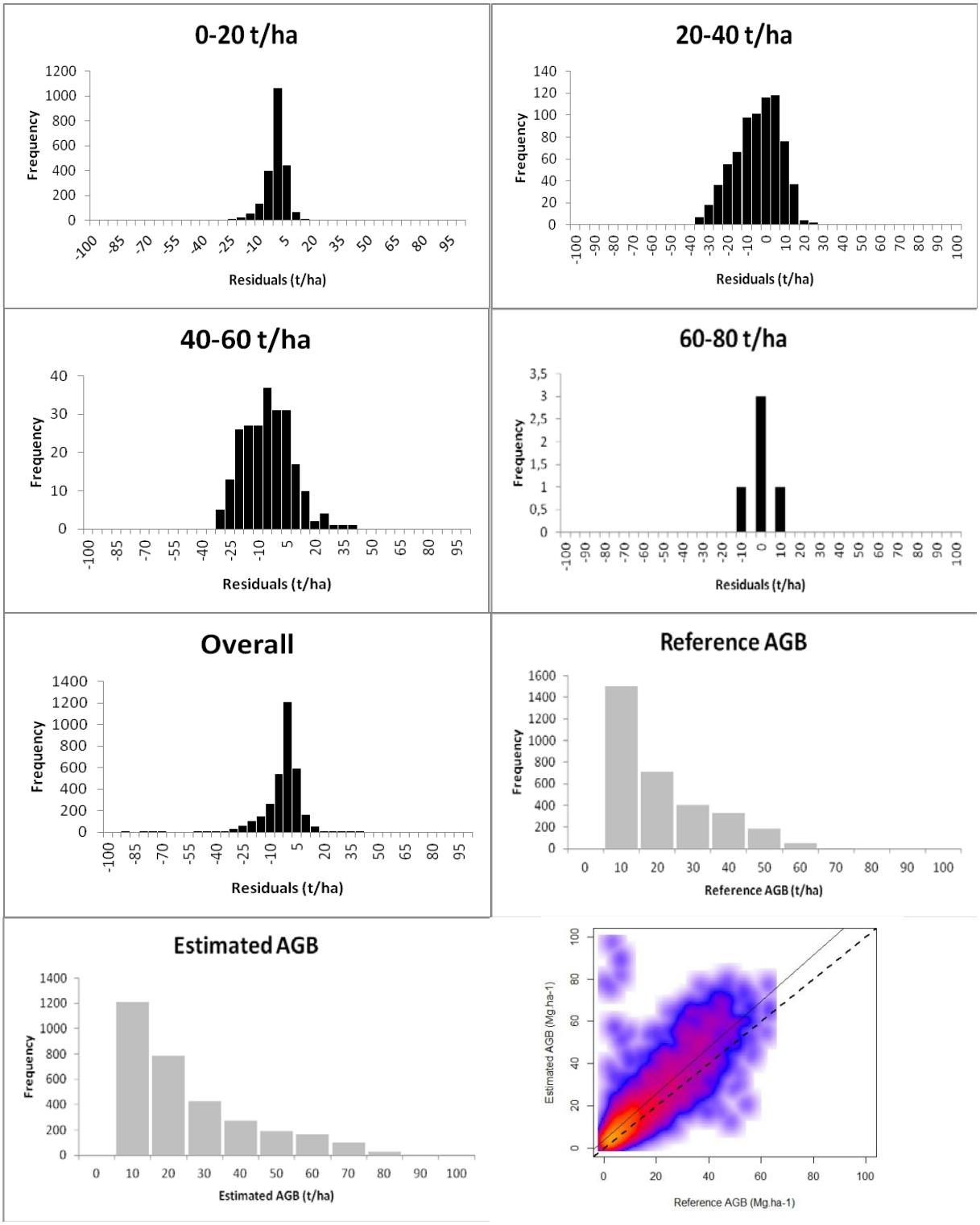


Figure 5-5b. Histograms of above-ground biomass (AGB) residuals (reference minus estimated) by AGB class (0-20 t/ha, 20-40 t/ha, 40-60 t/ha, and 60-80 t/ha) and overall, histograms of reference and estimated AGB data, and scatterplot of estimated against reference AGB values for epoch 2010 in South Africa.

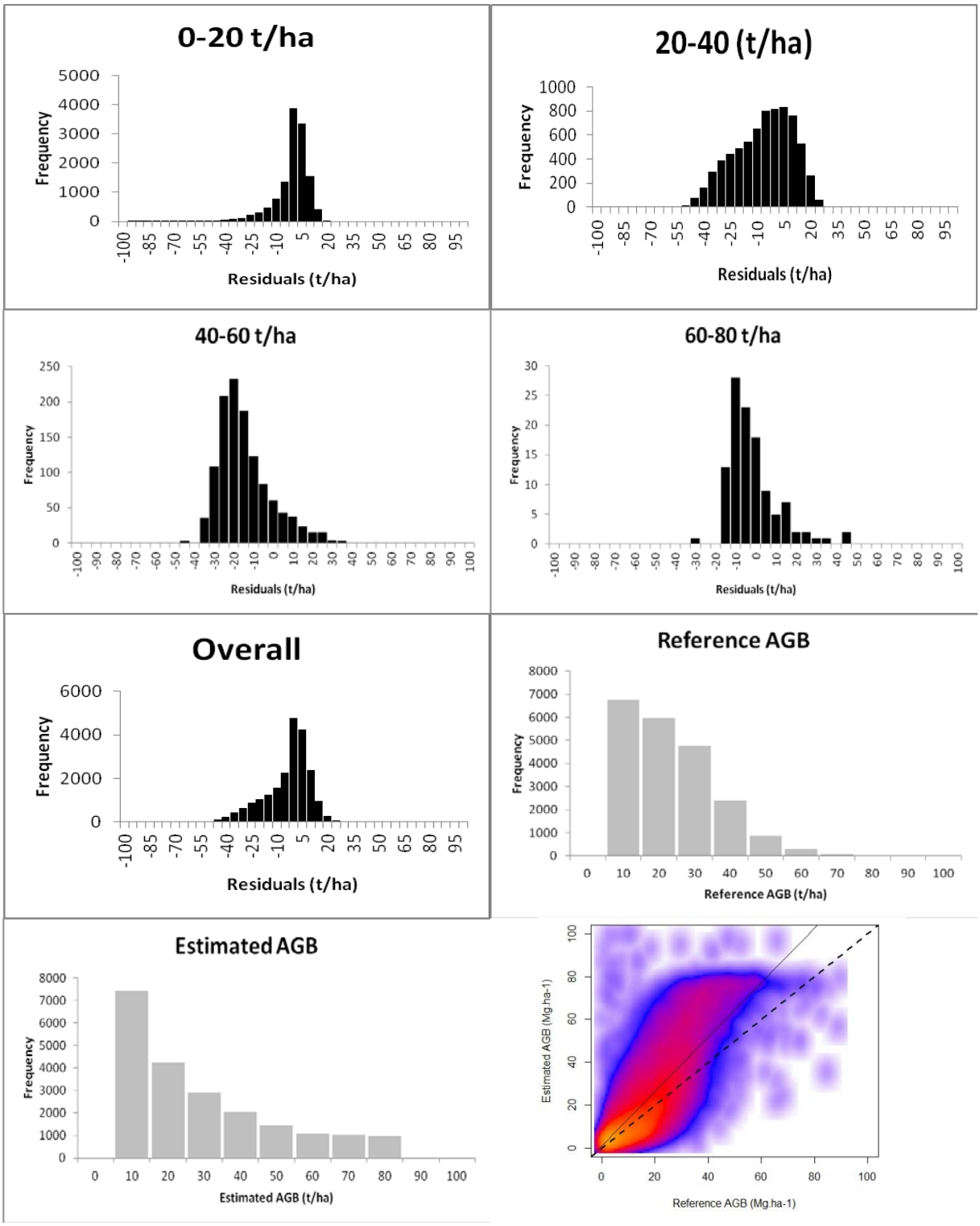


Figure 5-5c. Histograms of above-ground biomass (AGB) residuals (reference minus estimated) by AGB class (0-20 t/ha, 20-40 t/ha, 40-60 t/ha, and 60-80 t/ha) and overall, histograms of reference and estimated AGB data, and scatterplot of estimated against reference AGB values for epoch 2015 in South Africa.

5.7 Summary and conclusions

The analysis in this section has revealed important properties of the biomass inversions. Common to all regions is that lower biomass values (below around 100 t/ha) are overestimated. Although direct regression is not being used, this means that the procedure producing a fit to the backscatter-biomass relation has an intercept at zero biomass that is too large and/or the gradient of the fitting curve (i.e. $d\gamma^o/dB$, where γ^o is the backscattering coefficient and B is the biomass) is too large for lower biomass values. For higher biomass values, saturation of the L-band signal gives low sensitivity to biomass and leads to underestimation of the true biomass in the methods used in Poland, Sweden and Mexico. The Kalimantan inversions show little bias in the higher biomass range, but this seems to be because the fitting is largely determined by the larger number of points at high biomass. This will tend to force the mean values of the reference and estimated biomass to be close together for high biomass, giving low bias. The low sensitivity to biomass however gives rise to large scatter about the true value. For South Africa there are no high values of biomass.

Analysis of the structure of the errors reveals that for Poland the major errors, as expressed in RMSE, come from bias, i.e. the fitting curve is not a good representation of the true biomass-backscatter relation. This means that there is scope to improve the Polish results by use of an improved fitting curve. For other regions the problem is mainly due to scatter in inversions, i.e. although the sensitivity to biomass is being properly represented, the inversions are too noisy. It is not obvious how to reduce this noise, except perhaps by reducing the spatial resolution. This may be effective if the noise is due to the system, e.g. arises from speckle. It should also be stressed that apparent errors may arise from inaccuracies in the ground data or, for example, geolocation errors, so that the reference data and inversions are spatially mismatched.

