

Savannah woody structure modelling and mapping using multi-frequency (X-, C- and L-band) Synthetic Aperture Radar (SAR) data

Laven Naidoo^a, Renaud Mathieu^a, Russell Main^a, Waldo Kleynhans^b, Konrad Wessels^b,
Gregory Asner^c, Brigitte Leblon^d

^aEcosystem Earth Observation, Natural Resources and the Environment, CSIR, Pretoria,
South Africa

Corresponding author contact details: LNaidoo@csir.co.za; (+27)12 841 2233

^bRemote Sensing Unit, Meraka Institute, CSIR, Pretoria, South Africa

^cDepartment of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA

^dFaculty of Forestry and Environmental Management, University of New Brunswick,
Fredericton, Canada

Abstract

Structural parameters of the woody component in African savannahs provide estimates of carbon stocks that are vital to the understanding of fuelwood reserves, which is the primary source of energy for 90% of households in South Africa (80% in Sub-Saharan Africa) and are at risk of over utilisation. The woody component can be characterized by various quantifiable woody structural parameters, such as tree cover, tree height, above ground biomass (AGB) or canopy volume, each been useful for different purposes. In contrast to the limited spatial coverage of ground-based approaches, remote sensing also has the ability to sense the high spatio-temporal variability of e.g. woody canopy height, cover and biomass, as well as species diversity and phenological status – a defining but challenging set of characteristics typical of African savannahs. Active remote sensing systems (e.g. Light Detection and Ranging – LiDAR; Synthetic Aperture Radar - SAR), on the other hand, may be more effective in quantifying the savannah woody component because of their ability to sense within-canopy properties of the vegetation and its insensitivity to atmosphere and clouds and shadows. Additionally, the various components of a particular target's structure can be sensed differently with SAR depending on the frequency or wavelength of the sensor being utilized. This study sought to test and compare the accuracy of modelling, in a Random Forest machine learning environment, woody above ground biomass (AGB), canopy cover (CC) and total canopy volume (TCV) in South African savannahs using a combination of X-band (TerraSAR-X), C-band (RADARSAT-2) and L-band (ALOS PALSAR) radar datasets.

Training and validation data were derived from airborne LiDAR data to evaluate the SAR modelling accuracies. It was concluded that the L-band SAR frequency was more effective in the modelling of the CC ($R^2=0.77$), TCV ($R^2=0.79$) and AGB ($R^2=0.78$) metrics in Southern African savannahs than the shorter wavelengths (X- and C-band) both as individual and combined (X+C-band) datasets. The addition of the shortest wavelengths also did not assist in the overall reduction of prediction error across different vegetation conditions (e.g. dense forested conditions, the dense shrubby layer and sparsely vegetated conditions). Although the integration of all three frequencies (X+C+L-band) yielded the best overall results for all three metrics ($R^2=0.83$ for CC and AGB and $R^2=0.85$ for TCV), the improvements were noticeable but marginal in comparison to the L-band alone. The results, thus, do not warrant the acquisition of all three SAR frequency datasets for tree structure monitoring in this environment.

Keywords: *Woody structure, Savannahs, SAR, Multi-frequency, LiDAR, Random Forest*

1. Introduction - Background, Aims and Objectives

Structural parameters of the woody component in African savannahs provide estimates of carbon stocks that are vital to the understanding of fuelwood reserves, which is the primary source of energy for 90% of households in South Africa (80% in Sub-Saharan Africa) and are at risk of over utilisation (Wessels et al., 2011, Wessels et al., 2013). The woody component in African savannahs is an important physical attribute for many ecological processes and impacts the fire regime, vegetation production, nutrient and water cycles (Silva et al., 2001). The density of woody plants can also severely compromise the availability of grazing resources, valuable for livestock populations and related livelihoods, through bush encroachment (Wigley et al., 2009). Within the context of climate change, the sequestration of carbon by growing vegetation is a significant mechanism for the removal of CO₂ from the atmosphere (Falkowski et al., 2000, Viergever et al, 2008). Understanding how carbon is stored as carbon sinks in vegetative biomass and thus quantifying this standing biomass is central to the understanding of the global carbon cycle. Vegetation clearing (e.g. for cultivation) and degradation (e.g. for timber or fuelwood) and the burning of biomass, which are prevalent in developing regions and savannah woodlands of Southern Africa, can

alter carbon stocks and emissions (Falkowski et al., 2000, Viergever et al, 2008). Based on the important environmental implications revolving around woody vegetation, there are growing initiatives aiming at forest and woodland conservation that require its active inventorying, mapping and subsequent monitoring such as the Reducing Emissions from Deforestation and Forest Degradation programme (REDD+) (Corbera & Schroeder, 2011, Kanowski et al., 2011, Asner et al., 2013).

The woody component can be characterized by various quantifiable woody structural parameters, such as tree cover, tree height, above ground biomass (AGB) or canopy volume, each been useful for different purposes. AGB is defined as the mass of live or dead organic matter above the ground surface (excluding roots etc.) and is usually expressed in tonnes per unit hectare (Bombelli et al, 2009). Woody canopy cover (i.e. the percentage area occupied by woody canopy or CC) is a key parameter used in monitoring vegetation change and can be combined with tree height to estimate approximate AGB (Colgan et al., 2012). Lastly, total woody canopy volume (TCV) indicates the volume of vegetation present within the vertical profile and serves as an alternative proxy for biomass density and distribution. Further, these metrics, both 2D (CC) or 3D (TCV and AGB) in nature can provide useful information regarding the prediction of density, habitat requirements and biodiversity assessments for conservation (Bradbury et al., 2005, Mueller et al., 2010 and Jung et al., 2012).

Remote Sensing has been used in numerous studies as the preferred tool for quantifying and mapping woody structural features due mainly to its superior information gathering capabilities, wide spatial coverage, cost effectiveness and revisit capacity (Lu, 2006). In contrast to the limited spatial coverage of ground-based approaches, remote sensing also has the ability to sense the high spatio-temporal variability of e.g. woody canopy height, cover and biomass, as well as species diversity and phenological status – a defining but challenging set of characteristics typical of African savannahs (Cho et al., 2012; Archibald & Scholes, 2007; Mills et al., 2006). Woody structural parameters have been successfully mapped using passive optical data at fine and coarse spatial scales (Boggs, 2010; Castillo-Santiago et al, 2010) by making use of textural (i.e. the local variance of an image related to

its spatial resolution – Nichol and Sarker, 2011) and/or spectral (e.g. spectral vegetation indices related to vegetation structure – Johansen & Phinn, 2006) approaches. Passive optical data are, however, adversely affected by high spectral variation (i.e. change in spectral properties or character of a target) due to seasonal dynamics, clouds and haze (prevalent in the rainy season of African summers and veld fires in the dry winter) and shadow (resulting from terrain topography and tree canopies) at fine resolutions and mixed wood-grass pixels at the medium and coarser resolutions. Active remote sensing systems (e.g. Light Detection and Ranging – LiDAR; Synthetic Aperture Radar - SAR), on the other hand, may be more effective in quantifying the savannah woody component because of their ability to sense within-canopy properties of the vegetation and its insensitivity to atmosphere and clouds and shadows.

Airborne LiDAR systems provide high-resolution geo-located measurements of a tree's vertical structure (upper and lower storey) and the ground elevations beneath dense canopies. SAR systems provide backscatter measurements that are sensitive to forest spatial structure and standing woody biomass due to its sensitivity to canopy density and geometry (Sun et al, 2011, Mitchard et al, 2011). Airborne LiDAR provides detailed tree structural products but it relies on the availability of aircraft infrastructure, which is not always available in Africa. Satellite LiDAR is also currently not available. A SAR-based approach, on the other hand, offers an all-weather capacity to map relatively large extents of the woody component, which cannot be easily achieved with either airborne or satellite LiDAR (Mitchard et al, 2011).

Polarization (orientation of the emitted and received signal) and frequency of SAR data play important roles in sensing vegetation structure. Multi-polarized SAR systems (emitting and receiving in HH, HV, VH and/or VV with H referring to a horizontal wave orientation and V referring to a vertical wave orientation) allow the more complete characterisation of the scattering properties of ground targets which in turn, enables the extraction of greater structural information. For instance, HV or VH are better linked to canopy structure because of the volumetric water content in the canopies architecture (Schmullius & Evans, 1997) which brings about volumetric scattering within the canopy (i.e. “random” scatterers),

which tends to change the polarization of the emitted wave (e.g. H to V or V to H). The various components of a particular target's structure can be sensed differently with SAR depending on the frequency or wavelength of the sensor being utilized. For example when sensing vegetation, the signal of shorter SAR wavelengths (e.g. X-band and C-band) interact with the fine leaf and branch elements of the vegetation resulting in canopy level backscattering with limited signal penetration. The signal of longer SAR wavelengths (e.g. P-band and L-band), on the other hand, can penetrate deeper into the vegetation with backscatter resulting from signal interactions with larger vegetation elements such as major branches and trunks (Vollrath, 2010; Mitchard et al., 2009). Consequently, the L-band frequency has been proven in numerous studies to be the most preferred (Carreiras et al., 2013, Mitchard et al., 2012, Santos et al., 2002, Ryan et al., 2011) and the most effective (Lucas et al., 2006) in estimating woody structure, particularly AGB with a higher saturation level at 80-85 tonnes/ha compared to the shorter wavelengths, in forested and savannah woodland environments. However, since woodlands and savannahs possess a sporadic combination of fine and large woody elements within individual tree canopies and a heterogeneous distribution of large trees and smaller shrubs throughout the landscape we hypothesized that combining the capabilities of these different SAR frequencies under a multi-sensor approach may enhance the sensing of the savannah woody element (Schmullius & Evans, 1997). Various studies have 'fused' or integrated multiple SAR frequency and polarimetric datasets for modelling and mapping of tree structural attributes across various environments from the coniferous temperate forests of North America to mangrove forests and to the open-forest woodlands of Australia (Tsui et al., 2012; Mougín et al., 1999; Collins et al., 2009). Despite the success achieved in these various studies via combining different SAR wavelengths (according to Mougín et al., 1999 & Tsiu et al., 2012), the combined strength of both shorter and longer SAR frequency sensor technologies, however, have yet to be assessed in the heterogeneous and complex Southern African savannah environment.

This study sought to test and compare the accuracy of modelling woody above ground biomass (AGB), canopy cover (CC) and total canopy volume (TCV) in South African savannahs using a combination of X-band (TerraSAR-X), C-band (RADARSAT-2) and L-band (ALOS

PALSAR) radar datasets. Training and validation data were derived from airborne LiDAR data to evaluate the SAR modelling accuracies. The research questions were:

- 1) How do various SAR frequencies (X- or C- or L-band) perform in predicting woody structural parameters (CC, TCV & AGB) in southern African savannahs?
- 2) Does combining SAR backscatter through different frequency combinations or scenarios (X+C or X+L or C+L band or X+C+L-band) improve the predictions of the various woody structural parameters and by how much?

We hypothesized that the combination of shorter wavelength (~3cm X-band and ~5cm C-band) with longer wavelength (~15cm L-band) SAR datasets, in a modelling approach, will yield an improved assessment of woody structure based on the assumption that X- and C-band SAR signals interact with the finer woody structural constituents (e.g. leaves and finer branchlets typical of the shrubby/thicket layer) while the L-band SAR signal interact with the major tree structural components (e.g. trunk and main branches typical of forested areas).

- 3) Finally, through the examination of the patterns of the prediction error, within the landscape for the different SAR frequency models, can the hypothesis, proposed above, be confirmed?

2. Study Area

The Kruger National Park regional study area is located in the Lowveld region of north-eastern South Africa, within the savannah biome (31°00' to 31°50' Long E, 24°33' to 25°00' Lat S). The study area included portions of the southern Kruger National Park, the neighbouring Sabi Sands Private Game Reserve, and the densely populated Bushbuckridge Municipal District (BBR) (figure 1). The area is characterised by short, dry winters and a wet summer with an annual precipitation varying from 235mm and 1000mm, and is representative of southern Africa savannahs. This rainfall range, together with grazing pressures, fire, geology, mega-herbivore activity and anthropogenic use (fuelwood collection and bush clearing for cultivation) govern the vegetation structure present in this biome. The vegetation comprise particularly of Clay Thornbush, Mixed Bushveld and Sweet and Sour Lowveld Bushveld (Mucina and Rutherford, 2006). The woody vegetation in the

region is generally characterized as open forest with a canopy cover ranging from 20-60%, a predominant height range of 2 to 5m and biomass below 60T/ha (Mathieu et al., 2013). The Sabi Sands Wildetuyn consists of a group of private owners with a strong eco-tourism based approach to conservation with the Kruger National Park being more geared towards large-scale public conservation via the inclusion of large tracts of land for protection. The communal rangelands of BBR are primarily utilised for livestock ranching, fuelwood harvesting and various non-commercial farming practices (Wessels et al., 2011, Wessels et al., 2013). This study region was selected to represent the differences in the woody structure (e.g. riparian zones, dense shrubs, sparse tall trees etc.) and spatial patterns of the different land management and disturbance regimes (communal rangeland management, private game reserve and national park management), varying vegetation types (lowveld savannah and mixed forest fringe species) and geological substrates (granite and gabbro).

Insert Figure 1

3. Materials and Methodology

The general methodology sought to develop woody structural metric models between collected field data and airborne LiDAR data for detailed localised metric maps (25m spatial resolution to match the field data plots) (figure 5). These LiDAR derived metric products (CC, TCV and AGB) were then used as the ground truth for model up-scaling at the regional scale using multi-frequency SAR datasets (X-, C- and L-band). This was achieved by integrating the LiDAR and SAR datasets with the use of a sampling grid and the extracted values were subjected to modelling using the Random Forest algorithm. Different SAR frequencies were modelled in the form of various SAR frequency combination scenarios. The SAR-derived woody structural metrics were then validated using the LiDAR-derived counterparts (error statistics and distribution).

