Elsevier Editorial System(tm) for ISPRS Journal of Photogrammetry and Remote Sensing

Manuscript Draft

Manuscript Number: PHOTO-D-14-00401R1

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

Article Type: Original Research Paper

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

Corresponding Author: Mr. Laven Naidoo, MSc

Corresponding Author's Institution: Council for Scientific and Industrial Research (CSIR)

First Author: Laven Naidoo, MSc

Order of Authors: Laven Naidoo, MSc; Renaud Mathieu, PhD; Russell Main, MSc; Waldo Kleynhans, PhD; Konrad Wessels, PhD; Gregory Asner, PhD; Brigitte Leblon, PhD

Manuscript ID: PHOTO-D-14-00401

ISPRS Journal of Photogrammetry and Remote Sensing

Type of manuscript: Journal Article

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

Authors: Laven Naidoo, Renaud Mathieu, Russell Main, Waldo Kleynhans, Konrad Wessels, Gregory Asner, Brigitte Leblon

Date: 4th March 2015

Dear Dr. Onisimo Mutanga, the Associate Editor, and Dr. Derek Lichti, the Editor-in-Chief

Thank you and the reviewers for the first round of comments and suggestions concerning our submitted manuscript. We have revised the manuscript accordingly by carefully taking into account the feedback put forward by the reviewers. Almost all of the suggestions were accepted for revision. The comments and the associated responses (in red) are mentioned below and the corrections included in the revised manuscript:

Reviewer #1:

- We would like to thank the reviewer for the positive comments pertaining to our manuscript
- As requested by the reviewer, the text in the manuscript have been aligned to both margins ("justified")

Main Points/concerns:

- **Contributions seem a little weak with regard to author's previous publications and other works (e.g. Urbazaev et al….). We would like to thank the reviewer for this comment. We believe the work has value and contributes to the scientific community as multi-frequency SAR and the usefulness of each of the frequencies (and their combinations) have not been investigated in this detail in the academic literature within the context of modelling savannah tree structural attributes (e.g. canopy cover and biomass etc.). The contributions (error distributed across structural attribute grouping and spatial patterning) of each frequency and their combinations could be seen as the major contribution of this manuscript rather than the actual modelling side which may resemble some of the previous publications work. A statement was added to the end of page 6 of the manuscript, accompanying research question #3, to further support the purpose of the study ["More specifically, the investigation of the interactions of the different SAR frequencies, and their possible combinations, across the different**

vegetation patterning and structural classes, such as grasslands, thickets and forests, will pin-point the effective application of the different SAR frequencies and their possible combinations in Southern African savannah landscapes."]

- **Presentation of Figure 8 and 9 are not easily understandable…from my point of view they add a little more confusion than clarification or support. Figure 9 and 10 may confuse the reader when it comes to the main error trends observed in the boxplot figures (Figure 8) but they both represent, visually, realistic examples of what is seen in the different landscape types associated with South African savannahs. The authors believe these figures have merit and should be included in the manuscript.**
- **Language issues related to the proper introduction of abbreviations, consistence when presenting certain concepts, long sentences and the overuse of parentheses. The authors would like to thank the reviewer for pointing out these language issues. The appropriate corrections were made throughout the manuscript to allow for improved sentence flow and structuring.**

Comments/Suggestions page by page:

- **P1L2: abbreviation of SAR might be redundant. Agreed, the SAR abbreviation was removed from the manuscript title**
- **P1L54: The paper is mainly written using British English. "utilized". The British spelling "utilised" was used here instead.**
- **P2L4: "R^2". This refers to a coefficient of determination which is a means of modelling accuracy. "Coefficient of determination" was added in the text for explanation.**
- **P2L38: I think multiple references should be separated by a semicolon. Check throughout text. Agreed, semicolons were added to separate multiple references throughout the text**
- **P2L51: "et al" missing period, check throughout text. We thank the review for pointing this error out. The period was added to all "et al."**
- **P3L35: "and" used for reference separation. Many thanks, the "and" was removed.**
- **P4 2nd Paragraph: For better structure of the paragraph, you might want to describe LiDAR first and SAR afterwards rather than mixing them both. The authors would like to thank the review for this suggestion. The paragraph was restructured to discuss LiDAR first followed by SAR.**
- **P5L23: "tonnes/ha", first of many different abbreviations to come (t/ha, T/ha), which are actually never properly introduced. Apologies for the inconsistency here. We expanded the unit of measure to "tonnes per hectare". The term was first and adequately introduced on Page 3 lines 10-12 but the inconsistency may have caused**

confusion. Paper was reviewed and all the different abbreviations of "tonnes per hectare" have been standardized to t/ha.

- **P5L25: "woodlands and savannahs possess". Agreed, the necessary change was made**
- **P5L28: "landscape, we" you have a tendency to (very) long sentences, which is sometimes hard to read. Check either if the usage of commas or splitting might be appropriate. The authors agree with the reviewer's recommendation. Commas were added to make the sentence flow a lot better. Throughout the manuscript, efforts have been made to reduce long sentences by splitting them up into multiple sentences.**
- **P6L14: long sentence due to heavy usage of () to include side information. This also makes it hard to follow the idea of the sentence. Agreed, parentheses have been removed and replaced with commas. The sentence was also split into two sentences to reduce length and improve word flow.**
- **P6L35: outline? What is to come in the next pages and sections? The authors appreciate the reviewer's suggestion. A paragraph outlining the contents of the following sections was added at the end of the introduction section on page 6.**
- **P6L47: "figure 1", starting with uppercase letter? Check throughout text. The authors agree with the suggestion and have modified all the "figure" and "table" labels in the text by using uppercase letters (e.g. Figure 1).**
- **P7L2: "T/ha". Many thanks for pointing this out. The tonnes per hectare abbreviations have all been standardised to "t/ha"**
- **P8L6: "figure 5", this is the 2nd reference to a figure. The authors understand the reviewer's issue here. To avoid confusion and inconsistency in the figure numbering order, the words "figure 5" were removed from the beginning paragraph of section 3 as the methodological schema is only presented at the end of the Section 3. The following sentence was also added at the end of Page 16 to properly introduce the methodology schema in Figure 5 ["The complete methodology have been summarized and compiled in the form a methodological schema (Figure 5)"].**
- **P8L14: "Random Forest algorithm", might be a good place for the reference, as it is the first mentioning. Agreed, the reference "Breiman, 2001" was added at Page 8 Line 8 in the manuscript.**
- **P9 Section 3.2: Some short description about time overlap of the field data and SAR data would be good, particularly as acquisition times are within almost 3 years. How is this handled, compensated, expected effects, influences? The authors would like to thank the reviewers for this suggestion. A paragraph on possible error sources was included in the discussion to address this issue. The difference in acquisition of the RADARSAT-2 and the collection of field / LiDAR data were unfortunately related to research project logistics such as the access to specific datasets through collaborations with the different**

space agencies. Also the last ALOS winter data were acquired in 2010 prior to the sensor failure. At the scale in which the study was carried out (i.e. Greater Kruger National Park region) it was expected the difference in vegetation structure between these different datasets (acquired at different years – 2009 to 2012) did not introduce any major noise or error in the results which is evident according to the results (e.g. 2010 L-band model trained and validated with 2012 LiDAR produced results expected for this type of environment – Colgan et al., 2012 and Mathieu et al., 2013).

- **P9L19: 100m X 100m vs. P9L25: 100 X 100m.** The authors would like to thank the reviewer for pointing this inconsistency out. All similar measurement entries have been standardised to the "100m X 100m" format.
- **P9L39: N?, R^2?, RSE? Is the weight per tree or species?** Apologies for the lack of explanation for these terms. The entire sentenced was modified to the follow: "Number of trees sampled =707; R^2 = 0.98; relative Root Squared Error = 52%; ranging from 0.2 – 4531 kg per tree". The abbreviation "R^2" was kept as is as it was clarified according to a previous comment.
- **P9L43: DBH?** The term DBH was expanded to "Diameter above Breast Height" which was then followed by the abbreviation "DBH".
- **P10L10: 25m X 25m vs. P10L26: 25m by 25m vs: P11L48: 25x25m.** All inconsistencies in displaying these types of terms have been standardised throughout the text according to the "25m X 25m" format.
- **P10L10: "small DBH".** This has been changed to "low DBH". The term "DBH" was elaborated upon according to the response to the comment P9L43.
- **P10L26: "DHB zones".** Apologies for the error. The words were changed to read "DBH zones".
- **P10L44: Y? assuming N is the same previously N=707?** The authors appreciate the reviewer's feedback. The Y/N here represents the presence of canopy cover (i.e. Y for Yes) or the absence of canopy cover (i.e. N for No). The text was modified to clarify this aspect. The reference to N is no longer used as potential confusion with N as number of sample.
- **P10L53: equation number? For instance: X+Y (1). And equations are part of the text, so check for usage of punctuation marks.** The authors thank the reviewer for this suggestion. We added an equation number ("Equation 3") and we made it part of the text with the appropriate punctuation marks ("ΣY/169X100")
- **P11L20: "5 X 5", P11L33: CAO?, P11L37: r^2? p?** Apologies for the lack of explanation for the abbreviations. The following text corrections were made: "5m X 5m", "Carnegie Airborne Observatory (CAO)", "R^2" and "p-value". The R2 term was elaborated upon according to the response to the comment P2L4.