It is clear from the findings above that overall RMSE (which varies between 33% and 75%) or relative bias gives little information on the true accuracy of the inversions and how they vary with biomass. Relative and absolute RMSE can vary widely with biomass level, but in many cases reported above relative RMSE exceeds 30% in all biomass ranges.

6 Regional Biomass Change Mapping

6.1 Introduction

When trying to detect and measure biomass change in a sequence of images, five types of change need to be considered:

1. Random variation due to system noise and, for radar images, speckle
2. Environmental changes that affect biomass estimates
3. Sudden high-intensity change due to deforestation and fire
4. Lower intensity, possibly progressive change caused by forest degradation
5. Biomass changes due to forest growth

The first two are nuisance factors when we are trying to measure changes in biomass. The third involves almost complete loss of biomass, so only requires an accurate estimate of biomass before the change event. The fourth is much harder to detect and measure, and there are few convincing studies that show this is possible except at low levels of accuracy, particularly in moderate to high biomass forests. The fifth is also hard to estimate accurately except in low biomass areas with rapid regrowth (such as regenerating forest in the tropics or short cycle plantations). So overall, except in case 3, measuring biomass change is a difficult problem.

Two routes are commonly pursued when investigating change in EO data:

1. Transform the data to a "product", e.g. land cover or biomass maps, and assess differences between the products
2. Detect change in the original data and then interpret it in terms of the desired quantities

Differences between products: This approach is straightforward and provides the Maximum Likelihood Estimate of the change in biomass (for unbiased products whatever the distributions of the random error in these products) but suffers from the fact that if the errors in the products are assumed to be statistically independent, then a pixel-wise difference will have a variance that is the sum of the variances of the errors in the individual products. As can be seen from the analysis in Section 5, this would lead to very large errors in the change product. Note that the above applies to the random error component in the products, but Section 5 shows clearly that bias is also an important component of the overall error. For the difference of two products, the pixel-wise bias (which, as have seen, is typically dependent on the true value of AGB) will be the difference of the individual biases. This could have the effect of compensating for bias errors in the individual products, and its importance depends on the extent to which errors in the products are dominated by bias (as for Poland and Sweden) or by random error (all other regions).

Note that the requirement for the biomass estimates was that they should meet a relative accuracy of 20% or better, and the brief analysis below discusses the implications of this for estimating the accuracy with which we can measure biomass change. It also assumes unbiased inversions. However, the analysis in Section 5 indicates that for most regions both assumptions are invalid. Bias is a major source of error for many regions and the precision being obtained in the inversions, measured as the standard deviation of the error, is roughly independent of biomass, i.e., it is not a relative precision but an absolute precision. Once the bias is removed, by methods still to be defined, the biomass–independent

random error will remain as the control on accuracy. Hence the analysis below will need to be modified once the bias-corrected inversions become available and the error analysis is redone.

Assigning an accuracy to a difference product becomes difficult when the accuracy of the individual products is only described as a *relative* accuracy, since then the variance is dependent on the true value of AGB, and the variance of the difference depends on the accuracy of both of the products (so cannot be characterised by a single number). This is explained further by the simple analysis below.

For the purpose of this analysis we assume that the AGB product is unbiased, and the target relative accuracy is *p*% at each pixel. The *p*% therefore corresponds to the precision of the measurement in the presence of random disturbances. For simplicity, the measure of precision we use here is the standard deviation (SD) of the biomass, rather than, for example, a Confidence Interval. Hence the SD of the estimates of a true AGB value *B* is *pB*, where *p* = 0.2 for 20% relative accuracy.

A further simplifying (but probably incorrect) assumption is that the estimates are normally distributed about the true value (this assumption can be assessed by examining the histograms of residuals, i.e. errors, in Section 5). Then if we have 2 estimates, $\hat{B}_1$ and $\hat{B}_2$, of AGB at a pixel at times sufficiently separated that the random errors decorrelate, their difference $\hat{B}_1 - \hat{B}_2$ is zero-mean normally distributed with variance $p^2 (B_1^2 + B_2^2)$, where B_1 and B_2 are the true AGB values and the relative accuracy, i.e. the SD/ mean, is:

$$\frac{p\sqrt{B_1^2 + B_2^2}}{B_1 - B_2}$$

This depends not only on $B_1 - B_2$ but also on at least one of the individual biomass values. The error properties of this difference are still being developed, and will form part of the final version of the ATBD.

Note that this approach to measuring change, possibly with some refinements, may be the only option unless the AGB estimator used to form the individual products is based on single data-types with well-characterised statistics.

Change detection in the data followed by interpretation: In this approach change is detected in the original time series of data and then interpreted in terms of biomass change (this type of approach is used in generating the MODIS burnt area product). Its first requirement is a well-founded way to detect change in a time series of data (which may have data only for two times). This requires the datasets used at each individual time to have known statistical properties (and if more than one dataset is being used at each time, their joint distribution is needed). In the context of GlobBiomass, this approach is most obviously applicable when the detection is based entirely on radar images, which have well characterised statistical properties. Such an approach is described in Mermoz et al (2016) and is proposed for change detection in the S. African region.

6.2 Poland

The AGB change maps were generated based on the subtraction of both AGB maps taking the RMSE for each AGB range (0-50 t/ha, 50-100 t/ha, 100-150 t/ha, 150-200 t/ha, 200-250 t/ha, and >250t/ha) into account (Section 5.4, Table 5-1).

For each pixel (AGB extimate) the new value is calculated by adding and subtracting the RMSE from the AGB estimate; range = [AGB_{est}-RMSE ; AGB_{est}+RMSE]. This is done for each pixel in both epochs. This results in calculating two additional raster layers:

$AGB_{est,min}$= AGB_{est}-RMSE

$AGB_{est,max}$= AGB_{est}+RMSE

The pixel is flagged as change when there is no overlap of the value ranges of the AGB estimates in the two epochs. The AGB increase and decrease are defined as follows:

- $AGB_{est,max,T1} < AGB_{est,min,T2}$ (increase)
- $AGB_{est,min,T1} > AGB_{est,max,T2}$(decrease)

Based on this assumption the change/no change masks (direction of the change - gain/loss) in binary and quantitative values were prepared for the period 2005-2010, 2010-2015 and 2005-2015. The MMU of 9 pixels was applied. The change maps are available at 25 meter spatial resolution (e.g., Figure 6-1).

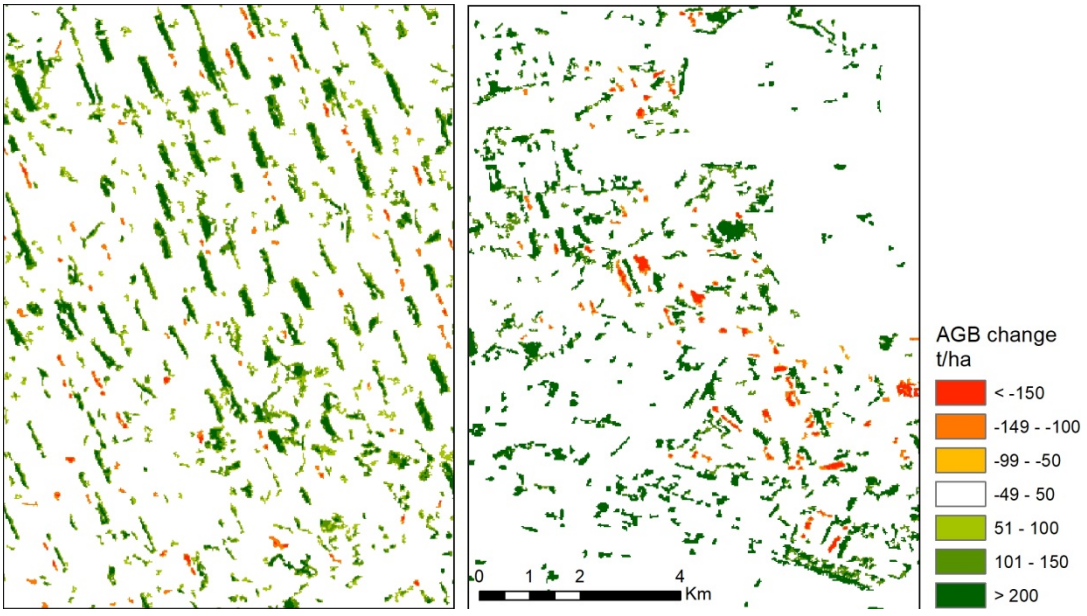


Figure 6-1. Example of the AGB change map (quantitative values) between 2005 and 2015.

6.3 Sweden

For biomass change, the difference between the merged biomass maps, between 2005 and 2010 (and possibly 2015, although it will be an entirely radar-based map) will be calculated. In addition, the

University of Chalmers will work on radar-based change algorithms for the years 2010 and 2015 in one or two smaller regions in Sweden (Remningstorp and Krycklan). In this case the approach used by CESBIO (Mermoz et al. 2016) may be appropriate.

6.4 Kalimantan (Indonesia)

The change between 2005 - 2010 and 2010 - 2015 was assessed using the RMSE in order to define a possible range for each estimate at pixel level. The RMSE is taken from the validation table, where the RMSE is calculated for five classes (0-50 t/ha, 50-100 t/ha, 100-150 t/ha, 150-200 t/ha and >200 t/ha). The RMSE value of the corresponding AGB range was subtracted and added to the estimated AGB on pixel basis in order to create two layers consisting the higher (H) and lower (L) possible values of the possible AGB range. In a next step, a change- no-change mask, with discrimination of increase and decrease, can be generated by using threshold values based on the calculated possible ranges. If Ht1 is higher than Lt2 we assume an overlap between the ranges and thus no-change. In cases where Ht1<Lt2, we suppose an increasing of the biomass and if Lt1>Ht2 biomass is decreasing between the two epochs, both indicates change (Figure 6-2). To remove areas smaller than nine pixels a minimum mapping unit of 3 x 3 pixels was applied to the final change map.