3.1 Remote sensing data

Five TerraSAR-X X-band dual-polarized (HH and HV), four RADARSAT-2 C-band quad-polarized (HH, VV, VH, and HV) and two ALOS PALSAR L-band dual-polarized (HH and HV) SAR datasets (summarized in table 1) were acquired for this research. Only dual polarized SAR data (HH and HV) was used because the HV polarization parameter is known to better model the structure of woody vegetation through volumetric backscatter interactions, while HH is also reported as been sensitive to structure although to a lesser extent than the cross-polarized band (Collins et al., 2009; Mitchard et al., 2009; Mathieu et al., 2013). Further, HH/HV was the common polarization configuration available for all three sensors. Winter seasonal SAR acquisitions were chosen because winter in the Lowveld is the dry season and exhibits the lowest level of moisture in the landscape. The tree leaves are off along with dry soil and dry grasses. This reduced the chance of interference of the SAR signal with variable moisture content while allowing a greater penetration of microwaves into the canopies. In the same region Mathieu et al. (2013) reported the best retrieval of woody structural parameters with RADARSAT-2 data acquired in winter. An extensive airborne LiDAR dataset (total coverage of c.a. 63000 ha) were acquired for this study (figure 1) by the Carnegie Airborne Observatory-2 AToMS sensor during April-May 2012. For our datasets, the LiDAR was operated at a pulse repetition frequency of 50 kHz with a 0.56m laser spot spacing and

an average point density of 6.4 points per m² from a flying altitude of 1000m above ground level (Asner et al., 2012).

3.2 Field data

Field data were collected in April – May, and November – December 2012 across 38 sampling sites (in figure 1). These sites provided ground truth data to model and validate the LiDAR derived woody structural metric products to be used to model the SAR-based woody structural metrics. Ground sampling sites were located to represent the diversity in woody structure of the different vegetation types, management regimes, and geological substrates mentioned above. Each site covered a 100m X 100m area and vegetation measurements were taken from four clustered 25m X 25m sampling plots (with minimum distance > 50m, identified from geostatistic range assessments, Wessels et al. 2011), located at each of the four corners of the site (Figure 2). The 100 x 100m sites were positioned using high resolution imagery from Google Earth as well as earlier LiDAR datasets acquired in 2008 – 2010 to ensure that they are representative of the surrounding landscape.

Field AGB estimates were derived from height and stem diameter measurements using an allometric biomass estimation equation (Colgan et al., 2013 – Appendix A). The allometric equation was developed following destructive harvesting of 17 savannah tree species present in the study area (N=707; R² = 0.98; relative RSE = 52%; ranging from 0.2 – 4531 kg) Colgan et al (2013). Tree height was measured using a height pole and Laser vertex/rangefinder, while stem diameter was measured using callipers and DBH tape. Stem diameter was measured at 10cm above the ground and for multi-stemmed plants every individual stem was measured as separate individuals (e.g. species such as *Dichrostachys cinerea*).

Due to logistical and time constraints associated with measuring every tree within the sample plot two main stem diameter ‘zones’ were identified inside the site to increase sampling efficiency while still yielding representative quantities of biomass estimates (figure 2). The first diameter zone was the 25m by 25m plot where all trees with a stem diameter

of 5cm and greater were recorded, provided that they had a height of 1.5m or greater, and the second diameter zone was a 10m by 10m area positioned at the inner corner of the 25m by 25m plot where all trees with a stem diameter between 3 and 5cm and greater than 1.5m were also recorded. This allowed catering for a few sites, mostly in the communal lands, where most of the AGB consisted of dense stands of multi-stemmed plants (coppicing) with small DBH (Matsika et al., 2012). A total of 152 25m X 25m biomass plots were sampled. Individual tree level AGB was derived using Colgan’s allometric equation (Colgan et al. 2013). AGB was then calculated for each diameter zone by summing the relevant tree level AGB values which was then subjected to particular AGB up-scaling factor (Appendix B). The complete plot level AGB was calculated by summing all the corrected AGB subtotals for the stem diameter zones.

One or two sampling plots were chosen for most sites for CC data collection – the north east 25m by 25m plot and/or the south west 25m by 25m plot (DHB zone 2 – figure 2). CC values were estimated following the vertical densitometer protocol (Stumpf 1993, Ko et al. 2009), conceptually a point intercept sampling approach, and one of the most time-efficient techniques to implement. The point intercept method is a small angle approach well suited to measure the vertical canopy cover – i.e. vertical projection of canopy foliage onto a horizontal surface –, and as such is the most directly comparable with cover derived from remote sensing imagery such as LiDAR (Fiala et al., 2006). The sampling procedure involved laying down transects along a fixed 25m measuring tape orientated from north to south and moving from west to east within the subplot at 2m increments (figure 2). Along these transects, canopy cover (presence/absence i.e. Y/N) was determined using a 5m pole placed vertically above each sampled points every 2m along the transects. For plot level canopy cover, in terms of percentage at the 25m by 25m scale, the CC presence and absence data were subjected to the formula below (equation 1):

$$\frac{\sum Y}{169} \times 100$$

Where Y represents the presence of cover data. The total number of sampling points in a 25m by 25m plot conducted at 2m sampling increments is 169. A total of 37 (25m by 25m) plots of CC were recorded during the field campaign.

Insert Figure 2

3.3 LiDAR data processing, woody structural metrics and validation

Two LiDAR datasets were utilised to derive the LiDAR tree structure metrics. For the first dataset, ~1m Digital Elevation Models (DEM) and top-of-canopy surface models (CSM) were created by processing the raw LiDAR point clouds according to the steps outlined in Asner et al. (2012). Canopy height models (CHM, pixel size of 1.12m) were computed by subtracting the DEM from the CSM. For the second dataset, the raw point cloud data were further processed to pseudo waveforms, in which the LiDAR hits or returns falling within a cube placed above the ground were binned into volumetric pixels (voxels of 5 X 5 horizontal X 1m vertical) and weighted relative to the total number of hits within the vertical column (the result – LiDAR slicer data) (Asner et al., 2009).

Three woody structural metrics were derived from the processed LiDAR datasets. The derivation of the three metrics excluded all woody vegetation below a height threshold of 0.5m as to exclude the grassy savannah component. The CAO LiDAR data were validated against field height measurements of approximately 800 trees. There was a strong relationship ($r^2 = 0.93$, $p < 0.001$) but a fraction of woody plants below 1.5-1.7m were not detected by the LiDAR (Wessels et al., 2011). This would introduce a source of error in the modelling process. However, since our objective was to investigate the potential contribution of short microwaves (X-band and/or C-band) in detecting the shrubby layer we still preferred to use a 0.5m height threshold over a higher height threshold at 1.5m. In addition, all metric products have been resampled and computed at the 25m spatial resolution to correspond with the ground data measurements (plot size of 25x25m) collected in the field for metric validation. These metrics are described in detail below:

- 1) Woody Canopy Cover (CC) is defined as the area vertically projected on a horizontal plane by woody plant canopies (Jennings et al., 1999). The metric was created by first applying a data mask to the LiDAR CHM image in order to create a spatial array of 0s (no woody canopy) and 1s (presence of a woody canopy). A percentage woody

cover distribution image (summing all the 1's and dividing by 625 and then percentage) was calculated at a spatial resolution of 25m. This metric was validated against the 37 25m by 25m CC ground truth plots (figure 3). Results yielded a strong, positive, unbiased relationship ($R^2=0.79$) with a low Root Mean Squared Error (RMSE) (12.4%) and Standard Error of Prediction (SEP) (23%).

Insert Figure 3

- 2) Total Canopy Volume (TCV) is a metric which approximates the area under the curve of the pseudo waveform (i.e. a plot displaying the LiDAR return frequency-by-height; Muss et al., 2011) and indicates the volume occupied by vegetation matter within the vertical profile. The metric was computed from the pseudo waveform LiDAR data (i.e. voxel) by the addition of the within-canopy LiDAR returns at different heights or slices (incrementally increasing by 1m) above 0.5m (Asner et al., 2009), and the value was converted to hectare. The TCV LiDAR metric was not validated with ground collected data as a suitable field sampling approach was yet to be defined for this type of savannah environment. However, in Mathieu et al. (2013), the TCV metric, in comparison to all the other metrics, was best correlated with RADARSAT-2 backscatter and was thus considered a suitable metric in this study.
- 3) Above ground woody biomass (AGB) is defined as the mass of live organic matter present above the ground surface (Bombelli et al, 2009) and is expressed in this study as tonnes per hectare (t/ha). The AGB LiDAR derived metric was modelled using a linear regression, ground estimated AGB (within 25m field plots) and a simple $H \times CC$ LiDAR metric (where H is the mean top-of-canopy height and CC is the canopy cover of a 25m pixel resolution) (Colgan et al., 2012). 65% of the 152 ground estimated AGB was used for model development while the remaining 35% was used for model validation. The validation results of ground versus LiDAR AGB (Figure 4) indicate a moderate positive correlation ($R^2=0.63$). With the use of allometric equations from Colgan et al. (2013) for ground AGB estimation, the RMSE (19.2t/ha) and SEP (63.8%) is, however, high with underestimation at high biomass levels by the LiDAR. Due to the intensive and time consuming nature of sampling these very high

biomass plots, an insufficient number of these plots may have been sampled to suitably train the model which thus led to such a deviation from the 1:1 line at the high biomass levels in figure 4. In the absence of better biomass estimates, the LiDAR derived AGB metric was deemed sufficient for the modelling and validation.

Insert Figure 4

3.4 SAR data and processing

The SAR imagery (X-, C- and L-band) were pre-processed according to the steps: multi-looking, radiometric calibration (conversion of raw digital numbers into σ^0 backscatter values), geocoding, topographic normalization of the backscatter and filtering. These steps were compiled in the form of scripts in GAMMA™ radar processing software for the Dual Pol. TerraSAR-X X-band (StripMap, Level 1b, MGD), Fine Quad Pol. RADARSAT-2 C-band (SLC) and Dual Pol. ALOS PALSAR L-band (Level 1.1) data. A 20m DEM and a 90m SRTM DEM were both used for the geocoding and orthorectification of the X-, C- and L-band SAR imagery. The 20m DEM was computed from South African 1:50 000 scale topographic maps (20m digital contours, spot-heights, coastline and inland water area data – ComputaMaps; www.computamaps.com) with RMS planimetric error of 15.24m and a total vertical RMS error of 6.8m. The 90m (3 arc sec) STRM DEM was gap-filled using Aster GDEM data and was derived from 20m interval contour lines extracted from 1:50 000 topographical maps. An automated hydrological correction was applied to correct inaccuracies along river lines and tributaries (Weepener et al., 2011). The multi-looking factors and filtering were chosen to best minimize the effect of speckle while not deteriorating the spatial detail captured by the sensors. 4:4, 1:5 and 2:8 range and azimuth multi-looking factors were implemented for the X-, C- and L-band datasets respectively. A Lee filter (3X3 filtering window) (Lee, 1980) was applied to the images. All datasets were resampled to a final spatial resolution of 12.5m to match the spatial resolution of the coarsest dataset (the L-band dataset).

Insert Table 1

3.5 Data integration, modelling protocols and mapping

Before modelling could be conducted the different datasets had to be processed to a common spatial grid. A sampling grid strategy was implemented as the relationship between dependent (LiDAR) and independent (SAR backscatter) datasets were not evident on a pixel-by-pixel basis mainly due to issues of SAR speckle and pixel-level inaccuracy of co-registration between datasets. This strategy also served as a means of extracting information from various remote sensing datasets of varying spatial resolutions (see table 1) without the need for pixel level fusion procedures. A regular spatial grid made up of 105m resolution cells at 50m distance spacing was applied. The choice of the cell size was informed by Mathieu et al. (2013), who tested various grid sizes ranging from 15m and 495m with RADARSAT-2 C-band data, and reported the 105m grid size as the resolution which provided the best trade-off between the finest spatial resolution/mapping scale and strongest correlation with the LiDAR woody structure parameters. Similar results (50-125m grid size) were reported with ALOS PALSAR L-band data in the region (Urbazaev, 2013). The 50m distance spacing between the grid cells was chosen to avoid autocorrelation effects arising from the inherent distribution of the vegetation structural parameters across the landscape (Wessels et al., 2011). Informal settlements, the main roads and water surfaces such as rivers and dams were masked and excluded from the analysis. Mean values within each cell were extracted for the SAR (X-HH, X-HV, C-HH, C-HV, L-HH and L-HV) and LiDAR metric datasets (CC, TCV and AGB). Due to the differences in spatial coverage of the multi-frequency SAR datasets in relation to the LiDAR coverage (figure 1), a varying number of data records (21170 records for X-band, 17980 records for C-band and 21467 records for L-band) were obtained during aggregation to the 105m grid. Various data mining, regression and machine learning algorithms (linear regression, support vector machines, REP decision trees, artificial neural network and random forest) were tested in Naidoo et al. (2014) and Random Forest (Breiman, 2003) was found to be the most robust and efficient, in terms of running time and accuracies (Prasad et al., 2006; Ismail et al., 2010). RF was applied (using Rattle data mining software) to the data with 35% of the data being used for model training and the remaining 65% being used for model validation. The random forest model (built on 'ntrees' = 500 and 'mtry' = $\sqrt{\#}$ SAR predictors and the trees were allowed to grow without pruning) was applied to the imagery using the combination of ModelMap, Random Forest

and GDAL modules in R statistical software. The maps were displayed in discrete class intervals (total of 6 classes) to best illustrate the tree structural metric distribution representative of the entire modelled ranges.

For the final stage of model validation, the correlation coefficient (R^2), Root Mean Squared Error (RMSE) and Standard Error of Prediction (SEP in % which also known as the Relative RMSE) were computed and the modelling algorithm accuracies were compared for the individual SAR scenarios. Seven modelling SAR scenarios (X-band only, C-band only, L-band only, X+C-band, X+L-band, C+L-band and X+C+L-band) were chosen to investigate the relationships between the individual SAR frequencies alone and different multi-frequency SAR combinations correlated against the three LiDAR metrics.