- **P12L42: "tonnes per hectare (t/ha)", first proper introduction. The authors would like to thank the reviewer for pointing this matter out. The term "tonnes per hectare (t/ha)" was introduced on Page 3.**
- **P12L46: it is highly recommended to avoid using the same abbreviation for different things. H had to be introduced as horizontal polarization. Agreed, the abbreviation H was replaced with the abbreviation "HGT" to prevent confusion with the H from horizontal polarization.**
- **P13 Section 3.4: Make sure all used abbreviations are properly introduced. If only used once then maybe no abbreviation is needed. MGD? SLC? SRTM? RMS? GDEM? Agreed, all abbreviations have been revised and appropriately introduced in detail. No abbreviation, however, was given for GDEM as it stood for Global Digital Elevation Map which may have been in conflict with the Digital Elevation Model (DEM) term. It is also used only one time.**
- **P13L49: "3X3". This term was changed to "3 pixel X 3 pixel" for clarity and for keeping in line with the abbreviation format used throughout the manuscript.**
- **P14L13: "table 1". Many thanks, the upper case was applied to all figure and table labels throughout the text.**
- **P14L36: why not use the full polarimetric information available from Radarsat2? Maybe for comparison? The authors are grateful to the reviewer for bringing up this issue. The full polarimetric dataset was not used in this study as the goal of the study was to compare SAR datasets from the X-, C- and L-band frequencies, alone and in combination. We thus used similar standard products which could be compared and modelled in combination, and which were expected to produce the best results, i.e. dual polarisation HH-HV of winter (dry) datasets. Also full polarimetric modes with a large enough spatial coverage over our entire study area were not available for X- and L-band SAR datasets. For example, the fully polarimetric mode for ALOS PALSAR was limited in coverage (not continuous) and was only acquired during the spring season over our study area which was thus not ideal for our analyses (see results in Mathieu et al. 2013). Additionally, Mathieu et al. (2013) did report attempts to used polarimetric decompositions with the Radarsat-2 dataset, and showed that these did not yield higher modelling accuracies for mapping tree structural components compared the HH and HV polarisations.**
- **P14L50: "found to <u>be</u> the most". Agreed, the suggested change was made.**
- **P14L52: RF? Apologies, RF is the abbreviation for Random Forest. This was added to the term when it was introduced in response to comment P8L14.**

- **P15L10: "correlation coefficient (R^2)", little late? Agreed, R^2 was introduced earlier in the manuscript in response to the comment P2L4. Additionally, R^2 is the coefficient of determination and not the correlation coefficient. This error was corrected in the text.**
- **P17L34: "[…]", you might want to refer to Table 2 and the highlighted results for each metric. Agreed, to avoid restating the pertinent results for each metric, the following sentence was added to simplify matters: "(refer to the highlighted results for each metric in Table 2)"**
- **P17L40: "[…]", rather than listing the number, a little more description would be appreciated. Agreed, more descriptive but simple text was added to replace the numeric text. The following sentence was added: "[In comparison to the results for L-band alone, there was a relative improvement of 10% or greater for all three structural metrics in modelling accuracies]".**
- **P19L37: "however, still persist across riparian zones of minor tributaries."? Is this highlighted in the figure(s)? How do we know? How/where do we see it in the figure? The authors understand the confusion caused as these minor tributary features are not particular prominent in the map products. The minor tributaries can be identified by the linear features which branch off the main ridge (encircled area in figure 9iv). A rectangle was added to point out the most visible example of these features in figure 9iv and was also mentioned in the text for clarity.**
- **P19L49: AOI? Apologies, AOI stands for Area of Interest. This was included in the text.**
- **P20 Number of experiments/evaluation runs used to generate boxplots? 5, 10, 100. All the observations / SAR pixels corresponding to the LiDAR coverage were used to build the boxplots. In total 17559 observations/pixels were used to generate the boxplots with the outliers removed. This number was stated in the Figure 11 caption ("N=17 559") in the figure captions document which accompanied the manuscript submission. "N" was changed in the caption to "Number of pixels" to avoid confusion. Additional text was included at the end of section 3.6 to clarify this matter.**
- **P20L17: "spread", variance? The authors will like to thank the reviewer for the suggestion. Agreed. The word "spread" was changed to "variance".**
- **Regarding all results, how are they obtained? Single values or averages over multiple observations? The authors will like to thank the reviewer for this comment.**
- **How did you utilize the multi-temporal images for each SAR system? The authors would like to apologies if there was any confusion which came about regarding this matter. Multi-temporal image datasets for each sensor were not acquired or utilised in this study. We only acquired a single winter season mosaicked dataset for each of the SAR systems. These datasets differed in the fact that they were acquired at different years due to sensor constraints (e.g. ALOS PALSAR's last complete dataset for our area was**

2010 due to sensor failure in early 2011) and data access issues (RADARSAT images were only available from previous works).

- P21 1st paragraph could/should be moved to conclusions. Agreed, the 1st paragraph of the discussion section was moved to the beginning of the conclusion section
- P21L34: "of leaf-off of", please rephrase. Agreed, the text was rephrased to state the following: "Despite the leaf-off conditions of most trees…"
- P22L50: "will be needed"? Agreed, the text was modified to incorporate the reviewer's suggestion.
- P22L50: "these results but these results", please rephrase. Agreed, the text was rephrased to the following: "…exact cause of these trends but the overall results…"
- P23? Last paragraph: the second part of it introduces so much new information particularly about areas and their characterizations, which have not been mentioned earlier in any way. The authors thank the reviewer for pointing out this concern. Unfortunately this section included a lot of areas which were mostly familiar with the people with the local knowledge of the area and for the surrounding park managers (Kruger National Park and Sabi Sands Game Reserve) who are also researchers as well. This newly introduced information pertaining to the local areas within our study area was vital to verify the accuracy of our models and mapped products especially within the context of distinct landscape features (ecca shale patches and fence line contrasts etc.) known by local managers. This information can be removed from the manuscript, if recommended by the reviewer, to appeal to a wider audience who are not familiar with the study area.
- P23L14: "with higher SEP values". Agreed, the reviewer's suggestion was gladly accepted and the text modified accordingly.
- P24L4: "key area of interest E", so why hasn't it been used for experiments and evaluations? Agreed, the word "key" is misleading and was removed. AOI E represented a fence line contrast which consisted of high and low structural attributes within close proximity.
- References: Please check and make sure that reference style uses 'last name, first name initials, [first name initials, last name …]'. The authors would like to thank the reviewer for pointing this inconsistency out. The references list has been standardised to the following format: 'last name, first name initials [for the lead author], first name initials last name [for the secondary authors…]' (minus the comma after the second author's initials). This format is consistent throughout the reference list.
- P28: RF reference, consider L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, October 2001. Many thanks for the reference. The reference was added

to replace the website reference (Breiman, 2003). The appropriate changes were made in the text of the manuscript to reflect this replacement.

- **P33: Urbazaev reference…** Agreed, the reference recommended by the reviewer was added to replace the Urbazaev reference in the reference list.
- **Equation numbers for Appendix A and B:** The appropriate Equation numbers were added and their numbering order was also updated in the text.
- **Figures and their captions - Figure 2: DBH? 10X10m, 25X25m, use same presentation as in text:** The figure has been updated to display the changes suggested by the reviewer. A full name for DBH was also given in the figure text.
- **Figure 3, 4: N?** The authors agree with the reviewer's suggestion. "N" refers to the number of observations of the corresponding LiDAR metric and ground metric measurements used to make the graphs. This abbreviation "N" was expanded to "Number of observations" in the caption text.
- **Figure 7iii: "50-100" to ">50". Any reason that this interval is so large compared to the others? Add abbreviations from figure to caption.** The class intervals were chosen to best maximise the visual representation of the CC distributions which would complement the range expected on the ground. The study area is a landscape which has mostly a low to medium CC distribution with few high CC areas. Thus smaller intervals at the top of the CC range will make little difference to the overall landscape distribution of the CC trends and might even dilute the visibility of the lower to medium CC classes which are more prominent. Area of denser CC would also be more difficult to visualise. The abbreviations from the figure were updated in the caption as recommended by the reviewer.
- **Figure 8: legend for colours? Mentioned regions of plot in text could be marked and labelled. Caption: "region of AOI"? so AOI cannot be area of interest then.** The authors appreciated the feedback. A colour legend (e.g. "Low point density to High point density") was added to the figure 8. The following text was added on page 18, first paragraph of Section 4.2 to give more information in the manuscript regarding Figure 8 ("Figure 8 shows the AGB vs. CC scatterplot for AOI 'A', a dense forested site"). Additionally, white labels ("Tall tree forests" and "Coppicing thickets") were added to figure 8 to highlight the concepts discussed in the text. The words "region of" in the Figure 8 caption was removed as it caused confusion.
- **Recommended to mark areas in Figure 7. Otherwise it is quite difficult to see and compare, particularly with this over-/underestimation comparison. Also not 100% sure what you are trying to show here, as this over-/underestimation make it rather difficult.** The authors appreciate the comment. Over-/underestimation figures (Figures 9 and 10) were added to provide a visual representation of the model's performance in the two

main vegetation conditions (dense and sparse vegetation). These figures set up the more quantitative results of the boxplots in Figure 11. The authors still believe in the merit of keeping Figures 9 and 10 but if the potential confusion is too great for readers, the authors can consider removing the Figures 9 and 10 (and associated text), if the reviewer requests it. In the meantime, the AOI extents represented in Figures 9 and 10 have been roughly outlined as boxes near the letter 'A' and over the letter 'C' in Figure 7iii. This also removes the confusion of which AOI 'C' is being shown in Figure 10 as multiple 'C' AOIs have been shown in Figure 7 to indicate the different patches of gabbro intrusions. The caption for Figure 7 has also been updated.

- **Figure 9: colour legend, colours are similar, so hard to understand difference? The author would like to thank the reviewer for the comment. It is however, the authors' honest opinion that the colour legend used in both Figure 9 and 10 were distinct enough to separate the main error classes: overestimation, negligible error and underestimation. The minor over- and underestimation error classes, through different shades, can be distinguished in both figures.**
- **Figure 10: colour legend as in Figure 9, which area C in Figure 7, there are 2? The same treatment was considered for Figure 10's colour legend as in Figure 9. The confusion regarding which AOI 'C', in Figure 7, was represented in Figure 10 was resolved by adding their extent boxes in Figure 7.**
- **Table 2: if you provide the measure of a unit in the heading for each sub-table, why give it for each table entry again? Content of table is okay, it just feels a little unnatural to read. Quick comparison and evaluation for each metric is tricky as you have to scan 3 lines. This is of course a personal preference. The authors would like to thank the reviewer for the comment. The measurement units within each of the sub-tables have been removed as the unit of measure is already mentioned in the heading of each sub-table. The authors agree with the reviewer's comment about the awkwardness of Table 2. Another alternative was to reposition the sub-tables into a single table with the rows being the different frequencies/frequency combinations and the columns being the structural metrics. The new version of the table replaced the existing version.**

Reviewer #2:

- We would like to thank the reviewer for the positive comments pertaining to our manuscript. The feedback is also greatly appreciated.

Main Points/concerns:

- **The part concerning the SAR analysis and modelling is not clear which makes the procedure key element vague. The author appreciates the reviewer's comments. We**

believe that the revisions made in response to the main points, below, and the specific comments have added clarity to the SAR analysis and modelling sections.