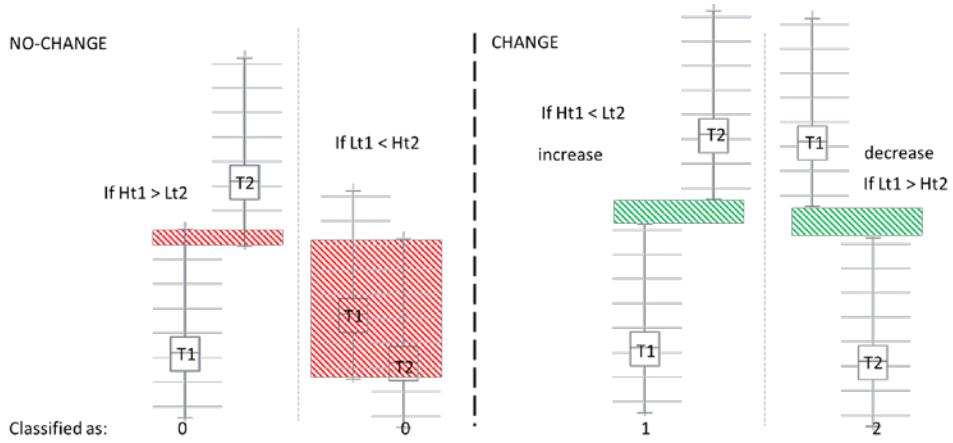
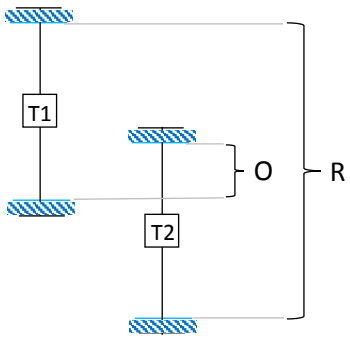


Figure 6-2. Schematic diagram of change and no-change approach.

The probability of change (0 - 1) was identified using a 95 % confidence interval. To achieve this, the AGB range is truncated by 2.5 % on both ends of the range (Figure 6-3). The overlap (O) of t1 and t2 divided by the range (R) of t1 and t2 shows the probability of no-change in a range between 0 - 1. Therefore, the probability of change is achieved by subtracting the probability of no-change from 1. If the probability of change is 1, then there is a change in the change- no-change mask.



$$Probability\ of\ change = 1 - O/R$$

Figure 6-3. Schematic diagram of creation of the probability of change layer (O = overlap, R = range).

6.5 Mexico

Methods

The AGB change maps will be generated based on the subtraction of both AGB epoch maps (AGBepoch2 – AGBepoch1) at pixel level, which will provide the amount and the direction of the change. Only forest AGB change is estimated.

Two sets of decision rules are generated to threshold those changes:

- Product 1 (based on independent validation per biomass range)
    1. The independent validation of both epochs combined per biomass range is used to define the standard deviation (SD) of the AGB map.
    2. AGB change occur under these conditions:
        - Significant AGB Loss: Loss is more than 2SD
        - Potential AGB Loss: Loss is more than 1SD
        - Significant AGB Gain: Gain is more than 2SD
        - Potential AGB Gain: Gain is more than 1SD
- Product 2 (based on uncertainty characterization at pixel level):
    1. The uncertainty layer generated for each epoch is used to define an interval around the AGB estimation (AGBepoch1 ± SDepoch1, and AGBepoch2 ± SDepoch2)
    2. AGB change is discarded if both intervals overlap. AGB change is only accepted if there is no overlap between the intervals. Estimation of AGBmax and AGBmin for each epoch based on uncertainty layer:
        - Significant AGB Loss: $AGB_{min\ epoch1} > AGB_{max\ epoch2}$
        - Potential AGB Loss: $AGB_{min\ epoch1} > AGB_{epoch2}$
        - Significant AGB Gain: $AGB_{max\ epoch1} < AGB_{min\ epoch2}$
        - Potential AGB Gain: $AGB_{max\ epoch1} < AGB_{epoch2}$

Validation

Due to the lack of long-term AGB measurements on permanent plots in the study areas accessible for the GlobBiomass team, it will be very challenging to properly validate this type of product. The AGB change map products will be validated using high resolution imagery for different epochs where available. This will allow validation of the complete AGB loss (deforestation) and the AGB gain from previous non-forest cover conditions (afforestation/reforestation), but it might not allow the validation of AGB loss (degradation) or gain (growth) in stable forest.

6.6 South Africa

Changes in forest biomass between 2007, 2010 and 2015 will be mapped at 25/50m resolution by using change detection algorithms. The classical change-detection approach in SAR remote sensing involves using the ratio of the local means in the neighborhood of each pair of colocated pixels because of the multiplicative nature of speckle. This approach assumes that a change in the scene will appear as a modification of the local mean value of the image. This detection method is robust to speckle noise, but is limited to the comparison of first-order statistics. Bujor et al. (2004) studied higher-order statistics for change detection in SAR images and concluded that the ratio of means was useful for step changes and that the second- and third-order log-cumulants were useful for progressive changes appearing in consecutive images in multi-temporal series. Because we aim to detect larges changes at

three to five-year temporal frequency, we assume that a modification of the local mean value is sufficiently reliable to detect changes.

The most obvious approach for change detection is to threshold the ratio image. However, a problem associated with thresholding approaches is that it tends to lose detail because of the linearly-fixed boundaries. Fuzzy set theory is conceptually different from the conventional crisp set theory in which an element either belongs or does not belong to a set. In fuzzy set theory, objects can be assigned grades of membership in a fuzzy set from zero to one. The method developed in Mermoz et al. (2016), based on fuzzy expectation-maximisation method, will be tested over the South-African regional site.

6.7 Discussion

Currently biomass change detection algorithms are under development. The most mature is that proposed for South Africa using the methodology described in Mermoz et al. (2016), which is applicable to time series of SAR data. This method first detects change, then interprets it in terms of change in biomass, where the biomass estimates are derived from inversion of the radar data using a regression equation. This appears not to be well-matched to methods being used for other regions that are not regression-based. However, because the change detection and change interpretation steps are independent, the detection phase can still be used as a prelude to applying other methods to estimate biomass.

The advantage of this approach is that it is statistically well-founded, at least for the detection phase, because the statistical properties of SAR data are so well characterised. Other proposed methods are based on subtracting biomass products. This suffers from two problems: (1) the difference product will have large errors compared to the individual biomass maps; (2) the accuracy of the difference is hard to characterise if the accuracy of the individual maps is expressed in terms of relative error, as is currently the case for GlobBiomass. However, we are considering use of such methods, but with the addition of ancillary knowledge to improve performance.

7 Relation of regional methods and maps to global product

The regional methods are essentially of two types, data-driven (Poland, Sweden kNN classification with optical data, Indonesia, Mexico) and model-driven (Sweden with PALSAR data, South Africa). Because the global methods are designed to function without extensive sets of ground data, they are model-driven, and hence there is no functional relationship between them and the data-driven approaches. They use instead effectively the same methods as those described for Sweden (with PALSAR data) and South Africa (indeed, as written, the "global" algorithm using MIPERS is identical to that for South Africa). The main data requirements and methods underlying the regional and global products are summarised in Table 7-1, together with comments on issues concerning high biomass forests (the regional section of this table is given as Table 7-1).

One of the key functions of the regional studies is to provide the most accurate biomass maps possible for their study areas, using whatever data are available, in order to provide an assessment of the accuracy of the global maps. Hence the accuracies of the regional maps must be well defined, and it must be shown that they themselves are accurate (at some appropriate scale). The global product must also have a well-defined measure of accuracy, otherwise there will be no way to assess whether it differs locally from the regional products, i.e., two estimates of biomass can only be compared in terms of the expected variability inherent in each estimate.

Table 7-1. The primary methods and data requirements of the regional and global approaches, together with comments on issues concerning high biomass forests

Region	Primary Data	Primary Method	High-Biomass	Ground Data
Poland	PALSAR 25-m mosaics HH+HV (2009+2010), additionally ERS-2 + ASAR	Random Forest, the whole country as an analysis unit	An implicit assumption is that Random Forest can overcome the known saturation issues around 100…150 m^3/ha even though the average stem volume is said to be up to 270 m^3/ha	Plots, r=12.6 m, screened to represent homogenous areas and aggregated
Sweden	Optical Spot4/5, PALSAR strips, 8+4 observations for 2014	kNN estimation with Spot 4/5 and NFI data + water cloud model and BIOMASAR approach with multi-temporal PALSAR data New Lidar data used to select the best alternative map	Saturation effects not commented but the model relies on stem volume of dense forest, which must be estimated for 30-km x 30-km processing tiles from external sources	National forest inventory plots with 7-m or 10-m radius
Indonesia	PALSAR 25-m mosaics HH+HV (2007, 2009+2010), additionally one ASAR/IM coverage 2004+2005	regression models, presumably the whole island as an analysis unit, ratio and texture bands at pixel spacing of 100 m $AGB = a_1 e^{v1} + a_2 e^{v2}…+c$ with so many input variables, over-fitting identified as a challenge	No saturation assumed based on the earlier work by the same researchers (and others but with lower biomass forest) but reference is made to (Mitchard et al 2011) where L-band saturation limit is stated to be around 200 tons/ha.	Plots, r=16…35 m
Mexico	PALSAR 25-m mosaics HH+HV, Landsat data and DEM data	MaxEnt for estimating probabilities of discrete biomass classes and the probability weighted average over the biomass classes, whole Mexico planned as an analysis unit	Above ground biomass over 200 tn/ha is assumed to be insignificant in Mexico	Plots from the Mexican national forest inventory, 1 ha sampled with 4 plots with a radius of 11.28 m
South Africa	PALSAR 25-m mosaics HH+HV for 2007 and 2010, PALSAR-1 dual-pol (strip?) data for 2015	Bayesian approach utilizing the water-cloud backscattering model	No saturation assumed in savanna areas, it remains unclear whether other forests are mapped or not	1-ha plots (37…120 of these)

Table 7-1. Cont.