3.6 Error assessment

The purpose of this section was to investigate the error produced by the different SAR models under varying tree structural scenarios, and to ascertain whether spatial patterns in error were associated with specific vegetation structural cohort types (e.g. grassland versus woodland conditions etc.). Error maps were created by subtracting the LiDAR-derived and SAR-derived woody (i.e. LiDAR – SAR) CC structural metric maps. CC was chosen as TCV lacked meaningful units for interpretation and AGB displayed high error in the dense forest canopies (plots not displayed but supported by the error observed between the ground AGB and LiDAR derived AGB in figure 4, before AGB up-scaling to the SAR). We assessed the following main vegetation structural cohort types typical of savannah landscapes: low cover and variable tree height (e.g. sparse veld), high cover and high tree height (e.g. forests) and high cover and low tree height (e.g. bush encroaching shrubs). Box and whisker plots were created from the mean LiDAR-SAR difference values (i.e. prediction error), which were extracted from the same sampling (105m) grid used in the predictor variable extraction process, and interpreted. Similar error assessment analyses were conducted over different landscape geologies (e.g. granite versus gabbro) and topographic features (e.g. crest, slope

and valleys) but the error distribution patterns were fairly similar without any distinct patterns to comment on.

Insert Figure 5

4. Results

4.1 Modelling Accuracy Assessment

Insert Table 2

Insert Figure 6A-G

Table 2 illustrates the validation performances of the different SAR predictors, under various multi-frequency SAR scenarios, in predicting the three woody structural LiDAR metrics (CC, TCV and AGB). When examining the individual SAR frequency performances for modelling all three metrics, the longer wavelength L-band PALSAR predictors consistently yielded higher accuracies in comparison to the shorter wavelength predictors of both X-band TerraSAR-X and C-band Radarsat-2. The X-band TerraSAR-X predictors by far consistently produced the lowest modelling accuracies. The combination of the short wavelength SAR datasets (X- and C-band) improved the tree structural modelling over the individual dataset accuracies results but never produced accuracies greater than the use of the L-band dataset alone. The combined use of all three SAR frequencies (X-, C- and L-band) data in the modelling process consistently yielded the highest accuracies for modelling all three structural metrics [CC~ $R^2=0.83$, RMSE (SEP)=8.76% (25.40%); TCV~ $R^2=0.85$, RMSE (SEP)=16443.57 no unit (15.96%); AGB~ $R^2=0.83$, RMSE (SEP)=5.20 T/ha (29.18%)]. In comparison to the results for L-band alone, there was noticeable improvement in modelling accuracies [CC~ R^2 improvement=0.06, SEP absolute improvement AI (relative improvement RI)=4.2% (14.3%); TCV~ $R^2=0.06$, SEP AI (RI)=2.9% (15.5%); AGB~ $R^2=0.05$, SEP AI (RI)=3.7% (11.3%)] when the shorter wavelength datasets (X- and C-band) were added but the inclusion of the L-band frequency contributed the most to the overall accuracies. Overall, the three metrics were modelled at high accuracies under the multi-frequency scenario (X-, C- and L-band) and with similar patterns when considering the various individual scenarios.

Figures 6A-G illustrates, by way of the 1:1 line, the extent of over-prediction and under-prediction by the models which is gradually reduced towards the multi-frequency scenarios. The TCV results were chosen for representation in figures 6A-G as the metric yielded the

highest overall modelled accuracies and the remaining metrics (CC and AGB) displayed similar trends throughout the different SAR frequency combinations. For TCV (figures 6A-G), general over-prediction is observed at values less than ± 100000 (no unit) TCV while general under-prediction is observed at values greater than this threshold.

4.2 Tree Structure Metric and Error Maps

Insert Figure 7i-iii

All three metrics were mapped for the study area (Figure 7i-iii) using the multi-frequency SAR models (X+C+L-band). Figures 7(i-iii) illustrate the spatial distributions of AGB (figure 7i), TCV (figure 7ii) and CC (figure 7iii) which overall were very similar with high and low AGB and TCV regions coinciding with high and low CC. The spatial distribution of these metrics, coupled with the authors' knowledge and observations, will be elaborated upon in detail in the discussion section. Figure 8 shows the ABG vs. CC scatterplot. The point cloud generally displays a high correlation between the 2D (CC) and 3D (ABG) variable, but also a triangular shape with an increasing base as the CC increases up to 75%. Hence, dense cover conditions (CC>70%) are characterized by ABG values varying from moderate (35-40t/ha) to high (>60t/ha), corresponding to a range of tree sizes from coppicing thicket and medium sized tree bush encroachment to taller tree forests.

Insert Figure 8

Examples of CC error maps for dense forested ('A' figure 7iii) and sparse gabbro ('C') sites are presented in figures 9 and 10, respectively. Total error statistics were calculated to investigate the contributions of the four main SAR frequencies scenarios (X-band, C-band, L-band and X+C+L-band) to the modelling and mapping error (Table 3).

Insert Table 3

Insert Figure 9i-v

Insert Figure 10i-v

In table 3, there is a noticeable decline in major CC overestimation (<-15%) and major CC underestimation (>15%) with an increase in negligible CC error (5 to 15%) from shorter wavelengths (X-band to C-band) to the longer wavelength (L-band). The X+C+L-band combined scenario further reduced major CC overestimation and marginally increased negligible CC error but at the cost of an increase in major CC underestimation in comparison to the L-band results. More specifically, under dense forested conditions (figures 9i-v), the X-band scenario (figure 9i) illustrate major CC underestimation. C-band results (figure 9ii) indicate an overall decrease of patches of major CC underestimation but some of these have been replaced with major CC overestimation across less dense patches of large trees (see encircled area in figure 9ii). Further improvement is visible for the L-band scenario (figure 9iii) with a noticeable increase in the minor CC underestimation (5 to 15%) and negligible CC error (evident in table 3). Finally, the X+C+L scenario in figure 9iv illustrated noticeable increases in the negligible CC error coverage (especially over the dense ridge visible in the LiDAR CC of figure 9v) but also indicated an increase in major CC underestimation over dense vegetation patches north of the ridge (see encircle area in figure 9iv). Patches of major CC overestimation, however, still persist across riparian zones of minor tributaries. Under sparse vegetated conditions across gabbro intrusions (figures 10-i-v), however, X-band and C-band scenarios (figures 10i and 10ii) indicate vast extents of major CC overestimation for the areas sparse vegetation and major CC underestimation for the dense forested patches (see encircled area in figure 10ii). The L-band scenario (figure 10iii) illustrates a drastic improvement with an extensive increase in negligible CC error across the AOI. Across patches of dense vegetation, major CC underestimation still persists (similar to the trend in figure 9). The X+C+L-band scenario (figure 10iv) also yields favourable results similar to the L-band scenario with no visible improvement but did illustrate a slight increase in minor overestimation throughout the sparse vegetation landscape (also evident in table 3). More quantitative results (box-plots, figures 11i-ii) were introduced next to further assess the individual SAR frequency CC error contributions under different sparse and dense vegetation conditions.

Insert Figure 11i-ii

CC error boxplots of the four main SAR frequency scenarios, figure 11, were chosen to investigate error across vegetation structural types, classified from the LiDAR CHM, and including sparse shrubs (CC<40% and height<3m) or trees (CC<40% and height>3m) (figure 11i), and dense forested (CC>70% and height>3m) or bush encroached (CC>70% and height<3m) conditions (figure 11ii). In general, SAR derived CC is mostly overestimated across sparse vegetation but is underestimated across conditions of dense cover which coincides with the main trends of figures 9i-v and 10i-v. The L-band scenario yielded the lowest overall CC errors (in terms of mean error or spread, or both) across both low levels of CC (<40%) and low height (<3m), and dense CC (>70%) across all height (<3m to >5m) in comparison to the X-band (highest variability and mean CC error) and C-band. Thus under sparse and low vegetation and bush encroaching conditions, it is the L-band which yields the lower levels of CC error and not the shorter wavelengths (X-band or C-band) as we may have expected. Also, the inclusion of the shorter wavelength datasets (X-band and C-band) with the L-band dataset led to minor improvements in the overall variability and mean of CC error across most sparse vegetation structural conditions (except regarding vegetation conditions with CC <40% and height >5m which is inconclusive) and across tall dense vegetation conditions (CC >70% and height >5m). Most significant improvement of the addition of the high frequency data occurred for the sparse and tallest trees (>3m) conditions.

5. Discussion

This study investigated the accuracy of modelling and mapping above ground biomass (AGB), woody canopy cover (CC) and total canopy volume (TCV) in heterogeneous South African savannahs using multi-frequency SAR datasets (X-band, C-band and L-band including their combinations). Various studies have implemented L-band SAR data for tree structural assessment in a savannah type environment (Carreiras et al., 2013, Mitchard et al., 2012) but the use of shorter wavelengths, such as C-band, have also been proven to perform relatively well (Mathieu et al., 2013). This study also served to compare the three SAR

frequency datasets (X-, C- and L-band) in the same study region of Mathieu et al. (2013) and is the first attempt in an African Savannah context. It was hypothesized that the shorter SAR wavelengths (e.g. X-band, C-band), since interacting with the finer woody plant elements (e.g. branchlets) would be useful for mapping the shrubby/thicket layer while the longer SAR wavelengths (e.g. L-band) would interact with larger vegetation elements such as major branches and trunks typical of forested areas (Vollrath, 2010; Mitchard et al, 2009). It was thus proposed that the combination of these different SAR frequencies would provide a better assessment of the savannah woody element than the individual SAR frequencies (Schmullius & Evans, 1997).

The modelling results indicated that it was the longer wavelength L-band dataset which contributed the most to the successful estimates of all three woody structural metrics. This finding agrees with other studies in the literature across a variety of ecosystem types such as coniferous forests (Dobson et al., 1992), boreal forest (Saatchi & Moghaddam, 2000) and temperate forests (Lucas et al., 2006). The results obtained for the L-band can be attributed to its ability to penetrate deeper into the canopy, allowing the signal to interact the most with the larger tree constituents such as the trunk and branches (Mitchard et al, 2009), and thus produces stronger correlations with the LiDAR metrics. Despite the status of leaf-off of most trees in winter, the shorter wavelengths (X- and C-band), 5.6cm for RADARSAT-2 and 3.1cm for TerraSAR-X, may have had a limited penetration of the canopy, and generally produced higher errors than the L-band for dense tree canopy (Figure 11ii). In the case of open woodlands ($CC < 40$, Figure 11i), results suggest that some penetration did occur through the larger gaps with some good performance of C- and X-band compared to L-band (see tree height > 3 m). However, C-band may have also been more sensitive to variability of surface roughness features (e.g. dense to sparse grass cover, fire scars etc.) which were too small to affect the coarser L-band (Wang et al., 2013, Bourgeau-Chavez et al., 2002, Menges et al., 2004). This interaction of the smaller wavelengths with these surface features may have introduced noise, which could have weakened correlations between the SAR signal and the LiDAR metrics.

The integration of the shorter wavelengths (e.g. X-band, C-band and X+C band), with L-band, yielded relatively small improvements in comparison to the L-band result alone (a reduction in SEP by ~3% and less for some metrics). The combination of all three frequencies yielded the highest overall accuracies for all metrics than each SAR frequency dataset alone. This result implies that the combination of short wavelength and long wavelength SAR datasets (X+C+L-band) does provide improved estimation in the modelling of the complete vegetation structure in terms of CC, TCV and AGB.

The CC error AOI maps (figures 9-10i-v) and the CC error box plots (figures 11i-ii) reaffirmed the modelling accuracy observations but provided greater insight into the specific SAR frequency contributions to the overall prediction error under a variety of woody structural conditions. The use of L-band alone and its integration with the shorter wavelengths reduced the overall CC overestimation error (mean error and variability) under sparse vegetation conditions while reducing overall CC underestimation under dense vegetated conditions, in comparison to the shorter wavelengths alone and their combinations. These observations thus go against the first part of the main hypothesis made in this study which hypothesised the importance of shorter wavelengths for interaction with the finer woody structural elements and shrubby vegetation cohorts as L-band appears to be more effective in this regard. The incorporation of the shorter wavelengths with the L-band improved the overall CC error budget by reducing the overall mean error and the overall variability of the error under most vegetation structural conditions. This result can be supported by the fact that the L-band was expected to penetrate deeper and interact with more of the lower levels of vegetation structure than the X- and C-band but the shorter wavelengths may have provided minor assistance to the L-band by interacting with certain canopy elements which the L-band may have missed due to wavelength size (Rosenqvist et al., 2003). Further investigation will need to be done to ascertain the exact cause of these results but these results, however, advocate the suitability of the L-band for analysing dense forested environments and thus confirms the second part of the main hypothesis which stated that the L-band SAR signal interacts with the major tree structural components (e.g. trunk and main branches typical of forested areas) (Lucas et al., 2006, Carreiras et al., 2013, Mitchard et al., 2012).

Among the three structural metrics, TCV was consistently modelled with higher accuracies, amongst all seven SAR scenarios (table 2). This result concurs with that of Mathieu et al. (2013). TCV is a metric which indicates the volume of vegetation present within the vertical structure and its higher modelled accuracies could be attributed to the leaf-off conditions typical of the dry winter season which allowed for greater wave penetration into the canopy for all wavelengths, even the shorter wavelengths. CC and AGB metrics yielded similar R^2 values with a higher SEP values observed for AGB which may be due to the associated error propagated through the allometric equation and the LiDAR model (results of figure 4). Since SAR is a system which utilises penetrating radio waves, the SAR signals will be expected to be more related to 3D structural metrics such as TCV and AGB rather than to the 2D CC metric (which achieved marginally poorer modelled results). This is due to the fact that CC is a metric for which the 2D horizontal coverage fluctuates seasonally depending on the phenological state of the vegetation, at least in comparison to TCV and AGB, which relies on the 3D nature of the woody structure which includes height and is thus more consistent across seasons (in the absence of disturbance).