- **I believe a detailed discussion of the merits of RF over other modelling methods will be very helpful. The authors agree with the reviewer. Although the performance of RF was tested against other algorithms in the Naidoo et al., 2014 IGARSS paper, a brief paragraph was added to the top paragraph on page 15 to highlight the advantages of RF over other machine learning algorithms. The following text was added: "Unlike other traditional and fast learning decision trees (e.g. Classification And Regression Trees or CART), RF is insensitive to small changes in the training datasets and are not prone to overfitting (Ismail et al., 2010; Prasad et al., 2006). Additionally, RF is less complex and less computer intensive in comparison to the high levels of customisation required for Artificial Neural Networks (ANN) and the long 'learning' or training times for Support Vector Machines (SVM) (Anguita et al., 2010). RF requires two main user-defined inputs – the number of trees built in the 'forest' or 'ntree' and the number of possible splitting variables for each node or 'mtry' (Ismail et al., 2010 & Prasad et al., 2006)."**
- **Modelling and analysis steps of RF were not fully discussed and the relation between the SAR data and retrieved structural elements is not clear. The modelling and analysis steps of the RF were elaborated upon in greater detail to remedy this issue. These detailed additions were made in section 3.5. Preliminary work indicated that the LiDAR was the preferred source for the calibration and validation of the SAR dataset modelling instead of the collected ground data (see last comment of this main points section for more information). Two sentences were added, in the discussion section, which summarised the relationship between ground data and the SAR directly (minus LiDAR use).**
- **The SAR data used in the modelling was not clear and did the authors use Intensity, Phase or the Complex SAR data in the modelling? Intensity SAR datasets were only used in the modelling procedures – corrections were made in the methodology sections to reaffirm this fact. The SAR datasets acquired for this study was stated in detail in Table 1.**
- **The methodology presented in this paper can only be applied if the LiDAR data is available along with the SAR data. During our preliminary analyses, we correlated the ground collected metric results (in this case, CC at the 100m plot scale) directly to the SAR dataset (in this case HH and HV L-band intensity data) and we found decent results (R^2 = 0.55; RMSE = 17.57%) but were not as high as when the LiDAR dataset was used an intermediate upscaling dataset (R^2 = 0.77; RMSE = 10.59%). If ground data is only available then it can be used without upscaling to the LiDAR which would, however, compromise the overall performance of the model. We used LiDAR as the**

source of model calibration and validation (after using the field data to validate the accuracy of the LiDAR first) as it yielded the highest accuracies in this study.

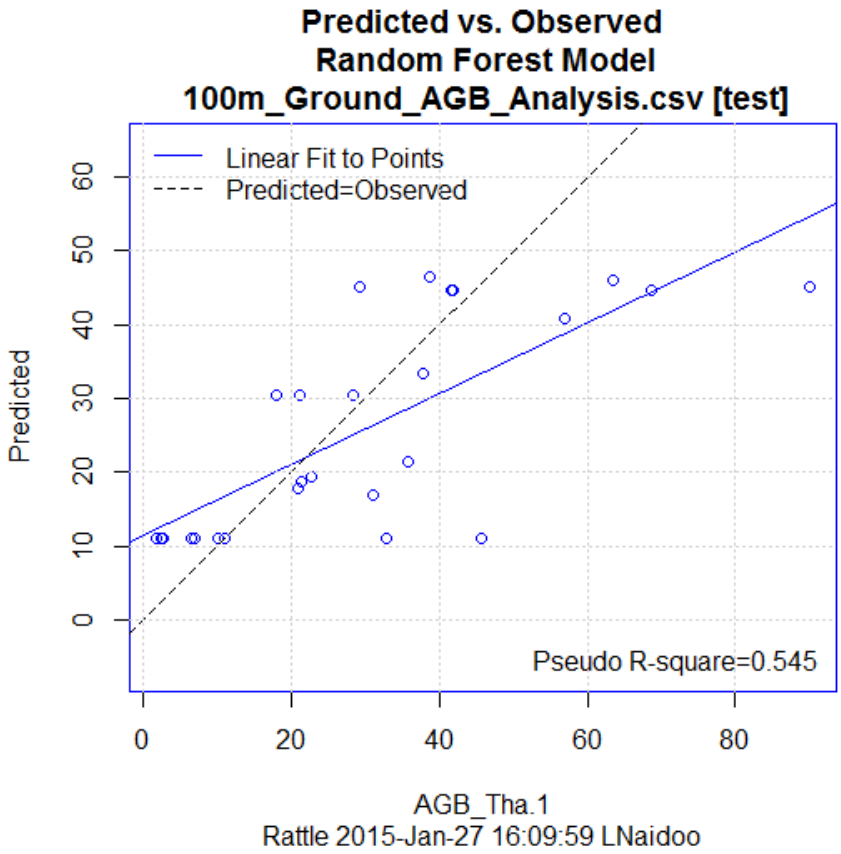
Specific comments:

- **P1L48: SAR phase is sensitive to atmosphere and cloud but intensity is not. It is not clear in the analysis which component of SAR data was used. The authors appreciate the comment. The insensitivity of the SAR intensity to atmosphere and cloud was added to the text, on page 4; 2nd paragraph, to clarify this fact: ["…offers an all-weather capacity, if SAR intensity information was implemented, to map…"]. The SAR intensity information was used in this study's analysis. This was clarified in the methodology section of the text.**
- **P6L17: wavelength (~15cm L-band)??? The wavelength of ALOS PALSAR L-band is 23.6cm an frequency 1270 MHz. Please cite references on the specifications of the SAR data used in this analysis. The authors apologise for the incorrect information displayed. The correct frequency was added ("~23cm L-band") and references were added to support the SAR dataset specifications in the Table 1 in the methodology section.**
- **P6L57: Is the accuracy of the proposed methodology depending on the canopy species? Not only on spatial distribution and density? The authors appreciate the comment made by the reviewer. It is believed by the authors that the accuracy of the methodology would be predominately influenced by the broader landscape group distributions (e.g. grasslands, woodlands, thickets and forest) rather than the canopy species (e.g. Acacia versus Combretum species) though the shape of the individual trees, which is a function of the tree's species, maturity and surrounding disturbances (e.g. coppicing through harvesting), could additionally influence the results. This needs to be explored in greater detail which falls beyond the scope of this paper.**
- **P8L18: What the authors mean by "the LiDAR-derived counterparts (error statistics and distribution)". The LiDAR derived counterparts refer to the LiDAR derived woody structural metrics/maps which was subtracted by the SAR derived woody structural metrics/maps to achieve the error maps discussed in section 3.6. The following was added to the text for clarification: "The SAR-derived woody structural metrics were then validated using the LiDAR-derived woody structural metrics (CC, TCV and AGB) to ascertain error statistics and error distribution."**
- **P8L57: the author stated that they made the observations in dry season (winter) but the LiDAR observations were made is late summer (table 1). What is the effect of this divergence on the overall results, especially that LiDAR interacts with the leaves and fine branches of the canopy? The LiDAR is the reference dataset. It was used to assess woody structural metrics which are generally required and measured by managers and**

ecologists (cover, woody biomass, volume from trees with leaves). A LiDAR dataset acquired in winter would not properly assess these woody structural metrics. We selected SAR datasets acquired in winter because the dry season have been shown to be the best to model cover and biomass (no noise from water content), see e.g. Mathieu et al., 2013 in the region, so winter SAR datasets act as the best predictor or proxy of the LiDAR-based metrics acquired when trees have leaves. This is not a source of error as such. One source of error in our reference dataset is the fact that the LiDAR was taken during the wet-dry transition season, and the yellowing or senescence process had just started (based on field observations) at that time especially in the Kruger region. This information was incorporated in the paragraph on the sources of error in the discussion section.

- **P8L57: the SAR acquisition dates have a gap of one or two years. Is the canopy parameters didn't change during these years? i.e. what effect on the overall accuracy of modelling?** The authors agree with the comment brought up by the reviewer. As with the comment above, the temporal difference in the acquisitions would have introduced a margin of error to the overall modelling results but the main tree structure would have been preserved between the years (which allow for the backscatter interactions) if the trees in the area haven't been extensively harvested. The authors relied upon the fact that the imagery were consistently acquired during the winter season (minimal interference) and that the study area did not undergo any major changes in the landscape (local knowledge of the area from the authors). The fact that the ALOS PALSAR satellite was not operational during this time was also a pity so the use of the selected L-band datasets was unavoidable. This information was incorporated in the paragraph on the sources of error in the discussion section.
- **P9L39: briefly explain the values in brackets.** The abbreviations (N, RSE etc.) were expanded and text added to bring extra clarity to the values.
- **P10L52, identify the equation by "eqn.1".** A correction was made in this regard. The equation was identified as "Equation 1".
- **P12L53: The validation of LiDAR against Ground truth AGB measurements (R2=0.63) and the validation of SAR against LiDAR (0.83). What do you expect of the validation of SAR against Ground truth measurements "the actual value I assume"?** The authors appreciate the reviewer's suggestion. Preliminary results utilising L-band SAR (best performing single SAR frequency) was obtained by modelling AGB (35% training, 65% validation), in RF, using the ground truth dataset, AGB collected at 100m plots, as the calibration and validation (cal/val) data sources and the L-band SAR intensity (HH and HV) as the input parameters. We found decent results (R^2 = 0.55; RMSE = 15.36t/ha; see graph below) but they were not as high as when the LiDAR dataset was used as an

intermediate upscaling dataset (R^2 = 0.78; RMSE = 6.05t/ha). The LiDAR was seen as the most appropriate dataset for model cal/val.



- **P13L53: …Please explain how the spatial resolution of ALOS is 12.5m as stated in the specified line.** **The authors understand the reviewer's concern. After the multi-looking step in the SAR processing, the three (X-, C- and L-band) SAR datasets were resampled to the final resolutions (see new Table 2). Oversampling factors were applied to each dataset, with respect to the resolution of the DEM used, to achieve this resampling which was conducted as part of the "gc_map" module in the GAMMA DIFF & GEO processing chain (http://members.chello.nl/~r.sugardiman/html/gc_map.html). Additional text was added to section 3.4 to clarify this matter. An additional table (a new Table 2) was added to document the changes in dataset spatial resolution throughout the different processing steps.**
- **P14L40, the spatial resolution "a varying number of data records…" please explain that. If the authors resampled all the data first then the aggregated grid cells should be equal. If the authors made the grid first, then how the authors co-registered the grid cells?** **The coverage of some of the SAR datasets didn't completely match the LiDAR coverage (as was the case with the C-band coverage which excluded the upper strip of LiDAR in Figure 1) and some of the SAR datasets had gaps between some of the acquired scenes (particularly the X-band coverage between two scenes). This difference**

in coverage resulted in some of the grid cells (which was developed according to the LiDAR coverage) being excluded from the analyses. The different SAR datasets were only clipped to the common coverage extent for the mapping process only but were kept separate with their full extents during the data extraction and modelling processes – this information was added to the manuscript in section 3.4 and section 3.5. The SAR datasets differed in coverage, especially C-band and L-band, so any resampling would not make the number of grid cells available consistent. Since the grid was developed in relation to the LiDAR (the grid started from the upper left most pixel position of the LiDAR's full coverage) and that the LiDAR was well aligned with the SAR datasets, any registration issues which did arose may have been negligible.