Region	Primary Data	Primary Method	High-Biomass	Ground Data
Global	ASAR GM (+higher res in places) 1-km data	BIOMASAR	Tile-wise changing maximum GSV	MODIS Vegetation Continuous Fields, FAO GEZ strata
	PALSAR 25-m mosaics HH+HV for 2007, 2008, 2009, and 2010	BIOMASAR	Tile-wise changing maximum GSV	MODIS Veget. Continuous Fields, FAO GEZ strata, GLAS tree height
	PALSAR 25-m mosaics HH+HV for 2007, 2008, 2009, and 2010	CESBIO MIPERS, combination of regression models and BIOMASAR-style determination of water-cloud parameters	Bayesian multiplication by P(AGB)	1-ha plots, other more heterogeneous plots?
	ASAR and PALSAR generated biomass maps	Cubist, a machine-learning black box for refinement of special scale from ASAR 1-km map to PALSAR 25-m (…200-m) map	Assumed saturation-free ASAR 1-km biomass map	Land cover map (auxiliary layer)

8 Conclusions

1. The regional biomass mapping methods presented here reflect both the data available and the varying experiences of the project team. A key factor is the amount of *in situ* data available: where there are large amounts of *in situ* data from forest inventory, data-driven methods, such as kNN, can be developed, but otherwise (as for the global product) a model-based parametric approach is needed. Sweden provides an interesting test-case, since both methods can be applied here.

2. The key dataset in all the regional methods is ALOS-PALSAR (mosaics in all cases except Sweden, where strip data were used). These data are also of great importance to the global product since, although it already has a strong backbone provided by biomass maps of the northern boreal and temperate forests based on C-band data, PALSAR will allow biomass to be estimated at the pixel size (25m) of the PALSAR mosaics.

3. For the model-based approaches, information provided by optical sensors, such as land cover and forest density, is essential for parameterisation, and such data are also exploited in the Mexican product. DEM data from SRTM are also important in many cases, both to correct for terrain effects and to mask areas of steep terrain. The global product also uses a wide variety of auxiliary datasets.

4. An outstanding question is whether any of the approaches presented here can circumvent the well-known saturation of L-band data at higher biomass levels. This is particularly relevant for dense tropical forest, which is a key biome where good biomass information is needed. For this biome, a particular challenge for GlobBiomass is to demonstrate that its products are superior to the pan-tropical maps already available from Saatchi et al. (2011) and Baccini et al. (2012) (noting that improved maps are already available from Saatchi [private communication]).

5. Accuracy has two components, bias and precision. Underlying all the analysis of accuracy in this report is the assumption that the biomass estimates are unbiased, so that accuracy can be estimated in terms only of a zero-mean variation about the true value. (So, for example, the variance of the total error due to independent error contributions is given by summing the individual variances.) However, for signals that tend to saturate, as occurs with L-band data for large biomass, there is a serious risk that estimates will be biased over at least part of the biomass range. This will also occur if the fitting function used in regression does not capture the true form of the data over some or all of the biomass range. Unlike the zero-mean element (the precision), bias cannot be removed by spatial averaging and represents an intrinsic error in the measurements. Hence a crucial concern for all the products must be to assess whether the methodology is likely to lead to biases and whether such biases can be quantified and removed.

6. A further issue is the relation between accuracy and the scale at which biomass information is presented. Although PALSAR data offers the possibility to provide biomass estimates within 25m pixels, the useful measurement scale may be significantly larger. This is similar to the case of SAR: a single-look image offers the highest resolution, but for most purposes does not allow an accurate enough estimate of the true backscattering coefficient in a distributed target. Spatial averaging is usually required, so the effective resolution of the measurement may be significantly reduced. As an example, although the basic BIOMASS data will be 6-look have a ground resolution of ~50m, recovering biomass with an accuracy of 20% will only be possible at the scale of 4 ha (200m x 200m).

References

Askne, J., Dammert, P. B. G., Ulander, L. M. H., Smith, G., "C-band repeat-pass interferometric SAR observations of the forest," IEEE Transactions on Geoscience and Remote Sensing, vol. 35, no. 1, pp. 25-35, 1997.

Askne, J., Santoro, M., Smith, G., Fransson, J. E. S., "Multitemporal repeat-pass SAR interferometry of boreal forests," IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 7, pp. 1540-1550, 2003.

Askne, J. and Santoro, M. (2009), Automatic model-based estimation of boreal forest stem volume from repeat pass C-band InSAR coherence. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 513-516.

Askne, J. I. A., Fransson, J. E. S., Santoro, M., Soja, M. J. and Ulander, L. M. H. (2013), Model-based biomass estimation of a hemi-boreal forest from multitemporal TanDEM-X acquisitions. *Remote Sensing*, 5, 5574-5597.

Askne, J., Santoro, M., Smith, G., Fransson, J. E. S., "Multitemporal repeat-pass SAR interferometry of boreal forests," IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 7, pp. 1540-1550, 2003.

Askne, J., Santoro, M., "Experiences in boreal forest stem volume estimation from multitemporal C-band InSAR," in Recent Interferometry Applications in Topography and Astronomy, Padron, I., Ed.: InTech, ISBN: 978-953-51-0404-9, available from: http://www.intechopen.com/books/recent-interferometry-applications-in-topography-and-astronomy/experience-of-boreal-forest-stem-volume-estimation-from-multitemporal-c-band-insar, 2012.

Asner, G.P., Powell, G.V.N., Mascaro, J., Knapp, D.E., Clark, J.K., Jacobson, J., et al. 2010 High-resolution forest carbon stocks and emissions in the Amazon. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16738–16742.

Attema, E. P. W. and Ulaby, F. T. (1978), Vegetation modeled as a water cloud. *Radio Science*, 13, 357-364.

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Asthon, P., Banin, L., Bayol, N., Berry, N. J., Boeckx, P., de Jong, B. H. J., DeVries, B., Girardin, C., Kearsley, E., Lindsell, J., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E. T. A., Nagy, L., Qie, L., Quinones, M., Ryan, C. M., Slik, F., Sunderland, T. C. H., Vaglio Laurin, G., Valentini, R., Verbeeck, H., Wijaya, A. and Willcock, S. (2015), An integrated pan-tropical biomass map using multiple reference datasets. *Global Change Biology*.

Avtar, R., Suzuki, R., Takeuchi, W. and Sawada, H. (2013), PALSAR 50 m mosaic data based national level biomass estimation in Cambodia for implementation of REDD+ mechanism. *PLoS ONE*, 8, e74807.

Baccini, A., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P. S. A., Dubayah, R., Friedl, M. A., Samanta, S. & Houghton, R. A. 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Clim. Change,* advance online publication, http://www.nature.com/nclimate/journal/vaop/ncurrent/abs/nclimate1354.html#supplementary-information, ISSN: 1758-6798.

Baker, J., & Luckman, A. (1999). Microwave observations of boreal forests in the NOPEX area of Sweden and a comparison with observations of a temperate plantation in the United Kingdom. *Agricultural and Forest Meteorology* **98–99**, 389–416.

Ballhorn, U., Jubanski, J., and Siegert, F. 2011 ICESat/GLAS Data as a measurement tool for peatland

topography and peat swamp forest biomass in Kalimantan, Indonesia. *Remote Sens.* **3**, 1957–1982.

Beaudoin, A., Bernier, P. Y., Guindon, L., Villemaire, P., Guo, X. J., Stinson, G., Bergeron, T., Magnussen, S. and Hall, R. J. (2014), Mapping attributes of Canada's forests at moderate resolution through *k*NN and MODIS imagery. *Canadian Journal of Forest Research*, 44, 521-532.

Blackard, J. A., Finco, M. V., Helmer, E. H., Holden, G. R., Hoppus, M. L., Jacobs, D. M., Lister, A.J., Moisen, G. G., Nelson, M. D., Riemann, R., Ruefenacht, B., Salajanu, D., Weyermann, D. L., Winterberger, K. C., Brandeis, T. J., Czaplewski, R. L., McRoberts, R. E., Patterson, P. L. and Tymcio, R. P. (2008), Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment*, 112, 1658-1677.

Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L. and Kimes, D. S. (2008), Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment*, 112, 3876-3890.

Bouvet et al. (2016). An African savanna biomass map at 25m resolution using ALOS PALSAR data. Remote Sensing of Environment. To be submitted.

Breiman, L., 2001, Random Forests, Machine Learning 45 (1): 5-32. doi:10.1023/A:1010933404324.

Brown, S. 1997 Estimation biomass and biomass change of tropical forests: a primer. *FAO For. Pap.* **134**, 1–55.

Bruniquel, J., & Lopes, A. (1997). Multi-variate optimal speckle reduction in SAR imagery. International Journal of Remote Sensing, 18(3), 603-627.

Bujor, F., Trouvé, E., Valet, L., Nicolas, J. M., & Rudant, J. P. (2004). Application of log-cumulants to the detection of spatiotemporal discontinuities in multitemporal SAR images. IEEE Transactions on Geoscience and Remote Sensing, 42(10), 2073-2084.

Carreiras, J. M. B., Melo, J. B. and Vasconcelos, M. J. (2013), Estimating the above-ground biomass in miombo savanna woodlands (Mozambique, East Africa) using L-band synthetic aperture radar data. *Remote Sensing*, 5, 1524-1548.

Cartus, O. (2010). Large area forest stem volume mapping using synergy of spaceborne interferometric radar and optical remote sensing: a case study of northeast China. *Dissertation, Friedrich-Schiller-University Jena, Identifier: urn:nbn:de:gbv:27-20100429-150007-0*.

Cartus, O., Santoro, M., Schmullius, C. and 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*, 115, 931-943.

Cartus, O., Santoro, M., Kellndorfer, J., "Mapping forest aboveground biomass in the Northeastern United States with ALOS PALSAR dual-polarization L-band," Remote Sensing of Environment, vol. 124, pp. 466-478, 2012.