The multi-frequency (X+C+L-band) model maps created for AGB (figure 7i), TCV (figure 7ii) and CC (figure 7iii) illustrate patterns and distributions resulting from influence of numerous biotic (mega-herbivore herbivory and anthropogenic pressures such as fuelwood extraction and cattle ranching) and abiotic factors (fire regimes, geology and topographic features) relevant to the study area. In order to discuss the common patterns in CC, TCV and AGB in these maps, it will be collectively referred to as “woody vegetation”. Dense woody vegetation patterns are observed in the protected forested woodlands (Bushbuckridge Nature Reserve) and in the exotic pine plantations within the vicinity of A. Generally, the riparian zones of major rivers and tributaries (e.g. B, the Sabie River catchment) have high values of CC, TCV and AGB compared to lower levels on the hill crests. In contrast to the vegetation occurring on granitic soils, the intrusions of the Timbavati gabbro geology group (figure 7 C) have very low woody CC, TCV and AGB. These geological substrates naturally support more open landscapes than the more densely vegetated granite soils. Rangeland areas in and within the vicinity of informal settlements, such as Justicea (F), also showed

lower levels of CC, TCV and AGB which could be linked to the heavy reliance of the local populace on fuelwood collection for energy requirements (Shackleton et al., 1994, Wessels et al., 2011, 2013). The key area of interest E (Athole area which consisted of historical rotational grazing camps which are currently inactive – Barend Erasmus, personal communication, 27/02/2013) possesses a sharp fence line contrast in tree structure between the dense woody vegetation evident in the northern extents of Athole (i.e. north of fence) and the sparse woody vegetation in Sabi Sands Private Game Reserve (i.e. south of fence). The extended absence of grazing and browsing pressures in the old pasture and paddock enclosures in the northern reaches of the Athole fence line boundary (figure 7 E) caused dense woody vegetation which contrasted sharply with the sparser woody vegetation in the more open and highly accessed areas south of the fence boundary. Additionally, the dense woody vegetation associated with the *Acacia welwitschii* thicket which dominates the ecca shales geological group of Southern Kruger National Park (outside map extents) was clearly visible at D (Mathieu et al., 2013). In conclusion, the accuracy and credibility of these maps and their trends have been supported by the various observations made during field visits and by the authors' general knowledge of the study area and the general range of these tree structural metric values agreed with the ranges reported in other related studies conducted in this savannah region (Mathieu et al., 2013 and Colgan et al., 2012).

6. Concluding Remarks

After reviewing all the modelling and error assessment results, it can be concluded the L-band SAR frequency was more effective in the modelling of the CC, TCV and AGB metrics in Southern African savannahs than the shorter wavelengths (X- and C-band) both as individual and combined (X+C-band) datasets. Although the integration of all three frequencies (X+C+L-band) yielded the best overall results for all three metrics, the improvements were noticeable but marginal in comparison to the L-band alone. The results do not warrant the acquisition of all three SAR frequency datasets for tree structure monitoring. Further the addition of the shortest wavelengths did not assist in the overall reduction of prediction

error specifically of the shrubby layer as hypothesized. With the imminent launch of the ALOS PALSAR-2 L-band sensor, the use of such L-band based models will be critical for future accurate tree structure modelling and monitoring at the regional and provincial scale. The modelling results obtained from the C-band SAR frequency alone, however, does yield promising results which would make the implementation of similar models to the free data obtained from the recently launched Sentinel-1 C-band sensor (launched in April 2014) viable when L-band datasets are not available. Sentinel-1 data are as far as we know the only upcoming operational, free and open access SAR dataset available in the near future, especially in Southern Africa. Building up of seasonal / annual time series may also improve on the performance of single date C-band imagery. The inclusion of seasonal optical datasets (e.g. reflectance bands, vegetation indices and textures derived from LandsAT platforms), which can provide more woody structural information, may also augment the modelling results.

As a way forward beyond this study, in order to reduce the error experienced in the AGB results (at field collection, LiDAR and SAR levels), new and more robust savannah tree allometric equations, with a greater range of representative tree stem and height sizes, will need to be produced but such efforts will require extensive ground level harvesting campaigns. Due to the success of this study, particularly the positive results using L-band SAR data, future work will seek to up-scale these results to greater regional and provincial areas using more extensive LiDAR calibration and validation datasets.

7. Acknowledgements

The authors would like to acknowledge the Council for Scientific and Industrial Research, the Department of Science and Technology, South Africa (grant agreement DST/CON 0119/2010, Earth Observation Application Development in Support of SAEOS) and the European Union's Seventh Framework Programme (FP7/2007-2013, grant agreement no. 282621, AGRICAB) for funding this study. The X-band StripMap TerraSAR-X scenes were acquired under a proposal submitted to the TerraSAR-X Science Service of the German

Aerospace Center (DLR). The C-band Quad-Pol RADARSAT-2 scenes were provided by MacDonald Dettwiler and Associates Ltd. – Geospatial Services Inc. (MDA GSI), the Canadian Space Agency (CSA), and the Natural Resources Canada’s Centre for Remote Sensing (CCRS) through the Science and Operational Applications Research (SOAR) programme. The L-band ALOS PALSAR FBD scenes were acquired under a K&C Phase 3 agreement with the Japanese Aerospace Exploration Agency (JAXA). The Carnegie Airborne Observatory is supported by the Avatar Alliance Foundation, John D. and Catherine T. MacArthur Foundation, Gordon and Betty Moore Foundation, W.M. Keck Foundation, the Margaret A. Cargill Foundation, Mary Anne Nyburg Baker and G. Leonard Baker Jr., and William R. Hearst III. The application of the CAO data in South Africa is made possible by the Andrew Mellon Foundation, Grantham Foundation for the Protection of the Environment, and the endowment of the Carnegie Institution for Science. The LiDAR data was processed by T. Kennedy-Bowdoin, D. Knapp, J. Jacobson and R. Emerson at the Carnegie Institution for Science. The authors would also like to acknowledge SANParks (Dr Izak Smit), Sabi Sands Game Reserve (Michael Grover), WITS Rural facility (Rhian Twine and Simon Khosa), SAEON (Patrick Ndlovu and Mightyman Mashele), CSIR EO colleagues and Bushbuckridge local authorities and personnel for arranging land access, field work expertise and providing logistical support. Personal thanks also go to Mr Mikhail Urbazaev for providing support in GAMMA scripting and processing of the SAR imagery.

8. References

- Anguita, D., A. Ghio, N. Greco, L. Oneto, S. Ridella, 2010. Model Selection for Support Vector Machines: Advantages and Disadvantages of the Machine Learning Theory. *International Joint Conference on Neural Networks*, IEEE, Barcelona, Spain, 18-23 July, 1-8
- Archibald, S., R.J. Scholes, 2007. Leaf green-up in a semi-arid African savannah – Separating tree and grass responses to environmental cues. *Journal of Vegetation Science*, 18, 583-594

- Asner, G.P., D.E. Knapp, T. Kennedy-Bowdoin, M.O. Jones, R.E. Martin, J. Boardman & C.B. Field, 2007. Carnegie Airborne Observatory: in-flight fusion of hyperspectral imaging and waveform LiDAR for 3D studies of ecosystems. *Journal of Applied Remote Sensing*, 1, 1-27
- Asner, G.P., S.R. Levick, T. Kennedy-Bowdoin, D.E. Knapp, R. Emerson, J. Jacobson et al., 2009. Large-scale impacts of herbivores on the structural diversity of African savannas. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 4947-4952
- Asner, G.P., D.E. Knapp, J. Boardman, R.O. Green, T. Kennedy-Bowdoin, M. Eastwood, R.E. Martin, C. Anderson, C.B. Field, 2012. Carnegie Airborne Observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sensing of Environment*, 124, 454-465
- Asner, G.P., J. Mascaro, C. Anderson, D.E. Knapp, R.E. Martin, T. Kennedy-Bowdoin, M. van Breugel, S. Davies, J.S. Hall, H.C. Muller-Landau, C. Potvin, W. Sousa, J. Wright, E. Bermingham, 2013. High-fidelity national carbon mapping for resource management and REDD+. *Carbon Balance and Management*, 8 (7), 1-14
- Boggs, G.S., 2010. Assessment of SPOT 5 and QuickBird remotely sensed imagery for mapping tree cover in savannas. *International Journal of Applied Earth Observation and Geoinformation*, 12 (4), 217-224
- Bombelli, A., V. Avitabile, H. Balzter, L.B. Marchesini, 2009. T12 Assessment of the status of the development of the standards for the terrestrial essential climate variables – Biomass. *GTOS*, 67 (10), 1-30
- Bourgeau-Chavez, L.L., E.S. Kasischke, S. Brunzell, J.P. Mudd, 2002. Mapping fire scars in global boreal forests using imaging radar data. *International Journal of Remote Sensing*, 33 (20), 4211-4234

- Bradbury, R.B., R.A. Hill, D.C. Mason, S.A. Hinsley, J.D. Wilson, H. Balzter et al., 2005. Modelling relationships between birds and vegetation structure using airborne LiDAR data: A review with case studies from agricultural and woodland environments. *Ibis*, 147, 443-452
- Breiman, L., 2003. Manual on setting up, using and understanding Random Forests v4.0, http://oz.berkeley.edu/users/breiman/Using_random_forests_v4.0.pdf (accessed 08.02.11)
- Carreira, J.M.B., J.B. Melo, M.J. Vasconcelos, 2013. Estimating the above-ground biomass in Miombo savannah woodlands (Mozambique, East Africa) using L-band synthetic aperture radar data. *Remote Sensing Open Access*, 5, 1524-1548
- Castillo-Santiago, M.A., M. Ricker, B.H.J. de Jong, 2010. Estimation of tropical forest structure from SPOT-5 satellite images. *International Journal of Remote Sensing*, 31 (10), 2767-2782
- Cho, M.A., R. Mathieu, G.P. Asner, L. Naidoo, J. van Aardt, A. Ramoelo 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
- Colgen, M.S., G.P. Asner, S.R. Levick, R.E. Martin, O.A. Chadwick, 2012. Topo-edaphic controls over woody plant biomass in South African savannas. *Biogeosciences*, 9, 1809-1821
- Colgen, M.S., G.P. Asner, T. Swemmer, 2013. Harvesting tree biomass at the stand-level to assess the accuracy of field and airborne biomass estimation in savannas. *Ecological Applications*, 23 (5), 1170-1184
- Collins, J.N., L.B. Hutley, R.J. Williams, G. Boggs, D. Bell, R. Bartolo, 2009. Estimating landscape-scale vegetation carbon stocks using airborne multi-frequency polarimetric synthetic aperture radar (SAR) in the savannahs of North Australia. *International Journal of Remote Sensing*, 30 (5), 1141-1159

- Corbera, E., H. Schroeder, 2011. Governing and implementing REDD+. *Environmental Science & Policy*, 14 (2), 89-99
- Dobson, M.C., F.T. Ulaby, T. Le Toan, A. Beaudoin, E.S. Kasischke, N. Christensen, 1992. Dependence of radar backscatter on coniferous forest biomass. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 412-415
- Falkowski, P., R.J. Scholes, E. Boyle, J. Canadell, D.Canfield et al., 2000. The global carbon cycle: a test of our knowledge of Earth as a system. *Science*, 290, 291-296
- Fiala, A.C.S., S.L. Garman, A. Gray, 2006. Comparison of five canopy-cover estimation techniques in the western Oregon Cascades. *Forest Ecology Management*, 232, 188-197
- Ismail, R., O. Mutanga, L. Kumar, 2010. Modelling the potential distribution of pine forests susceptible to *Sirex Noctilo* infestations in Mpumalanga, South Africa. *Transactions in GIS*, 14 (5), 709-726
- Jennings, S.B., N.D. Brown, D. Sheil, 1999. Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. *Forestry*, 72, 59-73
- Johansen, K., S. Phinn, 2006. Mapping structural parameters and species composition of riparian vegetation using IKONOS and Landsat ETM+ data in Australian tropical savannahs. *Photogrammetric Engineering & Remote Sensing*, 72 (1), 71-80
- Jung, K., S. Kaiser, S. Boehm, J. Nieschulze, E.K.V. Kalko, 2012. Moving in three dimensions: effects of structural complexity on occurrence and activity of insectivorous bats in managed forest stands. *Journal of Applied Ecology*, 49, 523-531
- Kanowski, P.J., C.L. McDermott, B.W. Cashore, 2011. Implementing REDD+: lessons from analysis of forest governance. *Environmental Science & Policy*, 14 (2), 111-117

- Ko, D., N. Bristow, D. Greenwood, P. Weisberg, 2009. Canopy cover estimation in semiarid woodlands: comparison of field-based and remote sensing methods. *Forest Science*, 55 (2), 132-141
- Lee, J.S., 1980. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2 (2), 165-168
- Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation, *International Journal of Remote Sensing*, 27 (7), 1297-1328
- Lucas, R.M., N. Cronin, A. Lee, M. Moghaddam, C. Witte, P. Tickle, 2006. Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, Queensland, Australia. *Remote Sensing of Environment*, 100, 407-425
- Mathieu, R., L. Naidoo, M.A. Cho, B. Leblon, R. Main, K. Wessels, G.P. Asner, J. Buckley, J. Van Aardt, B.F.N. Erasmus, I.P.J. Smit, 2013. Toward structural assessment of semi-arid African savannahs and woodlands: the potential of multitemporal polarimetric RADARSAT-2 fine beam images. *Remote Sensing of Environment*, 138, 215-231
- Matsika, R., B.F.N. Erasmus, W.C. Twine, 2012. A tale of two villages: assessing the dynamics of fuelwood supply in communal landscapes in South Africa. *Environmental Conservation*, 40 (1), 71-83
- Menges, C.H., R.E. Bartolo, D. Bell, G.J.E. Hill, 2004. The effect of savannah fires on SAR backscatter in northern Australia, *International Journal of Remote Sensing*, 25 (22), 4857-4871
- Mills, A.J., K.H. Rogers, M. Stalmans, E.T.F. Witkowski, 2006. A framework for exploring the determinants of savanna and grassland distribution. *Bioscience*, 56 (7), 579-589
- Mitchard, E.T.A., S.S. Saatchi, I.H. Woodhouse, G. Nangendo, N.S. Ribeiro, M. Williams, C.M. Ryan, S.L. Lewis, T.R. Feldpausch, P. Meir, 2009. Using satellite radar backscatter to

predict above-ground woody biomass: A consistent relationship across four different African landscapes, *Geophysical Research Letters*, 36, 1-6