- **P14L52 to P15L4: The modelling procedure is not clear; please explain with further details to facilitate the reproduction for other researchers.** Agreed, this section of the manuscript has been reworked and expanded to give more detailed information on the RF modelling process.
- **P15L36, Figure 9 and Figure 10: The main advantage of using SAR is the 3d detection ability (i.e. AGB and TCV) and you chose to present CC for accuracy check, which can be easily estimated using other remote sensing methods. Why didn't you present an accuracy check of AGB and TCV? This should give an insight on the actual advantage of using SAR.** The authors understand the issue raised by the reviewer. In response, we replaced the original CC Error Figures 9 and 10 with the TCV counterparts. Additionally we added AGB and TCV error statistics to Table 4 (the old Table 3). The text associated with these figures and tables have been changed extensively in the text (methods, results and discussion sections). The CC box plots (Figure 11), however, remained unchanged as the CC ranges, coupled with vegetation height for a 3D interpretation, best illustrated the different vegetation structural cohorts (e.g. sparse forests and sparse veld, bush encroached thickets and closed forests etc.).
- **P15L38: Why did you use TCV in Figure 6 for model accuracy verification? While you stated that it lacked meaningful units. And P12L28 the author state that TCV was not validated with ground collected data.** The authors appreciate the comments made by the reviewer. TCV was chosen merely to complement the results of Table 2 which showed TCV yielding the highest accuracies compared to the other metrics. The TCV metric results, in Figure 6, showed a tighter and neater distribution in relation to the 1:1 line than other metric scatterplots, which made the comparison between the different SAR frequency and combinations easier. TCV was not validated due to the fact that it was particularly hard to measure the TCV metric on the ground as it was more related to the point cloud dataset obtained by the LiDAR.

- **P17L34: the authors repeating the information in Table 2. Agreed, the repeated text was removed from the manuscript.**
- **P17L40: It will be much easier to read if the authors presented the improvements in table form. Agreed, but, as requested by one of other reviewers, this part of the text was simplified and modified to be more descriptive rather than stating numeric improvements.**
- **P18L32, L34, L38: is it (ABG) or (AGB). The authors would like to thank the reviewer for pointing this inconsistency out. The correct abbreviation is "AGB" which has now been standardised in the text.**
- **P19L20: the authors repeating the information in Table 3 and figures 9,10,11. It would be more interesting and easier to read if the authors gave an insight of the results in the discussion. For example, if the authors recommended a favourable CC boundary to use SAR data and the expected accuracy of the results within this boundary, it would be much more meaningful. The authors appreciate the suggestions made by the reviewer. Unfortunately, the observed trends suggest more complex SAR interactions with the landscape which requires further investigation (e.g. through the simulation of SAR backscatter in a simulated environment) and could not be certainly summarised with convenient boundaries and thresholds. Such boundaries will only be speculative but additional information regarding broader SAR dataset usage recommendations were given on page 23, at the end of the first paragraph.**
- **P22L51: the authors recommended using L-band in dense forested environment while the accuracy shown in figure 9 is very poor? Please explain. Also please define "dense forested". How dense? Please give a metric value. Figure 9 did not report accuracy values, but showed over and underestimation of cover for various SAR frequency combinations. The L-band results did show some signs of major overestimation in the dense ridge but the results were considerably better than the X- and C-band results which showed widespread major underestimation. We made clear that the performance was ascertained relative to the other frequencies in the text. Figure 11 also illustrated a better performance of L-band in dense forested conditions. According to the criteria used in Figure 11ii, a "dense forested" environment was defined as ">70% CC". This was added to the text.**

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

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

[a]Ecosystem Earth Observation, Natural Resources and the Environment, CSIR, Pretoria, South Africa
Corresponding author contact details: LNaidoo@csir.co.za; (+27)12 841 2233
[b]Department of Geography, University of Pretoria, Pretoria, South Africa
[c]Remote Sensing Unit, Meraka Institute, CSIR, Pretoria, South Africa

[d]Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA
[e]Faculty 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 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 utilised. 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 (coefficient of determination or R^2 of 0.77), TCV (R^2 of 0.79) and AGB (R^2 of 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 CO_2 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 woody canopy cover (CC), tree height, above ground biomass (AGB) or total woody canopy volume (TCV), 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 hectare or t/ha (Bombelli et al., 2009). Woody canopy cover (i.e. the percentage area occupied by woody canopy) 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 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; 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 (the local variance of an image related to its spatial resolution – Nichol & 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, which refers to the change in spectral properties or character of a target, due to seasonal dynamics, clouds and haze. These spectral variations are prevalent in the rainy season of African summers with veld fires in the dry winter, and in shadowed areas, which results from terrain topography and tree canopies, at fine resolutions and in mixed wood-grass pixels at the medium and coarser resolutions. Active remote sensing systems such as Light Detection and Ranging (LiDAR) and 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. Although airborne LiDAR provides detailed tree structural products it relies on the availability of aircraft infrastructure, which is not always available in Africa. Satellite LiDAR is also currently not available. On the other hand, 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). A SAR-based approach offers an all-weather capacity, when using SAR intensity, to map relatively large extents of the woody component, which cannot be easily achieved with airborne LiDAR (Mitchard et al., 2011).

Polarization, which refers to the orientation of the emitted and received signal, and frequency of SAR data play important roles in sensing vegetation structure. Multi-polarized SAR systems emit and receive in HH, HV, VH and/or VV with H referring to a horizontal wave orientation and V referring to a vertical wave orientation. This allows 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 and its "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, such as 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, such as 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 per hectare 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; Mougin et al., 1999; Collins et al., 2009). Despite the success achieved in these various studies via combining different SAR wavelengths (Mougin 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 and 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, ~23cm L-band, SAR datasets, in a modelling approach, will yield an improved assessment of woody structure. This idea is based on the assumption that X- and C-band SAR signals interact with the finer woody structural constituents such as leaves and finer branchlets, typical of the shrubby/thicket layer, while the L-band SAR signal interact with the major tree structural components such as trunk and main branches which are 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?

   More specifically, the investigation of the interactions of the different SAR frequencies, and their possible combinations, across the different vegetation patterning and structural classes, such as grasslands, thickets and forests, will pin-point the effective application of the different SAR frequencies and their possible combinations in Southern African savannah landscapes.

The study is broken down into various sections. Section 2 describes the study area under investigation. Section 3 and subsections focus on the material and methodology which outlines the remote sensing datasets used, field datasets collected, LiDAR and SAR pre-processing and metric generation, modelling protocols, mapping and finally validation and

error assessment. Section 4 describes the modelling, mapping and error results while sections 5 and 6 discuss the main study outcomes and concluding remarks, respectively.

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' E longitude, 24°33' to 25°00' S latitude). 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 60 t/ha (Mathieu et al., 2013). The Sabi Sands Wildetuin 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). 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 intensity backscatter 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 (RF) algorithm (Breiman, 2001). 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 woody structural metrics (CC, TCV and AGB) to ascertain error statistics and error 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 intensity datasets (summarized in Table 1) were acquired to cover the study transect shown in Figure 1. 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 m^2 from a flying altitude of 1000m above ground level (Asner et al., 2012). In comparison with the LiDAR dataset, the SAR images were acquired during the winter 2009 (RADARSAT-2), 2010 (ALOS PALSAR), and 2012 (TerraSAR-X). Unfortunately, the last ALOS PALSAR winter scenes were acquired during 2010 in the study area, and no RADARSAT imagery were available closer to 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 100m 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 – Equation 1 in Appendix A). The allometric equation was developed following destructive harvesting of 17 savannah tree species present in the study area (Number of trees sampled =707; R^2 = 0.98; relative Root Square Error = 52%; ranging from 0.2 – 4531 kg per tree, Colgan et al., 2013). Tree height was measured using a height pole and Laser vertex/rangefinder, while stem diameter was measured using callipers and Diameter above Breast Height (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 constrains 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 X 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 X 10m area positioned at the inner corner of the 25m X 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 low 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 (Equation 2 in 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 X 25m plot and/or the south west 25m X 25m plot (DBH 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, the presence of canopy cover was determined using a 5m pole placed vertically above each sampled points every 2m along the transects. At each sampled point the presence of cover was coded as Y. For plot level canopy cover, in terms of percentage at the

25m X 25m scale, the CC presence and absence data were subjected to the formula below (Equation 3):

$$(\Sigma Y/169)X100 \qquad \text{Equation 3}$$

Where Y represents the presence of cover data. The value 169 represents the total number of sampling points in a 25m X 25m plot conducted at 2m sampling increments. A total of 37 (25m X 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 5m X 5m 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 Carnegie Airborne Observatory (CAO) LiDAR data were validated against field height measurements of approximately 800 trees. There was a strong relationship (R^2 = 0.93, p-value < 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 25m X 25m) 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 X 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 HGT x CC LiDAR metric (where HGT 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 (19.2 t/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 intensity images (X-, C- and L-band) were pre-processed according to the following steps: multi-looking, radiometric calibration (conversion of raw digital numbers into sigma naught (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 (Gamma Remote Sensing, Copyright © 2000-2011) for the Dual Polarised TerraSAR-X X-band (StripMap, Level 1b, Multi Look Ground Range Detected), Fine Quad Polarised RADARSAT-2 C-band (Single Look Complex) and Dual Polarised ALOS PALSAR L-band (Level 1.1) data. A 20m Digital Elevation Model (DEM) and a 90m Shuttle Radar Topography Mission (STRM) 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 Root Mean Square (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 Global Digital Elevation Map 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. All datasets were resampled, using a bicubic-log spline interpolation function, to their final map geometry resolutions. This was achieved by applying a DEM oversampling factor (DEM resolution / Final image resolution) to the multi-looked SAR datasets which was set in the "gc_map" module under the GAMMA Differential Interferometry and Geocoding package. The original pixel size, multi-looking factors used in the pre-processing, modified pixel size (after multi-looking) and the final pixel size (i.e. map geometry) of the different SAR datasets were summarised in Table 2. Finally, a Lee filter (3 pixel X 3 pixel filtering window) (Lee, 1980) was applied to the images. It is important to note that the full extents varied for the different SAR datasets due to sensor coverage programming and specifications (Figure 1).