Cartus, O., Kellndorfer, J., Walker, W., Franco, C., Bishop, J., Santos, L. & Fuentes, J. 2014. A National Detailed Map of Forest Aboveground Carbon Stocks in Mexico. *Remote Sensing,* 6**,** 5559-5588, ISSN: 2072-4292.

Castel, T., Beaudoin, A., Stach, N., Stussi, N., Le Toan, T. and Durand, P. (2001), Sensitivity of space-borne SAR data to forest parameters over sloping terrain. Theory and experiment. *International Journal of Remote Sensing*, 22, 2351-2376.

CCG, 2007. Canadian Digital Elevation Data, GeoBase, available at http://www.geobase.ca/geobase/en/index.html, last accessed December 2014.

Chauhan, N. S., Lang, R. H. and Ranson, K. J. (1991), Radar modeling of a boreal forest. *IEEE Transactions on Geoscience and Remote Sensing*, 29, 627-638.

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., & Perez, R. (2004). Error propagation and scaling for tropical forest biomass estimates. Philosophical Transactions of the Royal Society B: Biological Sciences, 359(1443), 409-420.

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., and Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, *145*(1), 87-99.

Chave, J., Condit, R., Lao, S., Caspersen, J. P., Foster, R. B. & Hubbell, S. P. 2003. Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. Journal of ecology, 91**,** 240-252, ISSN: 1365-2745.

Cho, Moses, Azong, et al, (2012). Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. *Remote Sensing of Environment* 125: 214-226.

CONAFOR (2012). Forest above-ground biomass stocks in Mexico for 2004-2012. Comision Nacional Forestal, Mexico.

Cordellier, M. & Pfenninger, M. 2009. Inferring the past to predict the future: climate modelling predictions and phylogeography for the freshwater gastropod Radix balthica (Pulmonata, Basommatophora). *Molecular Ecology,* 18**,** 534-544, http:/dx.doi.org/10.1111/j.1365-294X.2008.04042.x, ISSN: 1365-294X.

Couhert, A., (2009). Inversion Bayesienne en biomasse forestière des données de rétrodiffusion radar dans le cadre du projet de satellite BIOMASS. MSc thesis report, Université Paul Sabatier

de Ferranti, J., Digital Elevation Data, available at http://www.viewfinderpanoramas.org/dem3.html, last accessed on December 2014.

Della Pietra, S., Della Pietra, V. & Lafferty, J. 1997. Inducing features of random fields. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 19**,** 380-393, ISSN: 0162-8828.

Desnos, Y.-L., Buck, C., Guijarro, J., Suchail, J.-L., Torres, R. and Attema, E. (2000), ASAR - Envisat's Advanced Synthetic Aperture Radar. *ESA Bulletin*, 102, 91-100.

DiMiceli, C. M., Carroll, M. L., Sohlberg, R. A., Huang, C., Hansen, M. C. and Townshend, J. R. G., 2011. Annual Global Automated MODIS Vegetation Continuous Fields (MOD44B) at 250 m Spatial Resolution for Data Years Beginning Day 65, 2000 - 2010, Collection 5 Percent Tree Cover, University of Maryland, College Park, MD, USA,

Dobson, M. C., Ulaby, F. T., Le Toan, T., Beaudoin, A., Kasichke, E. S. and Christensen, N. (1992), Dependence of radar backscatter on coniferous forest biomass. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 412-416.

Du, L. and Lee, J. S. (1996), Fuzzy classification of earth terrain covers using complex polarimetric SAR data. *Internation Journal of Remote Sensing*, 17, 809-826.

Du, Y., Cihlar, J., Beaubien, J., & Latifovic, R. (2001). Radiometric normalization, compositing, and quality control for satellite high resolution image mosaics over large areas. Geoscience and Remote Sensing, IEEE Transactions on, 39(3), 623-634.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Simon, F., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., Zimmermann, N. E. & Araujo, M. 2006. Novel Methods Improve Prediction of Species' Distributions from

Occurrence Data. *Ecography,* 29**,** 129-151, http:/dx.doi.org/10.2307/3683475, ISSN: 09067590.

Elith, J. & Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics,* 40**,** 677.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. & Yates, C. J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions,* 17**,** 43-57, ISSN: 1472-4642.

Englhart, S., Jubanski, J., and Siegert, F. 2013 Quantifying Dynamics in Tropical Peat Swamp Forest Biomass with Multi-Temporal LiDAR Datasets. *Remote Sens.* **5**, 2368–2388.

FAO (2001), FRA 2000 - Global ecological zoning for the Global Forest Resources Assessment 2000. Final Report, Working Paper 56, Rome.

FAO, *Global Forest Resources Assessment 2005*, FAO Forestry Paper 147, Food and Agriculture Organization of the United Nations, Rome (2006).

FAO, *Global Forest Resources Assessment 2010*, FAO Forestry Paper 163, Food and Agriculture Organization of the United Nations, Rome (2010).

FAO 2015, Global Forest Resources Assessment 2015 – Country Report Poland, Rome, 2015.

FAO 2010a. FRA 2010, Informe Nacional Mexico. Rome: Departamento Forestal, Organización de las Naciones Unidas para la Agricultura y la Alimentación.

Fazakas, Z., Nilsson, M., Olsson, H., "Regional forest biomass and wood volume estimation using satellite data and ancillary data," Agricultural and Forest Meteorology, 98-99, pp.417–425, 1999.

Fithian, W. & Hastie, T. 2013. Finite-sample equivalence in statistical models for presence-only data. *The annals of applied statistics,* 7**,** 1917.

Fridman, J. et al., 2014. Adapting National Forest Inventories to changing requirements–the case of the Swedish National Forest Inventory at the turn of the 20th century. *Silva Fennica*, vol. 48, 3, pp. 1–29. Available at: http://www.silvafennica.fi/pdf/article1095.pdf.

Franke, J., Navratil, P., Keuck, V., Peterson, K., and Siegert, F. 2012 Monitoring Fire and Selective Logging Activities in Tropical Peat Swamp Forests. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **5**, 1811–1820.

Fransson, J. E. S., Israelsson, H., "Estimation of stem volume in boreal forests using ERS-1 C- and JERS-1 L-band SAR data," International Journal of Remote Sensing, vol. 20, no. 1, pp. 123-137, 1999.

Fransson, J. E. S., Santoro, M., "K&C Science Report - Phase 3, Advances in forestry applications using satellite ALOS PALSAR images," Scientific reports for K&C Phase 3, 2014.

Frey, O., Santoro, M., Werner, C. and Wegmuller, U. (2013), DEM-based SAR pixel-area estimation for enhanced geocoding refinement and radiometric normalization. *IEEE Geoscience and Remote Sensing Letters*, 10, 48-52.

Gallaun, H., Zanchi, G., Nabuurs, G. J., Hengeveld, G., Schardt, M. and Verkerk, P. J. (2010), EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. *Forest Ecology and Management*, 260, 252-261.

GCOS 2003. Second report on the Adequacy of the Global Observing System for Climate in Support of the UNFCCC, GCOS-82, WMO/TD no. 1143.

Gibbs, H. K. (2006), Olson's major world ecosystem complexes ranked by carbon in live vegetation: An updated database using the GLC2000 land cover product, available at http://cdiac.ornl.gov/epubs/ndp/ndp017/ndp017b.html.

Gillis, M. D., Omule, A. Y. and Brierley, T. (2005), Monitoring Canada's forests: The National Forest Inventory. *Forestry Chronicle*, 81, 214-221.

Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S. & Peterson, A. T. 2007. What matters for predicting the occurrences of trees: techniques, data, or species'characteristics? *Ecological Monographs,* 77**,** 615-630, ISSN: 0012-9615.

Hamdan, O., Khali Aziz, H., and Mohd Hasmadi, I. 2014 L-band ALOS PALSAR for biomass estimation of Matang Mangroves, Malaysia. *Remote Sens. Environ.* **155**, 69–78.

Hame, T. , Rauste, Y., Antropov, O., Ahola, H.A., Kilpi, J., 2013, Improved Mapping of Tropical Forests With Optical and SAR Imagery, Part II: Above Ground Biomass Estimation, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, FEBRUARY 2013.

Hansen, M., De Fries, R., Townshend, J. R., Carroll, M., Dimiceli, C., Sohlberg, R.,"2001 Percent Tree Cover, Vegetation Continuous Fields MOD44B," University of Maryland, 2003, http://glcf.umiacs.umd.edu/data/vcf/.

Hansen, M. C., De Fries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C. and Sohlberg, R. A. (2003), Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS Vegetation Continuous Field algorithm. *Earth Interactions*, 7, 1-15.

Hansen, M. C., Egorov, A., & Roy, D. (2011). Continuous fields of land cover for the conterminous United States using Landsat data : first results from the Web-Enabled Landsat Data (WELD) project. Remote Sensing Letters, (915935429).

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O. and Townshend, J. R. G. (2013), High-resolution global maps of 21-st century forest cover change. *Science*, 342, 850-853.

Haralick, R. M. Shanmugam, K. and Dinstein, I. H. "Textural Featuresfor Image Classification," Systems, Man and Cybernetics, IEEE Transactions on, vol. 3, pp. 610-621, 1973.

Harrell, P. A., Bourgeau-Chavez, L. L., Kasischke, E. S., French, N. H. F. and Christensen Jr., N. L. (1995), Sensitivity of ERS-1 and JERS-1 radar data to biomass and stand structure in Alaskan boreal forest. *Remote Sensing of Environment*, 54, 247-260.

Hastie, T., Tibshirani, R., Friedman, J. & Franklin, J. 2005. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer,* 27**,** 83-85, ISSN: 0343-6993

Hengl T, de Jesus JM, MacMillan RA, Batjes NH, Heuvelink GBM, et al. (2014**).** SoilGrids1km — Global Soil Information Based on Automated Mapping**.** *PLoS ONE 9*, e105992.

Hoekman, D. H. and Reiche, J. (2015), Multi-model radiometric slope correction of SAR image of complex terrain using a two-stage semi-empirical approach. *Remote Sensing of Environment*, 156, 1-10.