- Mitchard, E.T.A., S.S. Saatchi, S.L. Lewis, T.R. Feldpausch, I.H. Woodhouse, B. Sonke, C. Rowland, P. Meir, 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
- Mitchard, E.T.A, S.S. Saatchi, L.J.T. White, K.A. Abernethy, K.J. Jeffery, S.L. Lewis, M. Collins, M.A. Lefsky, M.E. Leal, I.H. Woodhouse, P. Meir, 2012. Mapping tropical forest biomass with radar and spaceborne LiDAR in Lope National Park, Gabon: overcoming problems of high biomass and persistent cloud. *Biogeosciences*, 9, 179-191
- Mougin, E., C. Proisy, G. Marty, F. Fromard, H. Puig, J.L. Betoulle, J.P. Rudant, 1999. Multifrequency and multipolarization radar backscattering from mangrove forests. *IEEE Transactions on Geoscience and Remote Sensing*, 37 (1), 94-102
- Mucina, L., M.C. Rutherford (Eds.), 2006. The vegetation of South Africa, Lesotho and Swaziland. South African National Biodiversity Institute, Pretoria
- Mueller, J., J. Stadler, R. Brandl, 2010. Composition versus physiognomy of vegetation as predictors of bird assemblages: The role of lidar. *Remote Sensing of Environment*, 114, 490-495
- Muss, J.D., D.J. Mladenoff, P.A. Townsend, 2011. A pseudo-waveform technique to assess forest structure using discrete lidar data. *Remote Sensing of Environment*, 115, 824-835
- Naidoo, L., R. Mathieu, R. Main, W. Kleynhans, K. Wessels, G.P. Asner, B. Leblon, 2014. The assessment of data mining algorithms for modelling savannah woody cover using multi-frequency (X-, C- and L-band) synthetic aperture radar (SAR) datasets. *IEEE*

International Geoscience and Remote Sensing Symposium, Quebec, Canada, 13-18 July, 2014, 1-4

- Nichol, J.E., L.R. Sarker, 2011. Improved biomass estimation using the texture parameters of two high-resolution optical sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 49 (3), 930-948
- Prasad, A.M., L.R. Iverson, A., Liaw, 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, 9 (2), 181-199
- Rosenqvist, A., A. Milner, R. Lucas, M. Imhoff, C. Dobson, 2003. A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science & Policy*, 6, 441-455
- Ryan, C.M., T. Hill, E. Woollen, C. Ghee, E. Mitchard, G. Cassells, J. Grace, I.H. Woodhouse, M. Williams, 2011. Quantifying small-scale deforestation and forest degradation in African woodlands using radar imagery. *Global Change Biology*, 18 (1), 243-257
- Saatchi, S.S., M. Moghaddam, 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
- Santos, J.R., M.S.P. Lacruz, L.S. Araujo, M. Keil, 2002. Savanna and tropical rainforest biomass estimation and spatialization using JERS-1 data. *International Journal of Remote Sensing*, 23, 1217-1229
- Schullius, C.C., D.L. Evans, 1997. Review article Synthetic aperture radar (SAR) frequency and polarization requirements for applications in ecology, geology, hydrology, and oceanography: A tabular status quo after SIR-C/X-SAR, *International Journal of Remote Sensing*, 18 (13), 2713-2722

- Shackleton, C.M., N.J. Griffin, D.I. Banks, J.M. Mavrandonis, S.E. Shackleton, 1994. Community structure and species composition along a disturbance gradient in a communally managed South African savannah. *Vegetatio*, 115, 157-167
- Silva, J.F., A. Zambrano, M.R. Farinas, 2001. Increase in the woody component of seasonal savannas under different fire regimes in Calabozo, Venezuela. *Journal of Biogeography*, 28, 977-983
- Stumpf, K.A., 1993. The estimation of forest vegetation cover descriptions using a vertical densitometer. [Internet site]. Available at: http://www.grsgis.com/publications/saf_93.html. [cited 30 June 2014]
- Sun, G., K. Jon Ranson, Z. Guo, Z. Zhang, P. Montesano, D. Kimes, 2011. Forest biomass mapping from LiDAR and Radar synergies. *Remote Sensing of Environment*, 115, 2906-2916
- Tsui, O.W., N.C. Coops, M.A. Wulder, P.L. Marshall, A. McCardle, 2012. Using multi-frequency radar and discrete-return LiDAR measurements to estimate above-ground biomass and biomass components in a coastal temperate forest. *ISPRS Journal of Photogrammetry and Remote Sensing*, 69, 121-133
- Tu, V.J., 1996. Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes. *Journal of Clinical Epidemiology*, 49 (11), 1225-1231
- Urbazaev, M., C. Thiel, C. Schmullius, R. Mathieu, L. Naidoo, S.R. Levick, I.P.J. Smit, G.P. Asner, 2013. Mapping of fractional woody cover using ALOS PALSAR L-band backscatter in southern African savannas. Master thesis (MSc), Friedrich-Schiller University Jena, Institute of Geography, Germany
- Viergever, K.M., I.H. Woodhouse, N. Stuart, 2008. Monitoring the world's savanna biomass by earth observation. *Scottish Geographical Journal*, 124 (2-3), 218-225

- Vollrath, A., 2010. Analysis of woody cover estimations with regard to different sensor parameters using the SIR-C/X-SAR dataset of Kruger National Park, RSA. Master thesis (MSc), Friedrich-Schiller University Jena, Institute of Geography, Germany, 1- 117
- Wang, X., L. Ge, X. Li, 2013. Pasture monitoring using SAR with COSMO-SkyMed, ENVISAT ASAR, and ALOS PALSAR in Otway, Australia. *Remote Sensing Open Access*, 5, 3611-3636
- Weepener, H.L., H.M. van den Berg, M. Metz, H. Hamandawana, 2011. The development of a hydrological improved Digital Elevation Model and derived products for South Africa based on the SRTM DEM. Water Research Commission, Report no. K5/1908, Pretoria, 1- 52
- Wessels, K.J., R. Mathieu, B.F.N. Erasmus, G.P. Asner, I.P.J. Smit, J.A.N. van Aardt et al., 2011. Impact of communal land use and conservation on woody vegetation structure in the Lowveld savannas of South Africa. *Forest Ecology and Management*, 261, 19-29
- Wessels, K.J., M.S. Colgan, B.F.N. Erasmus, G.P. Asner, W.C. Twine, R. Mathieu, J.A.N. van Aardt, J.T. Fisher, I.P.J. Smit, 2013. Unsustainable fuelwood extraction from South African savannas. *Environmental Research Letters*, 8 (1), 1-10
- Wigley, B.J, W.J Bond, M.T Hoffman, 2009. Bush encroachment under three contrasting land-use practices in a mesic South African savanna. *African Journal of Ecology*, 47 (1), 62-70
- Zhao, Y. & Y. Zhang, 2008. Comparison of decision tree methods for finding active objects. *Advances in Space Research*, 41, 1955-1959

Appendix A

$$M = 0.109D^{(1.39+0.14\ln(D))} H^{0.73} p^{0.80}$$

Where M = biomass in kg/Ha, D = Diameter above breast height (DBH) in cm, H = height of tree in metres and p = mean wood specific gravity (fixed at a mean value of 0.9) which is unitless.

Appendix B

Total 25m by 25m AGB plot = $X + Y + (Z \cdot 6.25)$

Where X is the total AGB of stems ≥ 10 cm DBH, Y is total AGB of stems between 5 and 10cm DBH and Z is the total AGB of stems between 3 and 5cm DBH. The up-scaling factor of 6.25 was used as stems between 3 and 5cm were only sampled within the 10 by 10m (i.e. DBH zone 1) subplot and not sampled for the rest of the 25 by 25m grid (i.e. DBH zone 2). So 625m^2 (i.e. total area of the 25 by 25m sample plot) divided by 100m^2 (area of the 10 by 10m subplot) is 6.25. All remaining stems within the 25 by 25m sample plot, which subscribed to the remaining DBH conditions (i.e. ≥ 5 cm DBH), were measured and therefore did not require any up-scaling factors.

Figure1: The Southern Kruger National Park region and the spatial coverage of all implemented remote sensing datasets. The solid red line indicates the coverage of the 2009 RADARSAT-2 scenes while the solid gold line indicates the two scenes of the 2010 ALOS dual-pol PALSAR imagery. The dashed grey line indicates the five scenes of the 2012 TerraSAR-X StripMap imagery. The shaded black areas represent the coverage of the 2012 CAO LiDAR sensor tree cover product. The red squares indicate the 38 sample sites where field data collections took place.

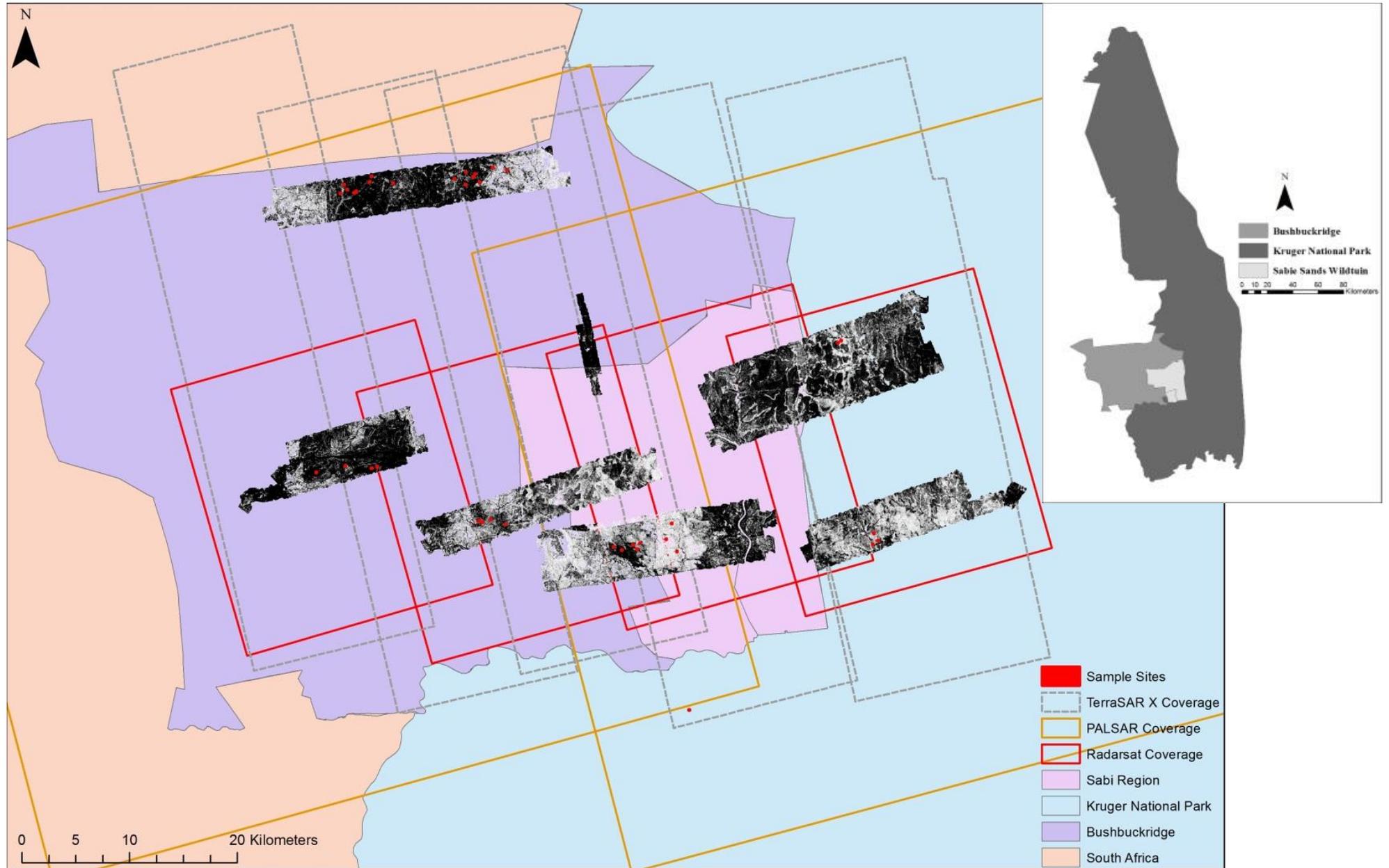
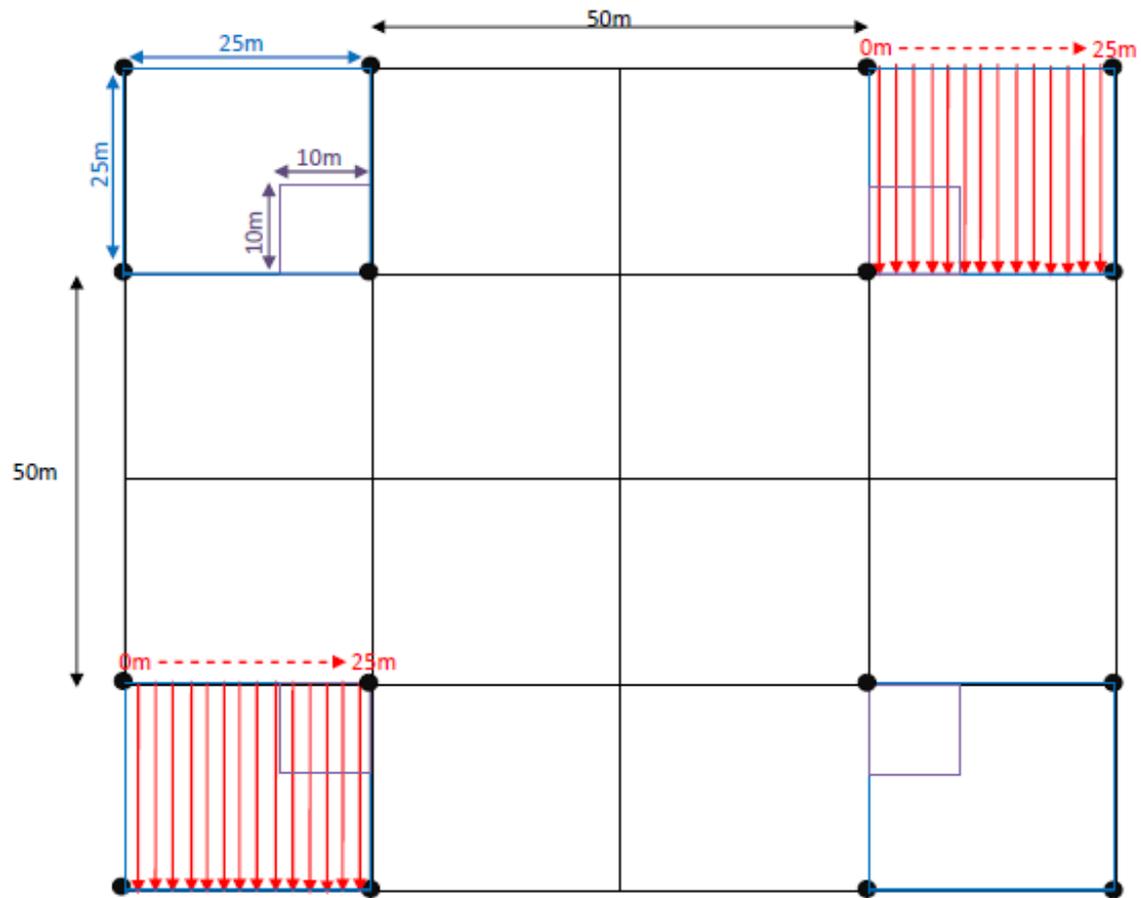