Insert Table 1

Insert Table 2

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 intensity) 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 and Table 2) without the need for pixel level fusion procedures. A regular spatial grid made up of 105m resolution cells at 50m distance spacing was created in QGIS 2.2 (Quantum GIS, Copyright © 2004-2014) and applied over the datasets. 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, 2001) was found to be the most robust and efficient, in terms of running time and accuracies (Prasad et al., 2006; Ismail et al., 2010). Unlike other traditional and fast learning decision trees (e.g. Classification And Regression Trees or CART), RF is insensitive to small changes in the training datasets and are not prone to overfitting (Ismail et al., 2010; Prasad et al., 2006). Additionally, RF is less complex and less computer intensive in comparison to the high levels of customisation required for Artificial Neural Networks (ANN) and the long 'learning' or training times for Support Vector Machines (SVM) (Anguita et al., 2010). RF requires two main user-defined inputs – the number of trees built in the 'forest' or 'ntree' and the number of possible splitting variables for each node or 'mtry' (Ismail et al., 2010 & Prasad et al., 2006).

RF was applied, using R rattle data mining software (Togaware Pty Ltd., Copyright © 2006-2014), to the data with 35% of the data being used for model training and the remaining 65% being used for model validation. For the modelling process, the SAR frequency datasets were selected as the input (independent) variables while the LiDAR derived metrics

were selected as the target (dependent) variables. The random forest models were built using the default setting parameters ('ntrees' = 500 and 'mtry' = √# SAR predictors) and the trees were allowed to grow without pruning. Predicted versus observed scatterplots and validation scores were outputted to calculate the model accuracy statistics. The coefficient of determination (R^2), Root Mean Square 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.

The best performing RF model, for each woody structural metric, was applied to the relevant SAR imagery, which were all clipped to a common coverage, resampled (pixel aggregate) to a common resolution of 12.5m to match the coarsest L-band and stacked, by using a mapping script. This script was developed in the R statistical software (Version 2.15.2, The R Foundation for Statistical Computing, Copyright © 2012) which utilised the combination of the 'ModelMap', 'Random Forest' and Geospatial Data Abstraction Library (GDAL) modules. The map products were imported into ArcMap 10.1 (ESRI, Copyright© 1995-2014) and displayed in discrete class intervals (total of 6 classes) to best illustrate the tree structural metric distribution representative of the entire modelled ranges.

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 statistics and maps were created by subtracting the LiDAR-derived and SAR-derived woody (LiDAR – SAR) structural metric maps for TCV, AGB and CC. The SAR derived metric maps were resampled to 25m, via pixel aggregate, to match the LiDAR metric spatial resolution first before the subtraction. The error statistics for all

metrics were documented but the TCV error maps were chosen for presentation over CC due to the metric's three dimensional properties which would best capture the SAR backscatter interactions. AGB error maps, however, were not displayed due to the 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). For ease of interpretation of the error statistics and maps, the error values were grouped into 5 main groups using intervals which best covered the error range observed in the different metrics. These groups were major overestimation, minor overestimation, negligible error, minor underestimation and major underestimation.

Additionally, 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). The combined use of CC and vegetation height metrics best described these structural cohorts than the use of AGB and/or TCV metrics. 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. A total of 17559 difference pixel values were used to generate the boxplots with the outlier values being removed. 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.

The complete methodology have been summarized and compiled in the form a methodological schema (Figure 5).

Insert Figure 5

4. Results

4.1 Modelling Accuracy Assessment

Insert Table 3

Insert Figure 6A-G

Table 3 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 (refer to the highlighted results for each metric in Table 3). In comparison to the results for L-band alone, there was a relative improvement of 10% or greater for all three structural metrics in modelling accuracies when the shorter wavelength datasets (X- and C-band) were added. However, 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 AGB vs. CC scatterplot for AOI 'A' (Figure 7), a dense forested site. The point cloud generally displays a high correlation between the 2D (CC) and 3D (AGB) variable, but also a triangular shape with an increasing base as the CC increases up to 75% (highlight by the white labels in figure 8). Hence, dense cover conditions (CC>70%) are characterized by AGB values varying from moderate (35-40 t/ha) to high (>60 t/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 TCV error maps for dense forested (black box near 'A' in Figure 7iii) and sparse gabbro (black box over 'C' in Figure 7iii) sites were presented in Figures 9 and 10, respectively. Total CC, TCV and AGB 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 4).

Insert Table 4

Insert Figure 9i-v

Insert Figure 10i-v

In Table 4, there is a noticeable decline in major overestimation and major underestimation with an increase in negligible error for all three metrics from shorter wavelengths (X-band to C-band) to the longer wavelength (L-band). For all metrics, the X+C+L-band combined scenario further reduced major overestimation and marginally increased negligible error but at the cost of an increase in major underestimation in comparison to the L-band results. The TCV metric, under L-band and X+C+L-band scenarios, illustrated the most noticeable reduction in major overestimation and underestimation, in comparison to the other metrics, but at the cost of a higher percentage of minor underestimation (~60% between 10 000 to 50 000 TCV units). The greatest percentage increase in negligible error (-5t/ha to 5t/ha) was noticed in AGB metric for the L-band and X+C+L-band combined scenarios. More specifically for the TCV metric, under dense forested conditions (Figures 9i-v), the X-band scenario (Figure 9i) illustrate major TCV underestimation. C-band results (Figure 9ii) indicate an overall decrease of patches of major TCV underestimation but some of these have been replaced with major TCV 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 TCV underestimation (10 000 to 50 000 TCV units) and negligible TCV error (evident in Table 4). Finally, the X+C+L scenario in Figure 9iv illustrated noticeable increases in the negligible TCV error coverage, especially over the dense green ridge visible in the LiDAR TCV of Figure 9v, but also indicated an increase in major TCV underestimation over dense vegetation patches north of the ridge (see encircle area in Figure 9iv). Patches of major TCV overestimation, however, still persist across riparian zones of minor tributaries (rectangle area in Figure 9iv). 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 TCV overestimation for the sparse vegetation areas and major TCV underestimation for the dense forested patches (see encircled area in Figure 10i). The L-band scenario (Figure 10iii) illustrates a drastic improvement with an extensive increase in negligible TCV error across the Area of Interest (AOI). Across patches of dense vegetation, major TCV 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. More quantitative results (box-plots, Figures 11i-ii) were introduced next to further assess the individual SAR frequency 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 variance, 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 (CC <40% and >3m) conditions.

5. Discussion

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 forests (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 leaf-off conditions 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. As an aside to the modelling results, CC and AGB field data were initially investigated as a LiDAR-substitute for SAR model calibration and validation but preliminary results showed poorer modelling accuracies (R^2<0.60) in comparison to the LiDAR derived results. This demonstrated the importance of extensive LiDAR coverage as the preferred source for modelling.

The three metric total percentage error statistics (Table 4), the TCV 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 metric overestimation error (mean error and variability) under sparse vegetation conditions while reducing overall metric 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 metric error budget by reducing the overall mean error and the overall variability of the error under most vegetation structural conditions. Additionally, L-band and X+C+L-band were more suited for assessing the 3D metrics (TCV and AGB) than the single 2D metric (CC) with the highest percentage of negligible AGB error and lowest percentages of major TCV under- and overestimation being observed. These results can be supported by the fact that the L-band was expected to penetrate deeper and interact more with 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 the smaller canopy elements (Rosenqvist et al., 2003). Further investigation will be needed to ascertain the exact cause of these trends but the overall results, however, advocate the suitability of the L-band over C- and X-band for analysing dense forested environments (>70% CC with an expected error ranging from ~7% to ~18%) 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). In the absence of L-band data, C-band has proven to be effective in sparser cover, i.e. less than 40% CC, savannah environments which coincided with the recommendations made by Mathieu et al., 2013.

Among the three structural metrics, TCV was consistently modelled with higher accuracies, amongst all seven SAR scenarios (Table 3). 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 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; Wessels et al., 2013). The 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. The general range of these tree structural metric values also agreed with the ranges reported in other related studies conducted in this savannah region (Mathieu et al., 2013; Colgan et al., 2012).

Although overall modelling and mapping results yielded favourable accuracies, it is, however, important to acknowledge the different sources of error which were introduced in this study. The first error source was the temporal difference between the acquisition of the SAR predictor datasets and the reference datasets such as collected field data and/or LiDAR datasets. This was unavoidable due to sensor failure (e.g. ALOS PALSAR in early 2011) and logistical restrictions to the current research project (e.g. specific RADARSAT-2 datasets available from collaborations). Although there has been documented evidence of big tree loss in the study region (Asner and Levick, 2012), no major error was observed in the modelling results, especially when the 2010 L-band model was trained and validated using 2012 LiDAR data which produced expected results for this environment (Colgan et al., 2012; Mathieu et al., 2013). This loss in trees which occurred during the different SAR dataset acquisitions times (between 2009 and 2012) may have also introduced a certain margin of error in the modelling results. It was expected, however, that the main structure of the remaining vegetation would not have changed prominently enough to extensively vary backscatter target interactions between the different acquisition times. A final source of

error was introduced by the fact that the LiDAR reference dataset, which was set to target woody canopies with complete foliage, was acquired during the wet-dry transition season where the senescence process had just started. This may have resulted in a distorted representation of the woody structural metrics expected on the ground. Understanding these sources of error will help improve future studies by promoting the creation of more accurate models.

6. Concluding Remarks

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).

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 recent 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., S.R. Levick, 2012. Landscape-scale effects of herbivores on treefall in African savannas. *Ecological Letters*, 15 (11), 1211-1217

- 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., 2001. Random forests. *Machine Learning*, 45 (1), 5-32

- 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. Multifequency 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

- Schmullius, 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, B. Leblon, 2013. Mapping of fractional woody cover using full, dual and single polarimetric L- and C-band datasets in the Kruger National Park region, SA. – Proceedings CD of ESA Living Planet Symposium, 09-13 September, Edinburgh, UK.

- 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

$$M = 0.109D^{(1.39+0.14\ln(D))}\ HGT^{0.73}\ p^{0.80} \qquad \text{Equation 1}$$

Where M = biomass in kilograms per hectare, D = Diameter above breast height (DBH) in centimetres, HGT = height of tree in metres and p = mean wood specific gravity (fixed at a mean value of 0.9) which is unitless.

Appendix B

$$\text{Total 25m X 25m AGB plot} = Q + S + (T*6.25) \qquad \text{Equation 2}$$

Where 'Q' is the total AGB of stems ≥ 10cm DBH, 'S' is total AGB of stems between 5 and 10cm DBH and 'T' 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 25m X 25m grid (i.e. DBH zone 2). So $625m^2$ (i.e. total area of the 25m X 25m sample plot) divided by $100m^2$ (area of the 10 X 10m subplot) is 6.25. All remaining stems within the 25m X 25m sample plot, which subscribed to the remaining DBH conditions (i.e. ≥5cm DBH), were measured and therefore did not require any up-scaling factors.