Hofton, M. A., Minister, J. B., and Blair, J. B.(2000). Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 1989-1996.

Houghton, R. A., Butman, D., Bunn, A. G., Krankina, O. N., Schlesinger, P. and Stone, T. A. (2007), Mapping Russian forest biomass with data from satellites and forest inventories. *Environmental Research Letters*, 2, 045032.

Hughes, R.F., Kauffman, J.B., and Jaramillo, V.J. 1999 Biomass, carbon, and nutrient dynamics of secondary forests in a humid tropical region of Mexico. *Ecology* **80**, 1892–1907.

Hernandez, P. A., Graham, C. H., Master, L. L. & Albert, D. L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography,* 29**,** 773-785, ISSN: 1600-0587.

Imhoff, M. L. (1995), Radar backscatter and biomass saturation: Ramifications for global biomass inventory. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 511-518.

INEGI 2009. Uso de suelo y vegetación. *In:* GEOGRAFIA, I. N. D. E. Y. (ed.). Mexico.

Instrukcja WISL wielkoobszarowej inwentaryzacji stanu lasu, Warszawa 2010.

IPCC (2006), *IPCC Guidelines for National Greenhouse Gas Inventories*: Institute for Global Environmental Strategies, Japan.

Jabłoński and Budniak, 2014, Szacowanie nadziemnej biomasy drzewnej lasów w Polsce na potrzeby sprawozdawczości EKG/FAO i UNFCCC (Estimating above-ground woody biomass of forests in Poland for UNECE/FAO and UNFCCC reporting (in Polish) for UNECE/FAO and UNFCCC reporting), Leśne Prace Badawcze, v.75 (3): 277-289.

Jaynes, E. T. 1957. Information Theory and Statistical Mechanics. *Physical Review,* 106**,** 620-630.

Jubanski, J., Ballhorn, U., Kronseder, K., and Siegert, F. 2013 Detection of large above-ground biomass variability in lowland forest ecosystems by airborne LiDAR. *Biogeosciences* **10**, 3917–3930.

Karam, M. A., Fung, A. K., Lang, R. H. and Chauhan, N. S. (1992), A microwave scattering model for layered vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 767-784.

Kellndorfer, J., Walker, W., LaPoint, E., Bishop, J., Cormier, T., Fiske, G., Hoppus, M., Kirsch, K. and Westfall, J., 2012. NACP Aboveground Biomass and Carbon Baseline Data (NBCD 2000), available at Available on-line at http://daac.ornl.gov from ORNL DAAC, Oak Ridge, Tennessee, U.S.A. http://dx.doi.org/10.3334/ORNLDAAC/1081,

Kindermann, G. E., McCallum, I., Fritz, S. and Obersteiner, M. (2008), A global forest growing stock, biomass and carbon map based on FAO statistics. *Silva Fennica*, 42, 387-396.

Kurum, M., Lang, R. H., O'Neill, P. E., Joseph, A. T., Jackson, T. J. and Cosh, M. H. (2009), L-Band radar estimation of forest attenuation for active/passive soil moisture inversion. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 3026-3040.

Kharouba, H. M., Algar, A. C. & KerR, J. T. 2009. Historically Calibrated Predictions of Butterfly Species' Range Shift Using Global Change as a Pseudo-Experiment. *Ecology,* 90**,** 2213-2222, http:/dx.doi.org/10.2307/25592737, ISSN: 00129658.

KOBiZE. 2014. Poland's National Inventory Report 2013. Greenhouse Gas Inventory for 1988–2012. Submission under the UN Framework Convention on Climate Change and its Kyoto Protocol. Warszawa.

Kronseder, K., Ballhorn, U., Böhm, V., and Siegert, F. 2012 Above ground biomass estimation across forest types at different degradation levels in Central Kalimantan using LiDAR data. *Int. J. Appl. Earth Obs. Geoinf.* **18**, 37–48.

Kurvonen, L., Pulliainen, J. and Hallikainen, M. (1999), Retrieval of biomass in boreal forests from multitemporal ERS-1 and JERS-1 SAR images. IEEE Transactions on Geoscience and Remote Sensing, 37, 198-205.

Lantmäteriet (2010), Produktbeskrivning: GSD-Höjddata - grid 50+ hdb, 2015-07-01, version 2.2 (in Swedish).

Le Toan, T., Beaudoin, A., Riom, J. and Guyon, D. (1992), Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 403-411.

Lefsky, M., Keller, M., Pang, Y., De Camargo, P. B., & Hunter, M. O. (2007). Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. Journal of Applied Remote Sensing, 1(1), 013537.

Lefsky, M. A. (2010), A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters*, 37, L15401.

Li, W. & Guo, Q. 2010. A maximum entropy approach to one-class classification of remote sensing imagery. *International Journal of Remote Sensing,* 31**,** 2227-2235, http:/dx.doi.org/10.1080/01431161003702245, ISSN: 0143-1161.

Lopes, A., Nezry, E. and Touzi, R. (1990), Adaptive filters and scene heterogeneity. *IEEE Transactions on Geoscience and Remote Sensing*, 28, 992-1000.

Los, S. O., Rosette, J., Kljun, N., North, P. R. J., Chasmer, L., Suárez, J. C., … Berni, J. a J. (2012). Vegetation height and cover fraction between 60° S and 60° N from ICESat GLAS data. Geoscientific Model Development, 5, 413–432.

Lucas, R., Armston, J., Fairfax, R., Fensham, R., Accad, A., Carreiras, J., Kelley, J., Bunting, P., Clewley, D., Bray, S., Metcalf, D., Dwyer, J., Bowen, M., Eyre, T., Laidlaw, M. and Shimada, M. (2010), An evaluation of the ALOS PALSAR L-band backscatter - Above ground biomass relationship Queensland, Australia: Impacts of surface moisture condition and vegetation structure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3, 576-593.

McRoberts, R.E. et al., "Using remotely sensed data to construct and assess forest attribute maps and related spatial products," Scandinavian Journal of Forest Research, vol. 25, no. 4, pp.340–367, 2010.

MacKinnon, K., Hatta, G., Halim, H., and Mangalik, A. 1996 The Ecology of Kalimantan - Indodesian Borneo Periplus Editions (HK) Ltd., Dalhousie University.

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., Gao, F., Kutler, J. & Lim, T.-K. 2006. A Landsat surface reflectance dataset for North America, 1990-2000. *Geoscience and Remote Sensing Letters, IEEE,* 3**,** 68-72, ISSN: 1545-598X.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. a., Cohen, W. B., and McKee, W. A. (1999). Use of large-footprint scanning airborne Lidar to estimate forest stand characteristics in the western cascades of Oregon. *Remote Sensing of Environment* 67, 298-308.

Mermoz, S., Le Toan, T., Villard, L., Réjou-Méchain, M., & Seifert-Granzin, J. (2014). Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote Sensing of Environment*, *155*, 109-119.

Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V., & Gourlet-Fleury, S. (2015). Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sensing of Environment, 159*, 307-317.

Mermoz, S., Le Toan, T., Villard, L., Réjou-Méchain, M., & Seifert-Granzin, J. (2014). Biomass assessment in the Cameroon savanna using ALOS PALSAR data. Remote Sensing of Environment, 155, 109-119.Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V., & Gourlet-Fleury, S. (2015). Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sensing of Environment, 159*, 307-317.

Mermoz, S., Le Toan, T., 2016. Forest Disturbances and Regrowth Assessment Using ALOS PALSAR Data from 2007 to 2010 in Vietnam, Cambodia and Lao PDR. Remote Sensing 8, 217. doi:10.3390/rs8030217

Mitchard, E. T. A., Saatchi, S. S., Lewis, S. L., Feldpausch, T. R., Woodhouse, I. H., Sonké, B., Rowland, C. & Meir, P. 2011. Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest–savanna boundary region of central Africa using multi-temporal L-band radar backscatter. *Remote Sensing of Environment,* 115**,** 2861-2873, http:/dx.doi.org/10.1016/j.rse.2010.02.022, ISSN: 0034-4257

Mitchard, E. T. A., Saatchi, S. S., Baccini, A., Asner, G. P., Goetz, S. J., Harris, N. L. and Brown, S. (2013), Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps. *Carbon Balance and Management*, 8:10.

Mitchard, E. T. A., and Coauthors (2014). Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Global Ecology & Biogeography*, **23**, Issue 8, 836–955.

Moeliono, M., Wollenberg, E., Limberg, G., and eds. 2009 The decentralization of forest governance: politics, economics and the fight for control of forests in Indonesian Borneo.

Mograbi, P. J., Erasmus, B. F., Witkowski, E. T. F., Asner, G. P., Wessels, K. J., Mathieu, R., ... & Main, R. (2015). Biomass Increases Go under Cover: Woody Vegetation Dynamics in South African Rangelands.

Moisen, G. G., Frescino, T. S., Huang, C., Vogelmann, J. E. and Zhu, Z. (2003), Predictive modeling of forest cover and tree canopy height in the central Rocky Mountains of Utah. *Proceedings of 2003 Meeting of the American Society of Photogrammetry and Remote Sensing, Anchorage.*

Montesano, P. M., Nelson, R., Sun, G., Margolis, H., Kerber, A. and Ranson, K. J. (2009), MODIS tree cover validation for the circumpolar taiga-tundra transition zone. *Remote Sensing of Environment*, 113, 2130-2141.

Murray-Smith, C., Brummitt, N. A., Oliveira-Filho, A. T., Bachman, S., Moat, J., Lughadha, E. M. N. & Lucas, E. J. 2009. Plant Diversity Hotspots in the Atlantic Coastal Forests of Brazil. Sitios de Importancia para la Conservación de la Diversidad de Plantas en los Bosques de la Costa del Atlántico de Brasil. *Conservation Biology,* 23**,** 151-163, http:/dx.doi.org/10.1111/j.1523-1739.2008.01075.x, ISSN: 1523-1739.