Figure2: Ground sampling design including ground tree biomass and tree cover collection protocols (50m spacing between sample plots coincide with the auto-correlation distance – refer to data integration section)



Legend and Sampling Protocols for Tree Biomass and Cover:

- Corner pole markers
- DBH Zone 1: 10X10m [Trees with DBH \geq 3cm measured]
- DBH Zone 2: 25X25m [Trees with DBH \geq 5cm measured]
- ▮ Line Transect and Vertical Densiometer methods (1m and/or 2m intervals)

Figure3: Validation results of field-measured CC (ground truth) versus LiDAR derived CC (above 0.5m height, N=37)

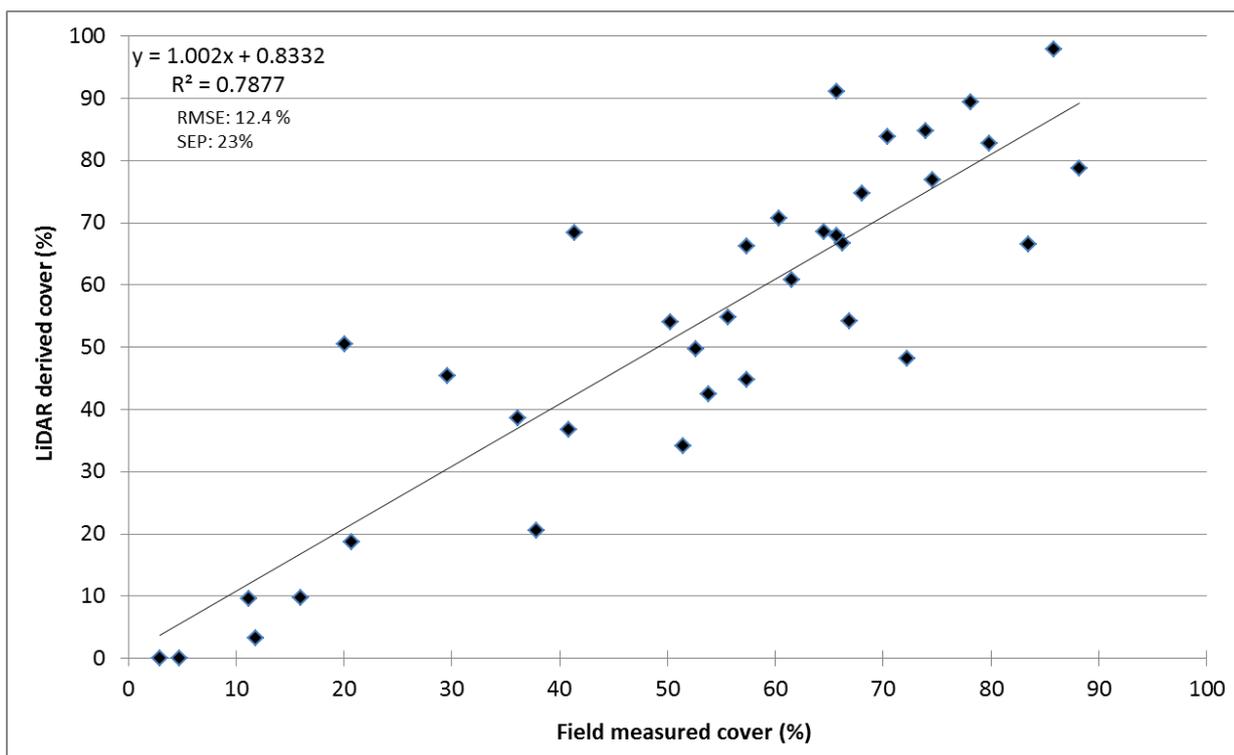


Figure 4: Validation results of field-measured AGB (ground truth) versus LiDAR derived AGB (above 0.5m height, N=53)

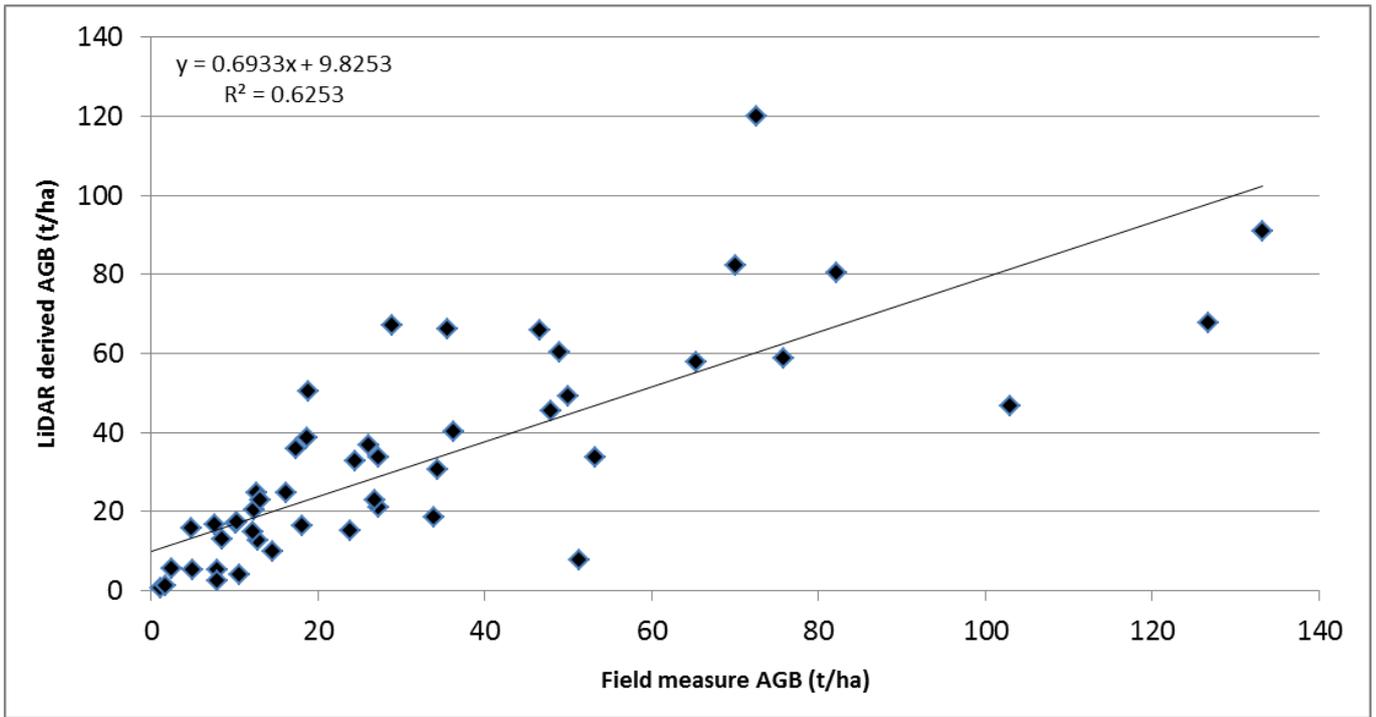


Figure 5: Methodology schema describing the data integration and modelling process

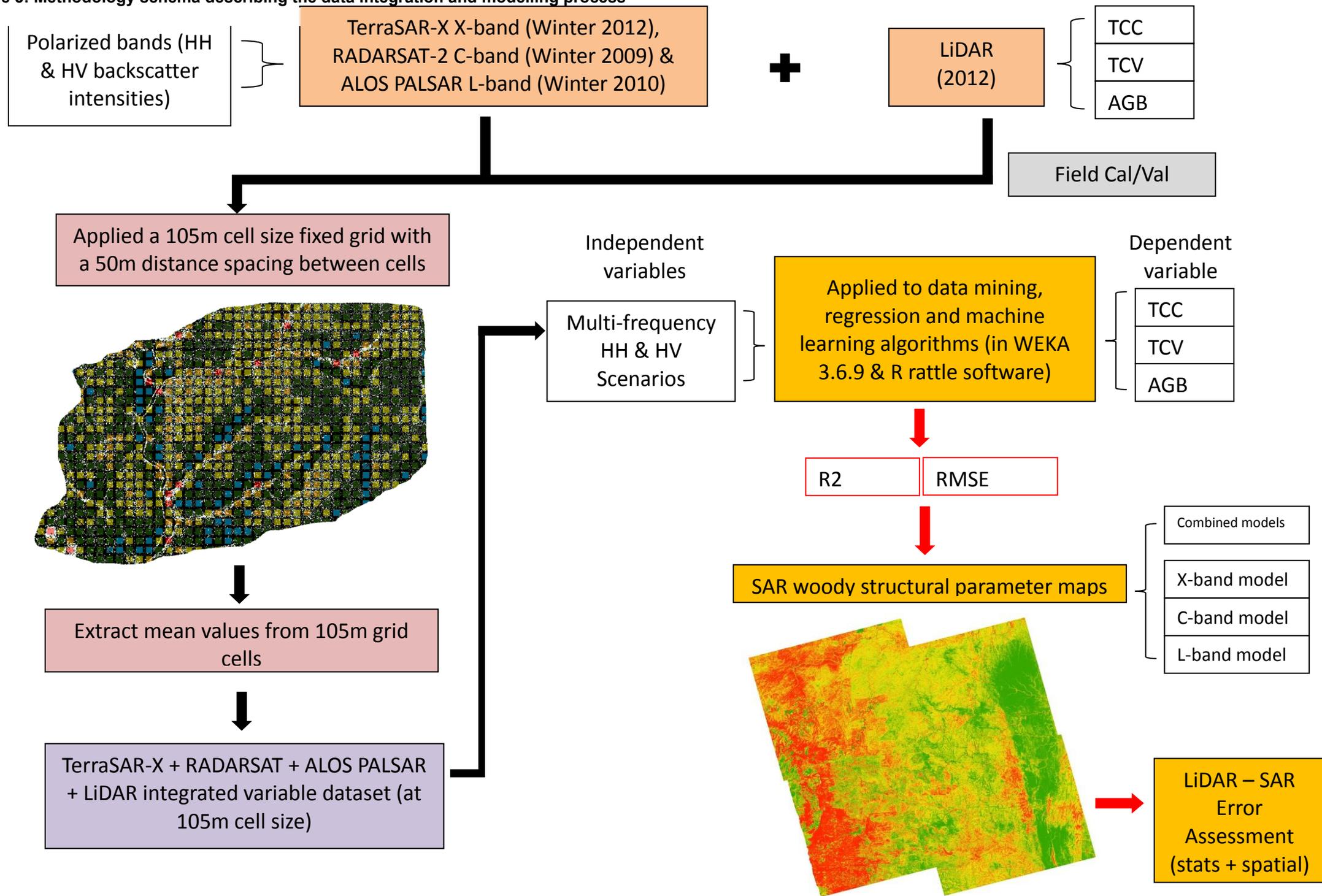


Figure 6A-G: Observed versus Predicted Total woody Canopy Volume (TCV) scatter density plots (dotted line is 1:1)

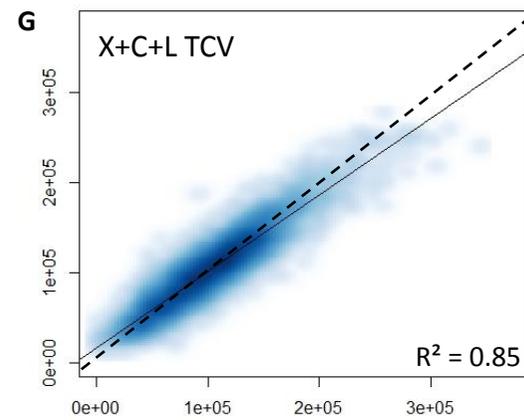
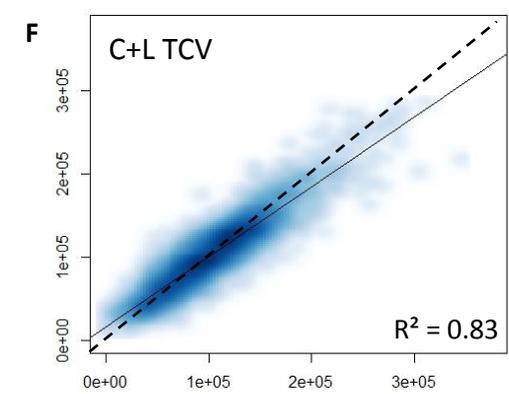
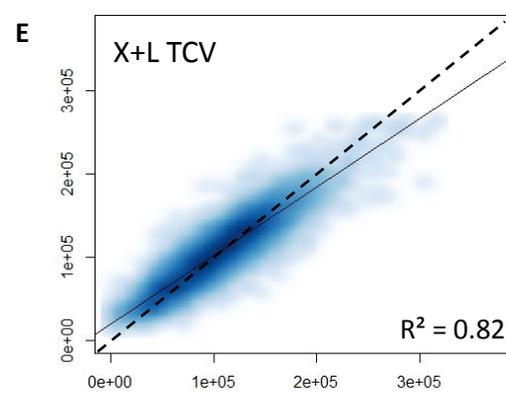
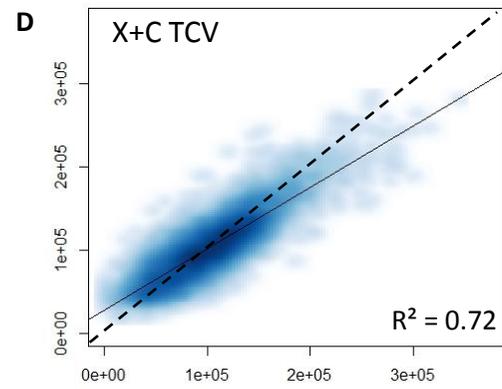
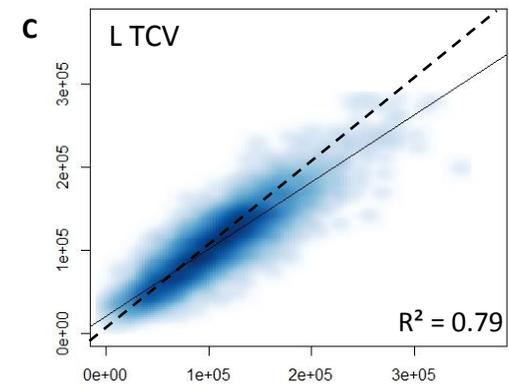
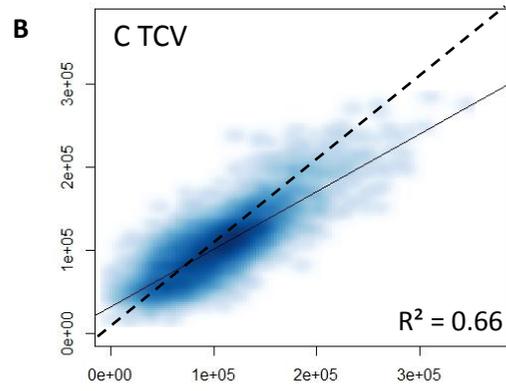
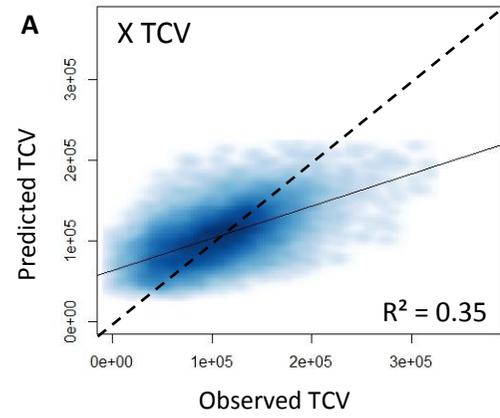