Figure captions

Click here to download Figure: Figure_captions.docx

Figure 1: 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.

Figure 2: 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)

Figure 3: Validation results of field-measured woody Canopy Cover (CC) versus LiDAR derived CC (above 0.5m height, Number of observations =37)

Figure 4: Validation results of field-measured Above Ground Biomass (AGB) versus LiDAR derived AGB (above 0.5m height, Number of observations =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 (AGB), ii) Total woody Canopy Volume (TCV) and iii) woody Canopy Cover (CC), using random forest. Letters A-F represents key areas of interest for discussion (for all three metrics). The black boxes represent the rough extents of the LiDAR-SAR CC scenario difference maps for Area of Interests 'A' and 'C'.

Figure 8: Scatterplot of Above Ground Biomass (AGB), y-axis, versus woody Canopy Cover (CC), x-axis, under dense cover conditions (plotted from pixels extracted from the Area of Interest 'A')

Figure 9i-v: LiDAR - SAR scenario difference (error) maps of Total woody Canopy Volume (TCV) for the Xanthia Forest Area of Interest (close to 'A'); v) 25m LiDAR-derived TCV map

Figure 10i-v: LiDAR - SAR scenario difference (error) maps of Total woody Canopy Volume (TCV) for the Gabbro Intrusions Area of Interest 'C'; v) 25m LiDAR-derived TCV map

Figures 11i-ii): Woody Canopy Cover (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) (Number of pixels = 17559)

Figure 1

Click here to download Figure: Figure1.docx

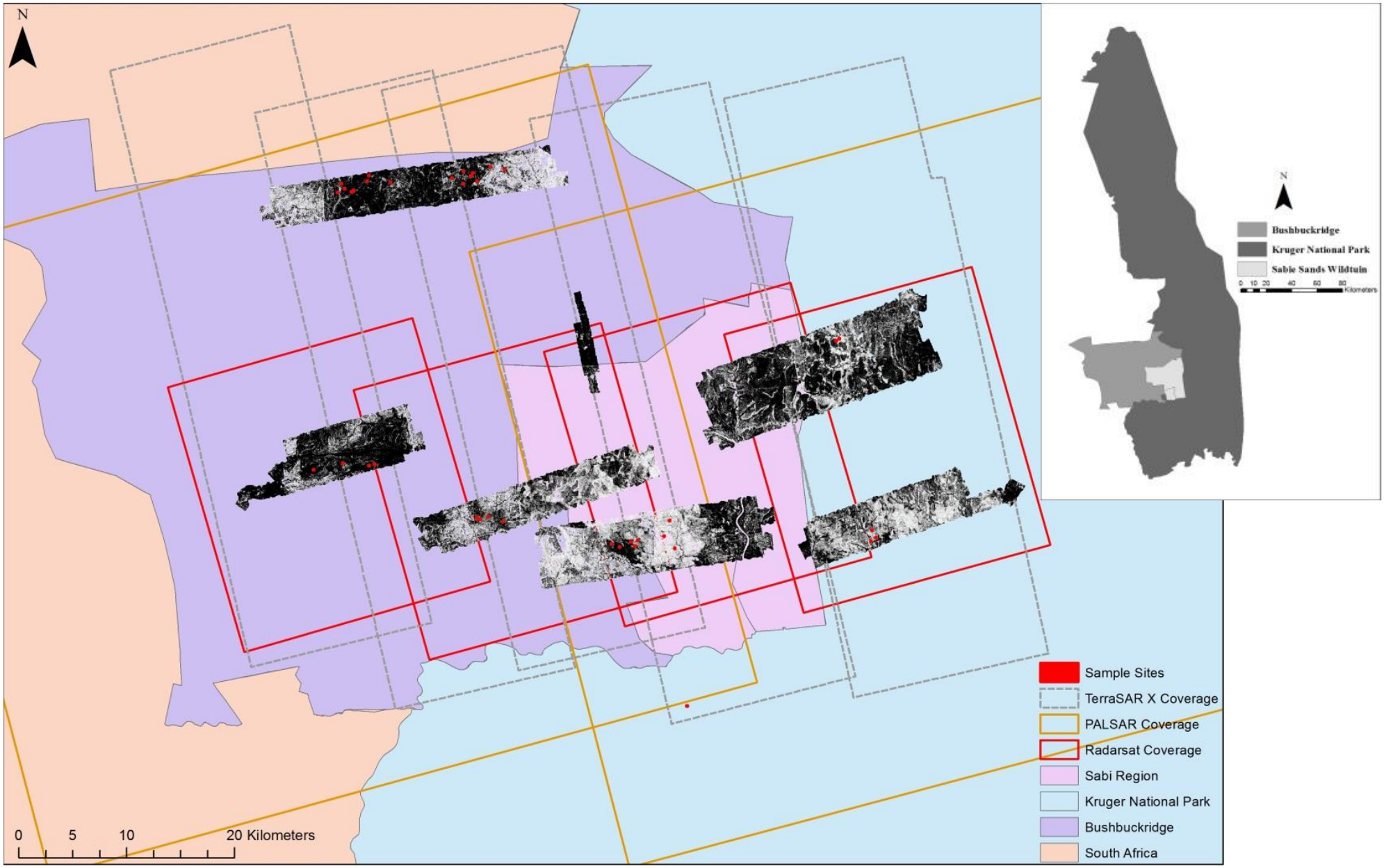
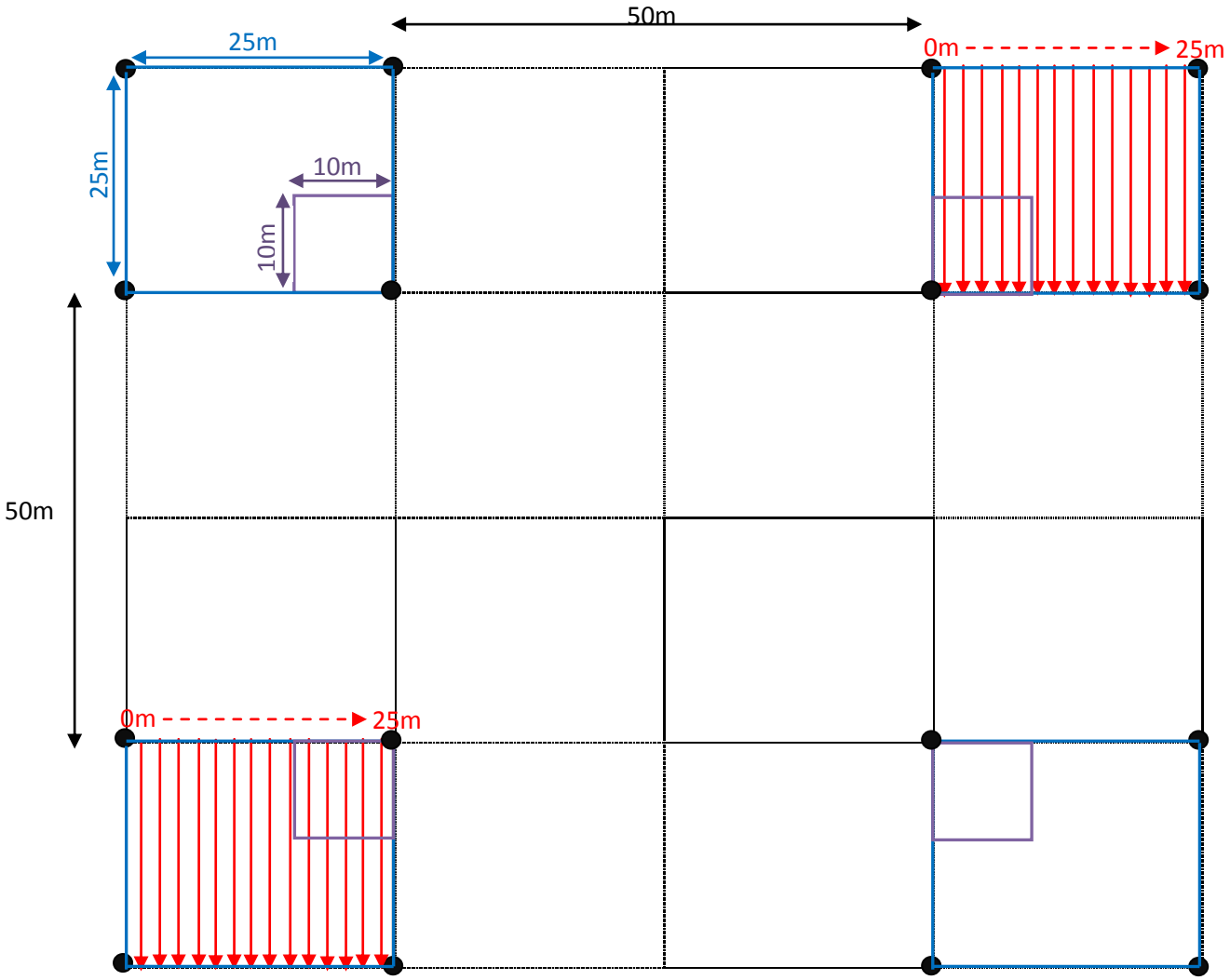


Figure 2

Click here to download Figure: Figure2.docx



Legend and Sampling Protocols for Tree Biomass and Cover:

- Corner pole markers
- DBH* Zone 1: 10m X 10m [Trees with DBH >= 3cm measured]
- DBH Zone 2: 25m X 25m [Trees with DBH >= 5cm measured]
- Line Transect and Vertical Densiometer methods (1m and/or 2m intervals)