Naidoo, Laven, et al, (2015). Savannah woody structure modelling and mapping using multi-frequency (X-, C-and L-band) Synthetic Aperture Radar data." *ISPRS Journal of Photogrammetry and Remote Sensing* 105: 234-250.

Nelson, R., Boudreau, J., Gregoire, T. G., Margolis, H., Næsset, E., Gobakken, T. and Ståhl, G. (2009), Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research*, 39, 862-881.

Nilsson, M., Olsson, H., 2010. Skogskarta. Available at: skogskarta.slu.se.

Notarnicola, C., & Posa, F. (2004). Bayesian algorithm for the estimation of the dielectric constant from active and passive remotely sensed data. IEEE Geoscience and Remote Sensing Letters, 1(3), 179-183.Quegan, S., Yu, J.J. 2001, Filtering of Multichannel SAR Images, IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 11, 2001

Oliver, C. and Quegan, S. 2004 Understanding Synthetic Aperture Radar Images.

Page, S.E., Siegert, F., Rieley, J.O., Boehm, H.D. V, Jaya, A., and Limin, S. 2002 The amount of carbon released from peat and forest fires in Indonesia during 1997. *Nature* **420**, 61–65.

Papathanassiou, K. P. and Cloude, S. R. (2001), Single-baseline polarimetric SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 2352-2363.

Pearson, R. G. 2010. Species' distribution modeling for conservation educators and practitioners. *Lessons in conservation,* 3**,** 54-89.

Pearson, R. G., Dawson, T. P., Berry, P. M. & Harrison, P. A. 2002. SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecological Modelling,* 154**,** 289-300, http:/dx.doi.org/http://dx.doi.org/10.1016/S0304-3800(02)00056-X, ISSN: 0304-3800.

Pearson, R. G., Dawson, T. P. & Liu, C. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography,* 27**,** 285-298, http:/dx.doi.org/10.1111/j.0906-7590.2004.03740.x, ISSN: 1600-0587.

Pearson, R. G., Raxworthy, C. J., Nakamura, M. & Townsend Peterson, A. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography,* 34**,** 102-117, ISSN: 1365-2699.

Penman J., Gytarsky M., Hiraishi T., Krug T., Kruger D., Pipatti R. et al. (red). 2003. IPCC Good Practice Guidance for Land Use, Land-Use Change and Forestry. IGES. Japan.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling,* 190**,** 231-259, http:/dx.doi.org/http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026, ISSN: 0304-3800.

Phillips, S. J., Dud, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. *Proceedings of the twenty-first international conference on Machine learning.* Banff, Alberta, Canada: ACM, http:/dx.doi.org/10.1145/1015330.1015412.

Picard, G., Le Toan, T., & Mattia, F. (2003). Understanding C-band radar backscatter from wheat canopy using a multiple-scattering coherent model. Geoscience and Remote Sensing, IEEE Transactions on, 41(7), 1583-1591.

Picard, G., Le Toan, T., Quegan, S., Caraglio, Y. and Castel, T. (2004), Radiative transfer modeling of cross-polarized backscatter from a pine forest using the discrete ordinate and eigenvalue method. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1720-1730.

Pierce, L. E., Bergen, K., Dobson, M. C., & Ulaby, F. T. (1998). Multitemporal land-cover classification using SIR-C/X-SAR Imagery. *Remote Sensing of Environment* **64**, 20–33.

Potapov, P., Turubanova, S. A., Hansen, M. C., Adusei, B., Broich, M., Altstatt, A., Mane, L., Justice, C. O. (2012). Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM+ data. Remote Sensing of Environment, 122, 106-116.

Potter, L. and Lee, J. 1998 Tree planting in Indonesia: trends, impacts and directions Center for International Forestry Research (CIFOR).

Praks, J., Antropov, O. and Hallikainen, M. (2012), LIDAR-aided SAR interferometry studies in boreal forest: Scattering phase center and extinction coefficient at X- and L-band. *Remote Sensing*, 50, 3831-3843.

Pulliainen, J. T., Mikkelä, P. J., Hallikainen, M. T. and 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, 758-770.

Pulliainen, J. T., Heiska, K., Hyyppä, J. and Hallikainen, M. T. (1994), Backscattering properties of boreal forests at the C- and X-bands. *IEEE Transactions on Geoscience and Remote Sensing*, 32, 1041-1050.

Pulliainen, J. T., Kurvonen, L., Hallikainen, M. T., "Multitemporal behavior of L- and C-band SAR observations of boreal forests," IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 2, pp. 927-937, 1999.

Quegan, S., Le Toan, T., Yu, J.J., Ribbes, F., and Floury, N. 2000 Multitemporal ERS SAR analysis applied to forest mapping. *IEEE Trans. Geosci. Remote Sens.* **38**, 741–753.

Quegan, S., Yu, J.J. 2001, Filtering of Multichannel SAR Images, IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 11, 2001

Quinlan, J. R. (1993), *C4.5: Programs for machine learning*. San Francisco, CA: Morgan Kaufmann Publishers Inc., pp.

Rabus, B., Eineder, M., Roth, A. and Bamler, R. (2003), The Shuttle Radar Topography Mission - A new class of digital elevation models acquired by spaceborne SAR. *ISPRS Journal of Photogrammetry & Remote Sensing*, 57, 241-262.

Ranneby, B. et al., 1987. *Designing a new national forest survey for Sweden*, SLU. Studia Forestalia Suecica, 177, Faculty of Forestry, Swedish University of Agricultural Sciences, Uppsala, Sweden. Available at: http://pub.epsilon.slu.se/4634/.

Ranson, K. J. and Sun, G. (1994), Mapping biomass of a northern forest using multifrequency SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 32, 388-396.

Ranson, K. J., Saatchi, S. and Sun, G. (1995), Boreal forest ecosystem characterization with SIR-C/XSAR. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 867-876.

Ranson, K. J. and Sun, G. (2000), Effects of environmental conditions on boreal forest classification and biomass estimates with SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 1242-1252.

Reese, H., Nilsson, M., Granqvist Pahlén, T., Hagner, O., Joyce, S., Tingelöf, U., Egberth, M. and Olsson, H. (2003), Countrywide estimates of forest variables using satellite data and field data from the National Forest Inventory. *Ambio*, 32, 542-548.

Reineke, L.H. (1933). Perfecting a stand-density index for even-aged forest. *Journal of Agricultural Research* **46**, 627-638.

Réjou-Méchain, M., Muller-Landau, H. C., Detto, M., Thomas, S. C., Toan, T. L., Saatchi, S. S., & Brockelman, W. Y. (2014). Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. Biogeosciences Discussions, 11, 5711.

Renner, I. W. & Warton, D. I. 2013. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. *Biometrics,* 69**,** 274-281, http:/dx.doi.org/10.1111/j.1541-0420.2012.01824.x, ISSN: 1541-0420.

Rodriguez-Veiga, P. 2015. *Large-Scale Mapping of Forest Above-Ground Biomass Retrieval from Maximum Entropy using SAR and Optical Satellite Data.* PhD in Physical Geography, University of Leicester.

Rosenqvist, A., Shimada, M., Suzuki, S., Ohgushi, F., Tadono, T., Watanabe, M., Tsuzuku, K., Watanabe, T., Kamijo, S. and Aoki, E. (2014), Operational performance of the ALOS global systematic acquisition strategy and observation plans for ALOS-2 PALSAR-2. *Remote Sensing of Environment*, 155, 3-12.

Ruesch, A. and Gibbs, H. K. (2008), New IPCC Tier-1 global biomass carbon map for the year 2000, available at http://cdiac.ornl.gov/epubs/ndp/global_carbon/carbon_documentation.html.

Saatchi, S. S. and Moghaddam, M. (2000), Estimation of crown and stem water content and biomass of boreal forest using polarimetric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 697-709.

Saatchi, S. S., Houghton, R. A., Dos Santos Alvalá, R. C., Soares, J. V. and Yu, Y. (2007), Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology*, 13, 816-837.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M. and Morel, A. 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. Proceedings of the National Academy of Sciences, 108**,** 9899-9904, http:/dx.doi.org/10.1073/pnas.1019576108.

Saatchi, S., Halligan, K., Despain, D. G., & Crabtree, R. L. (2007). Estimation of forest fuel load from radar remote sensing. *Geoscience and Remote Sensing, IEEE Transactions on*, *45*(6), 1726-1740

Salas, W. A., Ducey, M. J., Rignot, E. and Skole, D. (2002), Assessment of JERS-1 SAR for monitoring secondary vegetation in Amazonia: I. Spatial and temporal variability in backscatter across chrono-sequence of secondary vegetation stands in Rondonia. *International Journal of Remote Sensing*, 23, 1357-1379.

Santoro, M., Askne, J., Smith, G., Fransson, J. E. S., "Stem volume retrieval in boreal forests from ERS-1/2 interferometry," Remote Sensing of Environment, vol. 81, no. 1, pp. 19-35, 2002.

Santoro, M., Eriksson, L., Askne, J., Schmullius, C., "Assessment of stand-wise stem volume retrieval in boreal forest from JERS-1 L-band SAR backscatter," International Journal of Remote Sensing, vol. 27, no. 16, pp. 3425-3454, 2006.

Santoro, M., Fransson, J. E. S., Eriksson, L. E. B., Magnusson, M., Ulander, L. M. H., Olsson, H., "Signatures of ALOS PALSAR L-band backscatter in Swedish forest," IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 12, pp. 4001-4019, 2009.

Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., Wegmüller, U., Wiesmann, A., "Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements," Remote Sensing of Environment, vol. 115, no. 2, pp. 490-507, 2011.

Santoro, M., Beaudoin, A., Beer, C., Cartus, O., Fransson, J. E. S., Hall, R. J., Pathe, C., Schmullius, C., Schepaschenko, D., Shvidenko, A., Thurner, M., Wegmüller, U., "Forest growing stock volume of the northern hemisphere: spatially explicit estimates for 2010 derived from Envisat ASAR," Remote Sensing of Environment, vol. 168, pp. 316-334, 2015a.