Figure 7i-iii: X+C+L SAR derived tree structural metric maps, for i) Above Ground Biomass, ii) Total woody Canopy Volume and iii) woody Canopy Cover, using random forest. Letters A-F represents key areas of interest for discussion (for all three metrics)

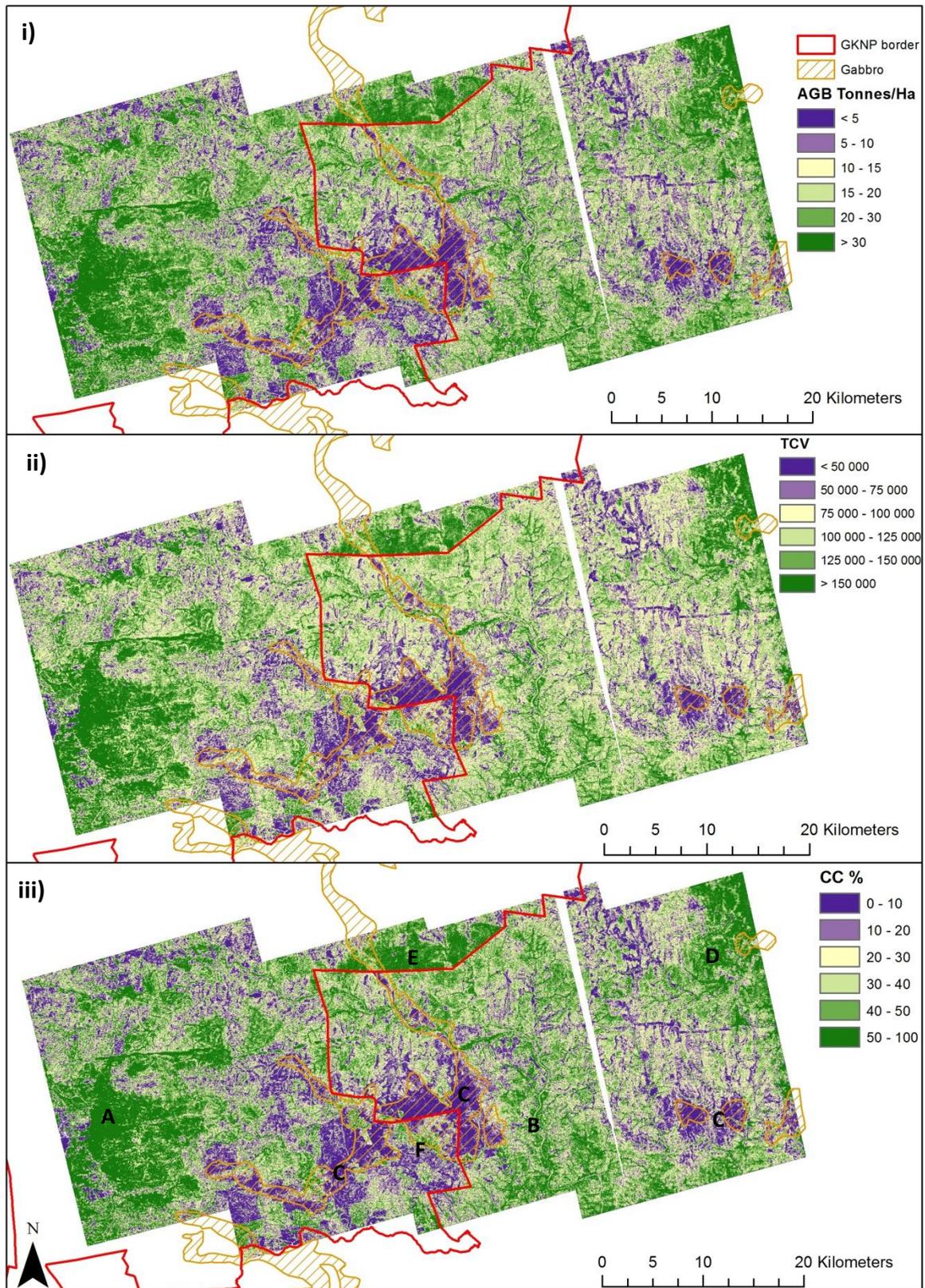


Figure 8: Scatterplot of AGB (y-axis) versus CC (x-axis) under dense cover conditions (plotted from pixels extracted from the region of AOI 'A')

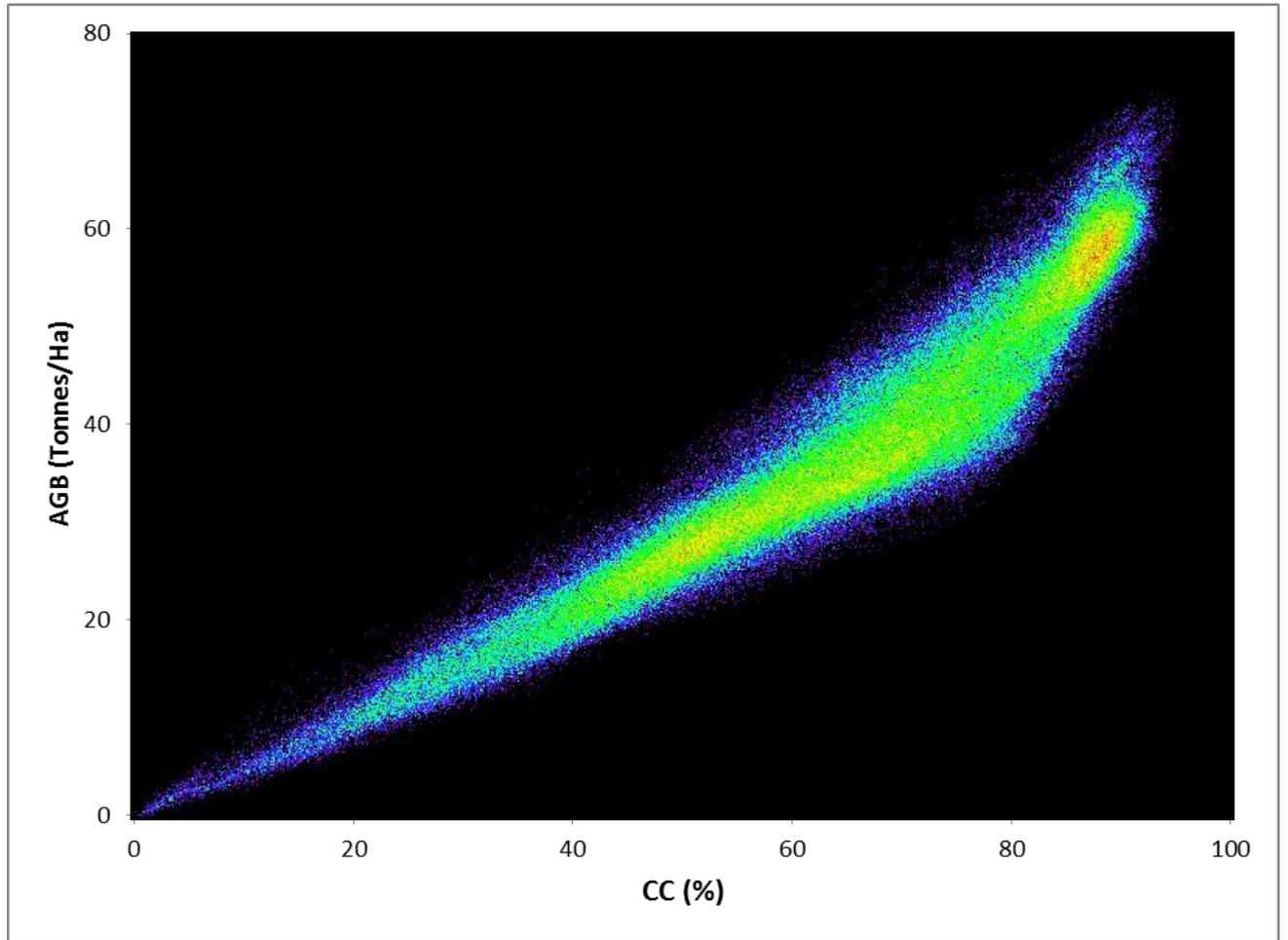


Figure 9i-v: LiDAR - SAR scenario difference (error) maps of CC for the Xanthia Forest Area of Interest (close to A); v) 1m LiDAR-derived CC map