** Note: DBH refers to Diameter above Breast Height (DBH)*

Figure 3

Click here to download Figure: Figure3.docx

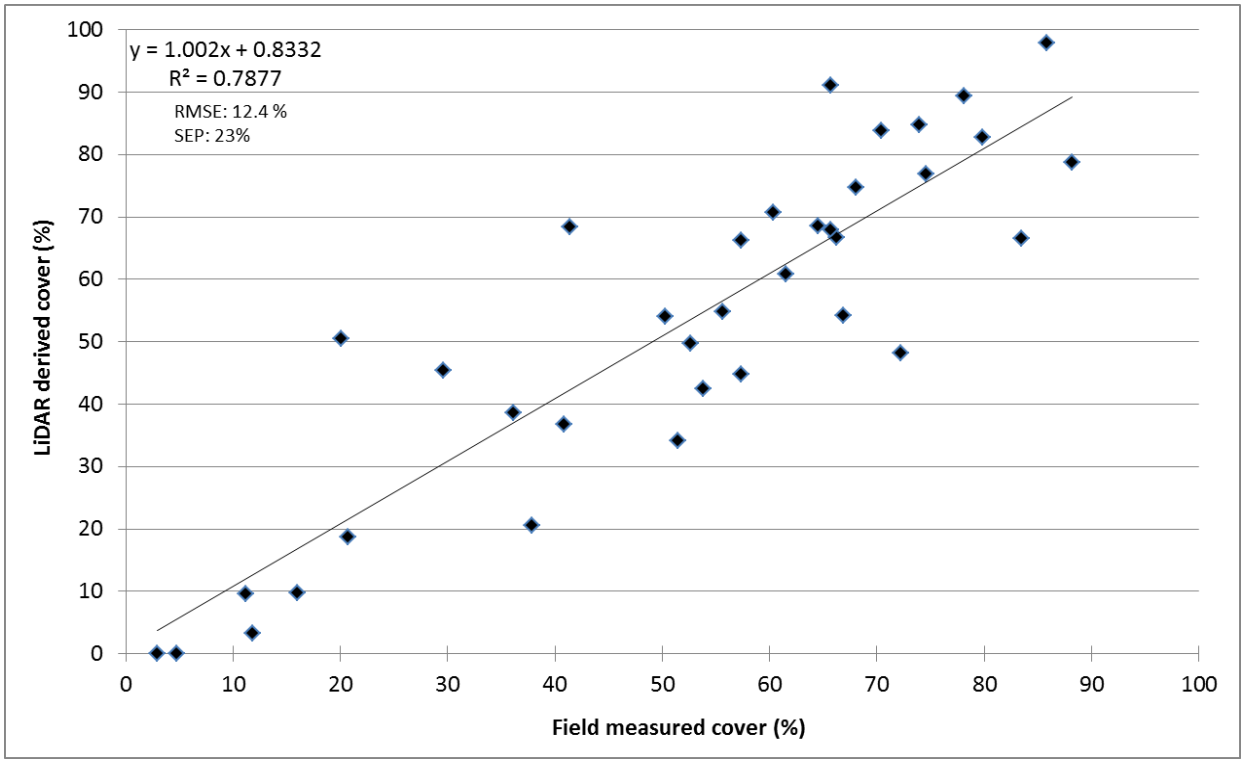


Figure 4

Click here to download Figure: Figure4.docx

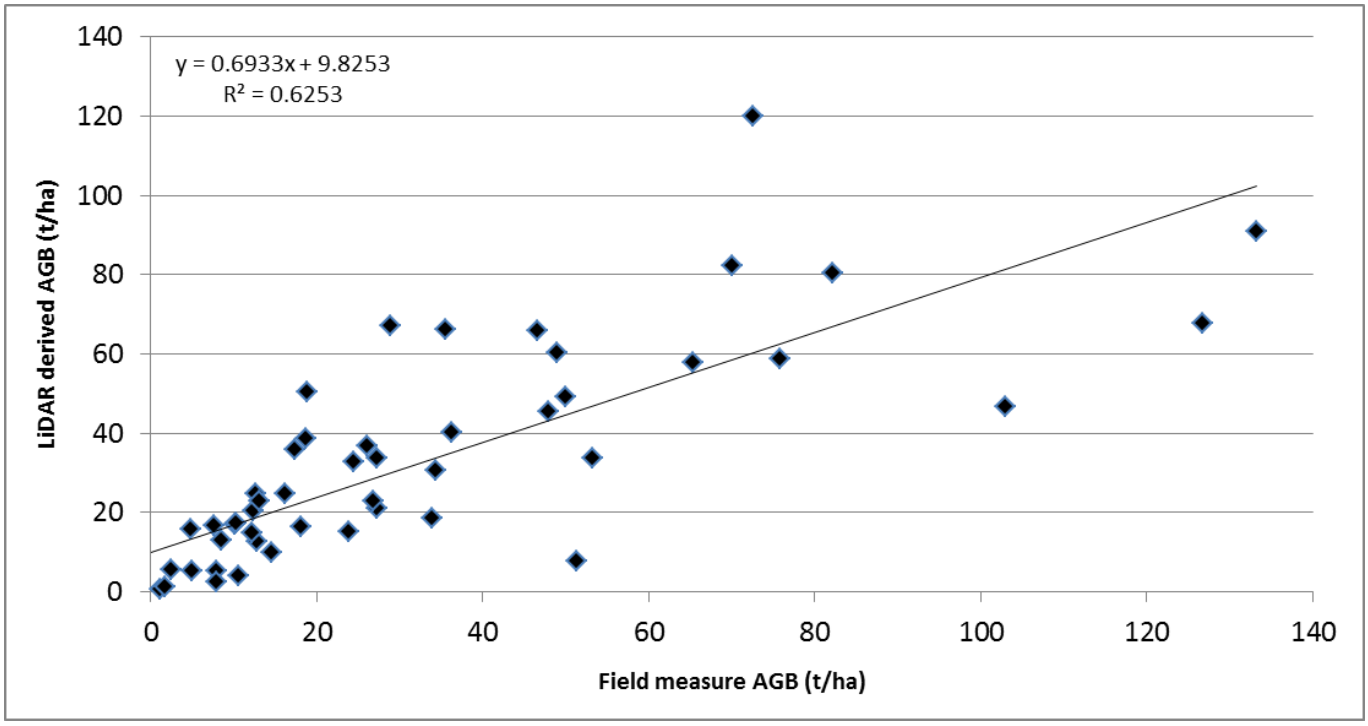


Figure 5

Click here to download Figure: Figure5.docx

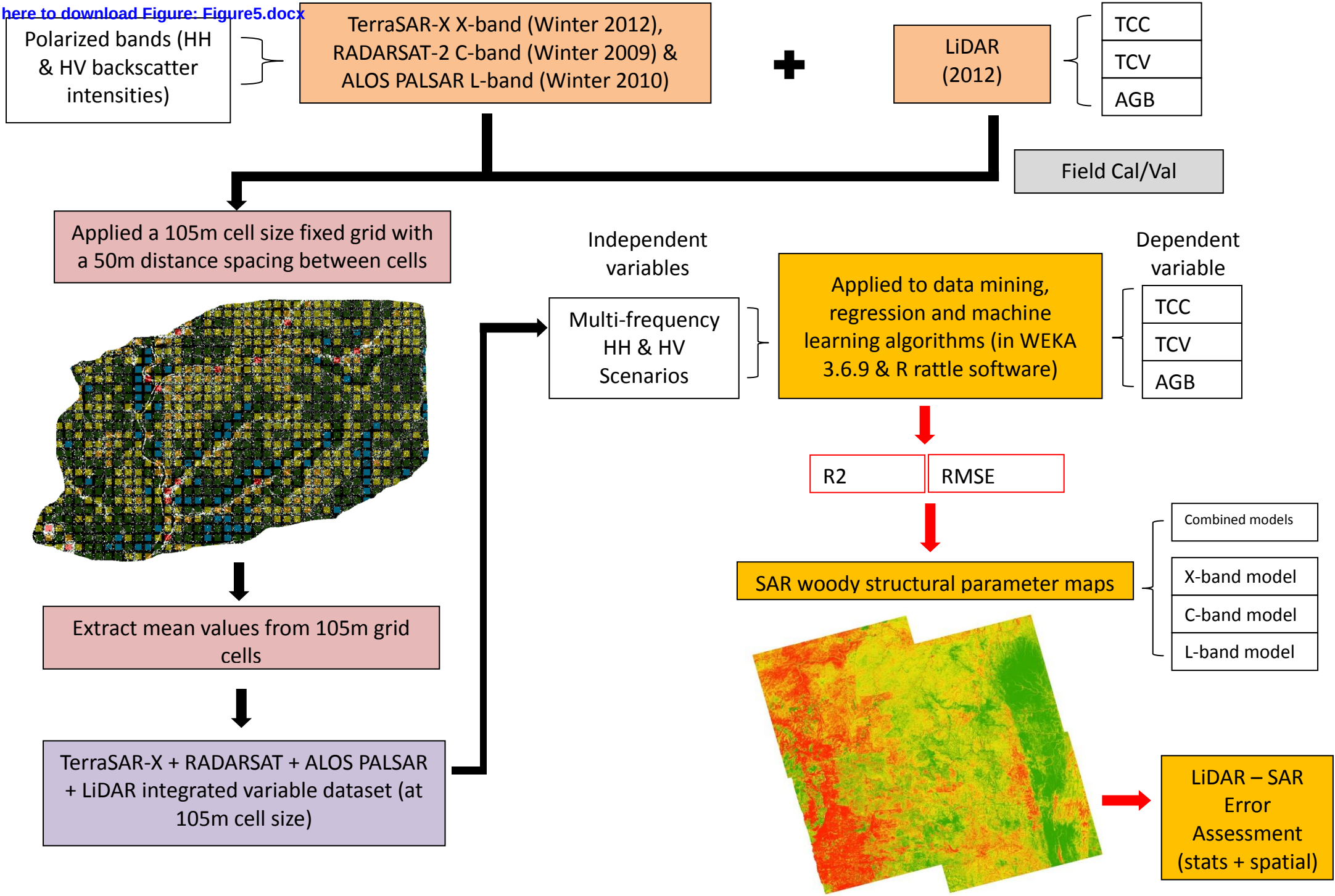