Santoro, M., Eriksson, L. E. B., Fransson, J. E. S., "Reviewing ALOS PALSAR backscatter observations for stem volume retrieval in Swedish forest," Remote Sensing, vol. 7, no. 4, pp. 4290-4317, 2015b.

Santoro, M. and Cartus, O. (2010), STSE-BIOMASAR: Validating a novel biomass retrieval algorithm based on hyper-temporal Wide-Swath and Global Monitoring Envisat ASAR datasets. Final Report, ESA ESRIN contract No. 21892/08/I-EC, 2010.

Santoro, M., Cartus, O., Fransson, J. E. S., Shvidenko, A., McCallum, I., Hall, R. J., Beaudoin, A., Beer, C. and Schmullius, C. (2013), Estimates of forest growing stock volume for Sweden, Central Siberia and Québec using Envisat Advanced Synthetic Aperture Radar backscatter data. *Remote Sensing*, 5, 4503-4532.

Santoro, M., Wegmuller, U., Fransson, J. E. S. and Schmullius, C. (2014), Regional mapping of forest growing stock volume with multitemporal ALOS PALSAR backscatter. *Proceedings of IGARSS'14, Quebec City, 13-18 July,* (pp. 2313-2316), IEEE Publications, Piscataway, NJ.

Santoro, M., Wegmuller, U., Lamarche, C., Bontemps, S., Defourny, P. and Arino, O. (2015c), Strengths and weaknesses of multi-year Envisat ASAR backscatter measurements to map permanent open water bodies at global scale. *Remote Sensing of Environment*, 171, 185-201.

Schepaschenko, D., Mc Callum, I., Shvidenko, A., Fritz, S., Kraxner, F. and Obersteiner, M. (2011), A new hybrid land cover dataset for Russia: A methodology for integrating statistics, remote sensing and *in situ* information. *Journal of Land Use Science*, 6, 245-259.

Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., Miller, C., Frankenberg, C., Hibbard, K. and Cox, P. (2015), Observing terrestrial ecosystems and the carbon cycle from space. Global Change Biology, 21, 1762–1776.

Semarnat, C., INEGI, INE, INIFAP 2004. Documento estratégico rector del Inventario Nacional Forestal y de Suelos

Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K. M., Channan, S., DiMiceli, C., Townshend, J. R., "Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error," International Journal of Digital Earth, vol. 6, pp. 427-448, 2013.

Sheen, D. R., Malinas, N. P., Kletzli, D. W., Lewis, T. B. and Roman, J. F. (1994), Foliage transmission measurements using a ground-based ultrawide band (300-1300 MHz) SAR system. *IEEE Transactions on Geoscience and Remote Sensing*, 32, 118-130.

Shimada, M. (2010), Ortho-rectification and slope correction of SAR data Using DEM and its accuracy evaluation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3, 657-671.

Shimada, M. & Isoguchi, O. 2002. JERS-1 SAR mosaics of Southeast Asia using calibrated path images. *International Journal of Remote Sensing,* 23**,** 1507-1526, ISSN: 0143-1161.

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R. & Lucas, R. 2014. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment,* 155**,** 13-31, http:/dx.doi.org/http://dx.doi.org/10.1016/j.rse.2014.04.014, ISSN: 0034-4257.

Shimada, M., & Ohtaki, T. (2010). Generating large-scale high-quality SAR mosaic datasets: Application to PALSAR data for global monitoring. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 3(4), 637-656.

Shimada, M., Isoguchi, O., Tadono, T., & Isono, K. (2009). PALSAR radiometric and geometric calibration. IEEE Transactions on Geoscience and Remote Sensing, 47(12), 3915-3932.

Shimada, M., & Ohtaki, T. (2010). Generating large-scale high-quality SAR mosaic datasets: Application to PALSAR data for global monitoring. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 3(4), 637-656.

Shinohara, H., Homma, T., Nohmi, H., Hirosawa, H. and Tagawa, T. (1992), Relation between L-band microwave penetration / backscattering characteristics and state of trees. *Proceedings of IGARSS'92, Houston, 26-29 May,* (pp. 539-541), IEEE Publications, Piscataway, NJ.

Siegert, F., Ruecker, G., Hinrichs, A., and Hoffmann, A.A. 2001 Increased damage from fires in logged forests during droughts caused by El Niño. *Nature* **414**, 437–440.

Simard, M., Pinto, N., Fisher, J. B. and Baccini, A. (2011), Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research - Biogeosciences*, 116, G04021.

Simons, H. (2001), FRA 2000. Global ecological zoning for the Global Forest Resources Assessment 2000. FRA Working Paper 56, FAO, Rome.

Skogsdata, “Aktuella uppgifter om de svenska skogarna från Riksskogstaxeringen,” Inst f. skoglig resurshushållning, Sveriges lantbruksuniversitet, Umeå (in Swedish), 2015.

Small, D. (2011), Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49, 3081-3093.

Tanase, M. A., Panciera, R., Lowell, K., Siyuan, T., Garcia-Martin, A., Walker, J. P., "Sensitivity of L-band radar backscatter to forest biomass in semiarid environments: A comparative analysis of

parametric and nonparametric models," IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 8, pp. 4671-4685, 2014.

Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation, SIAM.

Thapa, R.B., Watanabe, M., Motohka, T., and Shimada, M. 2015 Potential of high-resolution ALOS–PALSAR mosaic texture for aboveground forest carbon tracking in tropical region. *Remote Sens. Environ.* **160**, 122–133.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S. R. and Schmullius, C. (2014), Carbon stock and density of northern boreal and temperate forests. *Global Ecology and Biogeography*, **23** (3), 297-310.

Thurner, M., Beer, C., Carvalhais, N., Forkel, M., Santoro, M., Tum, M., Schmullius, C. (2016) Large-scale variation in forest carbon turnover rate is related to climate, *Nature Climate Change* (in press)

Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society. Series B (Methodological): 267-288.

Tomppo, E. et al. (2008), "Combining national forest inventory field plots and remote sensing data for forest databases," Remote Sensing of Environment, vol. 112, no. 5, pp. 1982–1999, 2008.

Tomppo, E., Katila, M., Mäkisara, K. and Peräsaari, J. (2013), The Multi-source National Forest Inventory of Finland – methods and results 2009. Metlan työraportteja / Working Papers of the Finnish Forest Research Institute, 273, 216 pages. ISBN 978-951-40-2428-3 (PDF), available at http://www.metla.fi/julkaisut/workingpapers/2013/mwp273.htm.

Twedt, D. J., Ayala, A. and Shickel, M. R. (2015), Leaf-on canopy closure in broadleaf deciduous forests predicted during winter. Forest Science, 61, 926-931.

Ulaby, F. T., Whitt, M. W. and Dobson, M. C. (1990), Measuring the propagation properties of a forest canopy using a polarimetric scatterometer. IEEE Transactions on Antennas and Propagation, 38, 251-258.

Ulander, L. M. (1996). Radiometric slope correction of synthetic-aperture radar images. Geoscience and Remote Sensing, IEEE Transactions on, 34(5), 1115-1122.

USGS, 2001. 7.5-Min DEM Native Format of the United States, available at http://www.webgis.com/, last accessed December 2014.

Villard, L. (2009). *Forward & Inverse Modeling for Synthetic Aperture Radar Observables in Bistatic Configuration. Applications in Forest Remote Sensing* (Doctoral dissertation, Institut Supérieur de l'Aéronautique et de l'Espace).

Walker, W. S., Kellndorfer, J. M., LaPoint, E., Hoppus, M. and Westfall, J. (2007), An empirical InSAR-optical fusion approach to mapping vegetation canopy height. *Remote Sensing of Environment*, 109, 482-499.

Wang, Y., Day, J. L. and Davis, F. W. (1998), Sensitivity of modeled C- and L-band radar backscatter to ground surface parameters in loblolly pine forest. *Remote Sensing of Environment*, 66, 331-342.

Way, J., Rignot, E. J. M., McDonald, K. C., Oren, R., Kwok, R., Bonan, G., Dobson, M. C., Viereck, L. A. and Roth, J. E. (1994), Evaluating the type and state of Alaska taiga forests with imaging radar for use in ecosystem models. *IEEE Transactions on Geoscience and Remote Sensing*, 32, 353-370.

Wegmüller, U., "Automated terrain corrected SAR geocoding", Proc. IGARSS'99, Hamburg, 28 June - 2 July, 1999.

Weisbin, C. R., W. Lincoln and S. Saatchi (2014). "A Systems Engineering Approach to Estimating Uncertainty in Above-Ground Biomass (AGB) Derived from Remote-Sensing Data." Systems Engineering **17**(3): 361-373.

Whitmore, T.C. 1984 Tropical rain forests of the Far East 2nd editio. Clarendon Press, Oxford.

Wiesmann, A., Wegmüller, U., Santoro, M., Strozzi, T. and Werner, C. (2004), Multi-temporal and multi-incidence angle ASAR Wide Swath data for land cover information. Proceedings of 4th International Symposium on Retrieval of Bio- and Geophysical Parameters from SAR Data for Land Applications, Innsbruck, 16-19 November, CD-ROM.Vepakomma, U., St-Onge, B., and Kneeshaw, D. 2011 Response of a boreal forest to canopy opening: assessing vertical and lateral tree growth with multi-temporal lidar data. Ecol. Appl. 21, 99–121

Wikström, P. et al., "The Heureka forestry decision support system: An overview," *Mathematical and Computational Forestry & Natural-Resource Sciences*, vol. 3, no. 2, pp.87–94, 2011.

Williams, J. N., C. Seo, J. Thorne, J. K. Nelson, S. Erwin, J. M. O'Brien and M. W. Schwartz (2009). "Using species distribution models to predict new occurrences for rare plants." Diversity and Distributions 15(4): 565-576.

Wollan, A. K., V. Bakkestuen, H. Kauserud, G. Gulden and R. Halvorsen (2008). "Modelling and predicting fungal distribution patterns using herbarium data." Journal of Biogeography 35(12): 2298-2310.