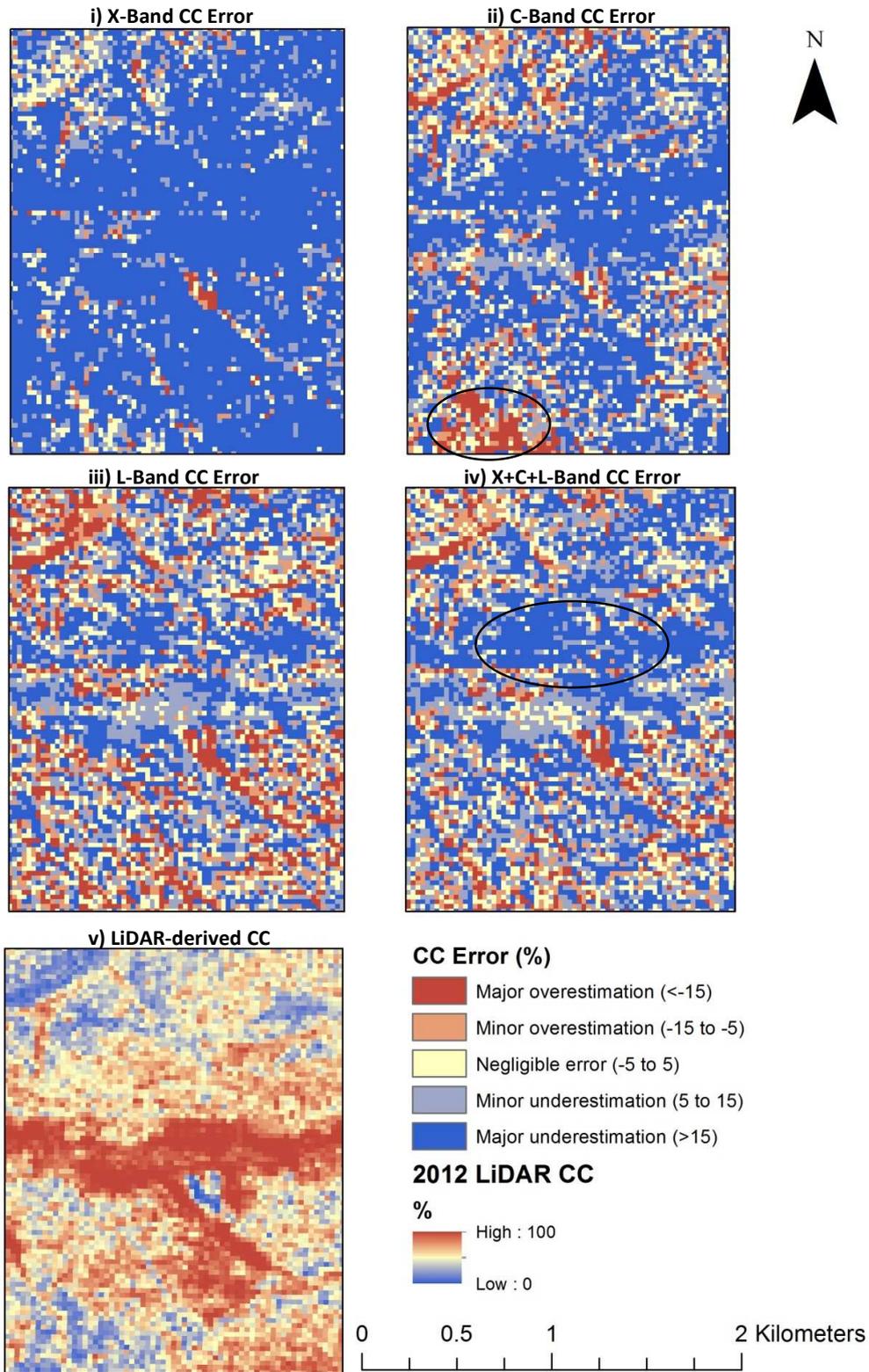
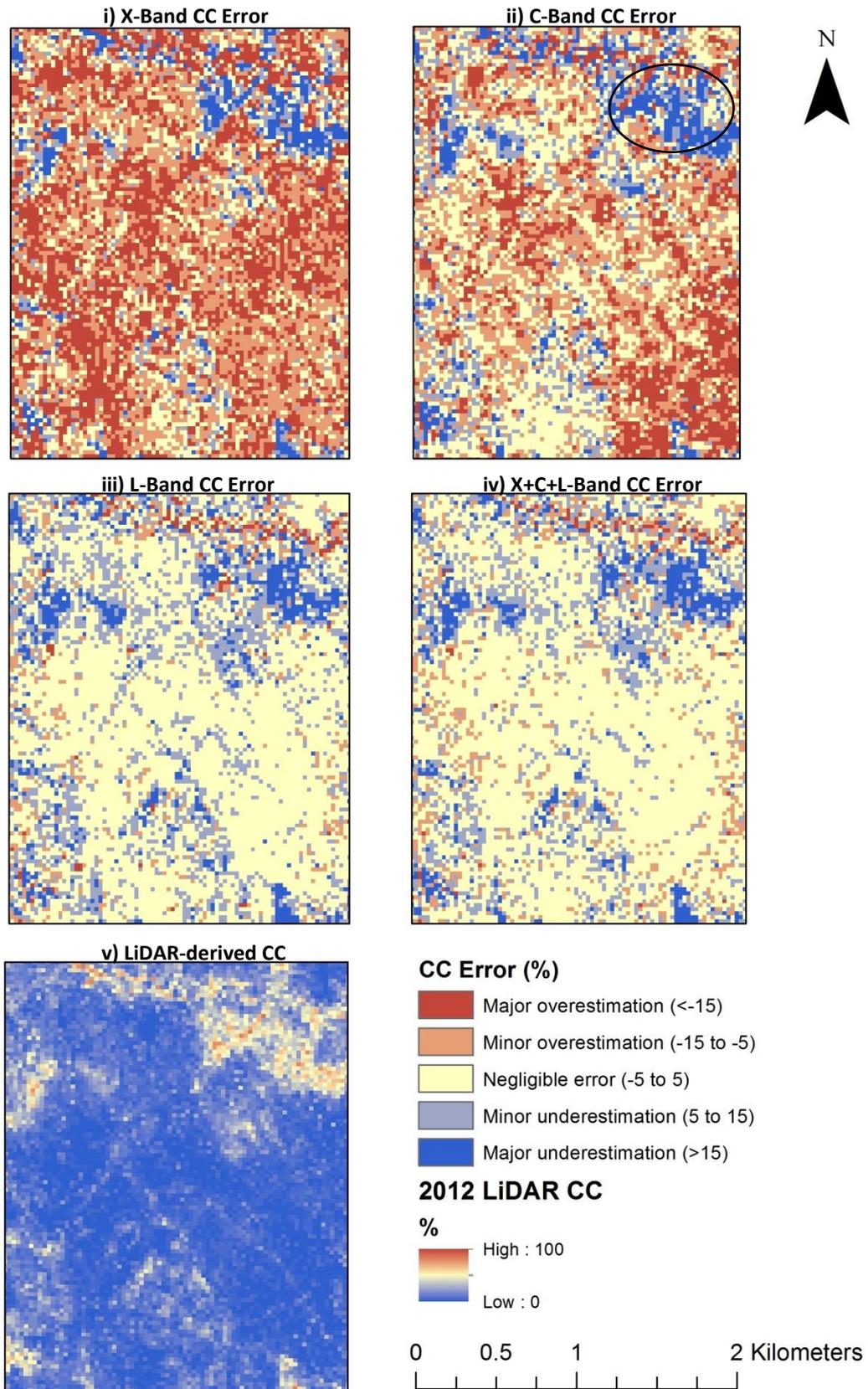


Figure 10i-v: LiDAR - SAR scenario difference (error) maps of CC for the Gabbro Intrusions Area of Interest (C); v) 1m LiDAR-derived CC map



Figures 11i-ii): CC Error box plots of: i) low LiDAR CC (<40%) and variable LiDAR vegetation height and ii) dense LiDAR CC (>70%) and variable LiDAR vegetation height (+'ve values = CC underestimation; -'ve values = CC overestimation; dashed line partitions the four main SAR scenarios across the x-axis classes, centre point = mean value, box = standard error and whiskers = standard deviation) (N= 17559)

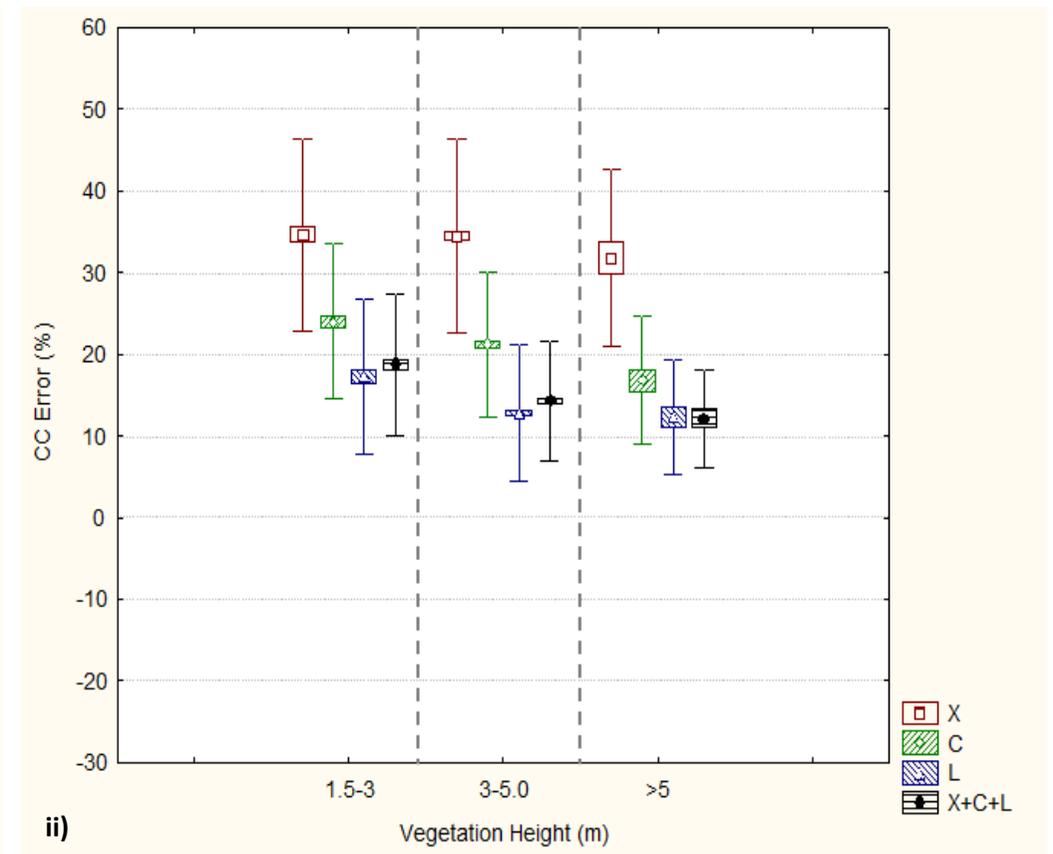
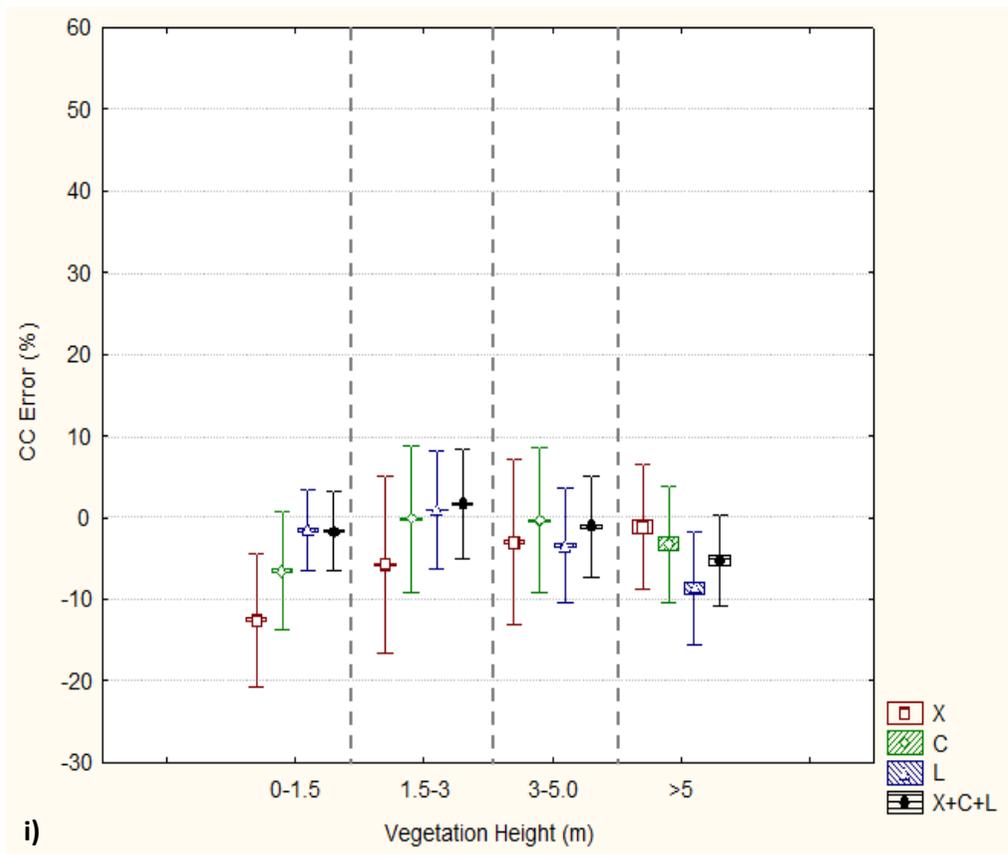


Table 1: SAR and LiDAR datasets acquired and utilised for the modelling of woody structural metrics

Imagery	Sensor	Mode	Original Spatial Resolution	Incidence angle	Acquisition time	Season
1 2 3 4 5	TerraSAR-X X-band	StripMap Dual Polarized (HH & HV)	3m	38.1-39.3° 21.3-22.8° 37.2-38.4° 36.2-37.4° 39.1-40.2°	08/09/2012 23/08/2012 28/08/2012 19/09/2012 30/09/2012	Late Winter 2012
1 2 3 4	RADARSAT-2 C-band	Quad Polarized (HH, HV, VH, VV) but only HH and HV used	5m	34.4 - 36.0° 39.3 - 40.1° 32.4 - 34.0° 37.4 - 38.9°	13/08/2009 06/08/2009 06/09/2009 30/08/2009	Winter 2009
1 2	ALOS PALSAR L-band	Dual Polarized (HH & HV)	12.5m	34.3°	14/08/2010 31/08/2010	Winter 2010
AGB (kg) Product CC (%) Product TCV Product	CAO LiDAR	Discrete Footprint	25m (1.1m original)	Nadir	1/04/2012- 24/05/2012	End summer 2012

Table 2: CC, TCV & AGB woody parameter modelling accuracy assessment (validation) results obtained from the Random Forest algorithm according to seven SAR frequency scenarios

CC Model Validation Results [split into 35% Training & 65% Validation]; Units = %					
X-band only		C-band only		L-band only	
R ²	RMSE (SEP)	R ²	RMSE (SEP)	R ²	RMSE (SEP)
0.34	18.12% (50.87%)	0.61	13.20% (38.50%)	0.77	10.59% (29.64%)
X+C band		X+L band		C+L band	
R ²	RMSE (SEP)	R ²	RMSE (SEP)	R ²	RMSE (SEP)
0.69	11.71% (33.94%)	0.80	9.90% (27.78%)	0.81	9.23% (26.94%)
		X+C+L band			
		R ²	RMSE (SEP)		
		0.83	8.76% (25.40%)		
TCV Model Validation Results [split into 35% Training & 65% Validation]; Units = unitless per hectare					
X-band only		C-band only		L-band only	
R ²	RMSE (SEP)	R ²	RMSE (SEP)	R ²	RMSE (SEP)
0.35	35534.50 (33.79%)	0.66	24731.06 (24.07%)	0.79	19902.79 (18.88%)
X+C band		X+L band		C+L band	
R ²	RMSE (SEP)	R ²	RMSE (SEP)	R ²	RMSE (SEP)
0.72	22243.64 (21.59%)	0.82	18609.04 (17.70%)	0.83	17236.50 (16.77%)
		X+C+L band			
		R ²	RMSE (SEP)		
		0.85	16443.57 (15.96%)		
AGB Model Validation Results [split into 35% Training & 65% Validation]; Units = tonnes/ha					
X-band only		C-band only		L-band only	
R ²	RMSE (SEP)	R ²	RMSE (SEP)	R ²	RMSE (SEP)
0.32	10.88t/ha (59.82%)	0.60	7.81t/ha (43.66%)	0.78	6.05t/ha (32.90%)
X+C band		X+L band		C+L band	
R ²	RMSE (SEP)	R ²	RMSE (SEP)	R ²	RMSE (SEP)
0.67	7.19t/ha (40.33%)	0.81	5.70t/ha (31.35%)	0.81	5.45t/ha (30.44%)
		X+C+L band			
		R ²	RMSE (SEP)		
		0.83	5.20t/ha (29.18%)		

Table 3: Total CC % error across the entire LiDAR-SAR coverage for the four main SAR frequency scenarios

Error Classes	X-band Error	C-band Error	L-band Error	X+C+L-band Error
Major overestimation (<-15)	21.02	13.87	12.78	9.43
Minor overestimation (-15 to -5)	17.30	16.38	16.74	16.85
Negligible error (-5 to 5)	19.52	24.58	31.34	31.84
Minor underestimation (5 to 15)	13.87	16.95	19.27	20.08
Major underestimation (>15)	28.29	28.21	19.87	21.80