Figure 6

Click here to download Figure: Figure6.docx

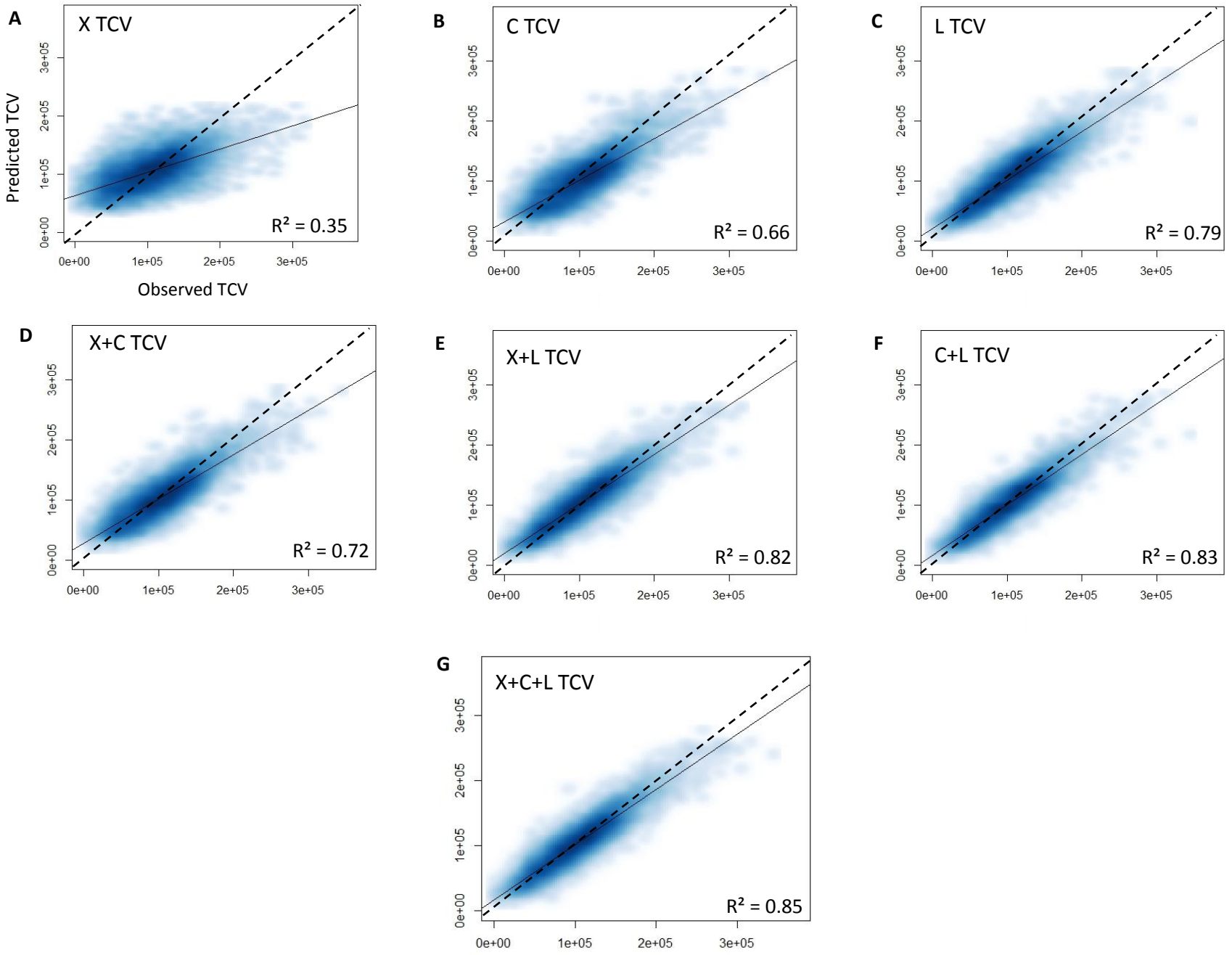


Figure 7

Click here to download Figure: Figure7.docx

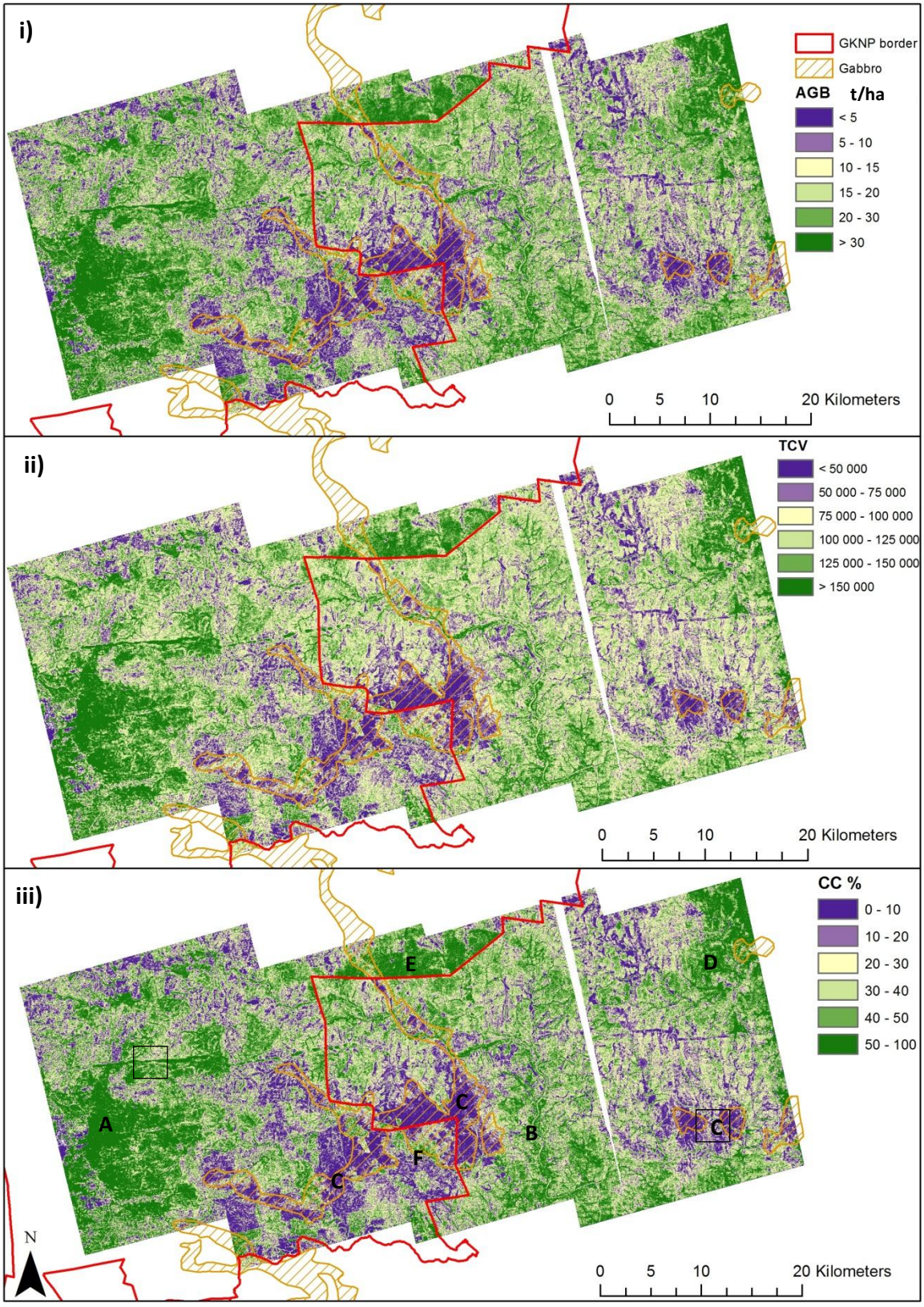


Figure 8

Click here to download Figure: Figure8.docx

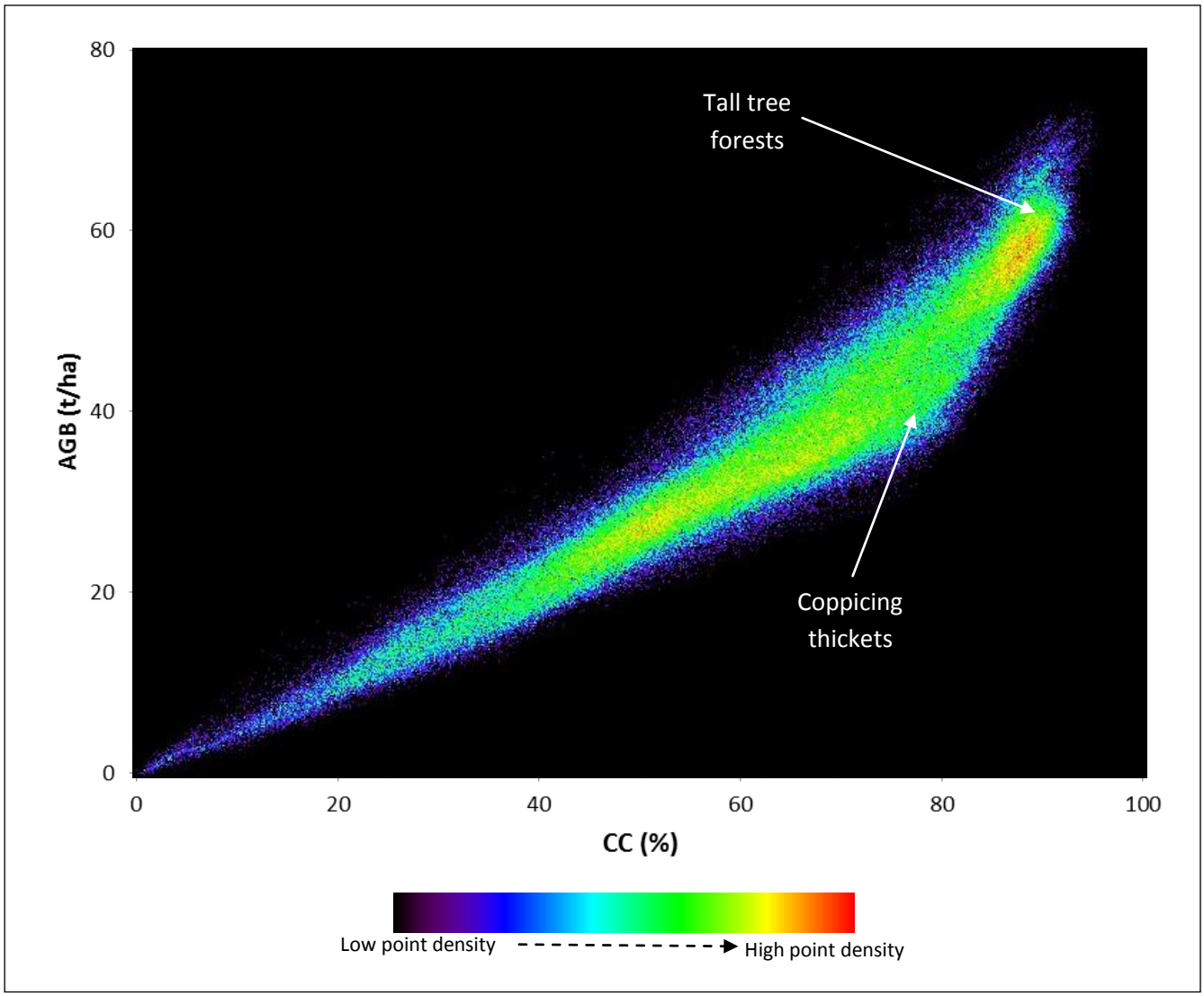


Figure 9

Click here to download Figure: Figure9_New.docx

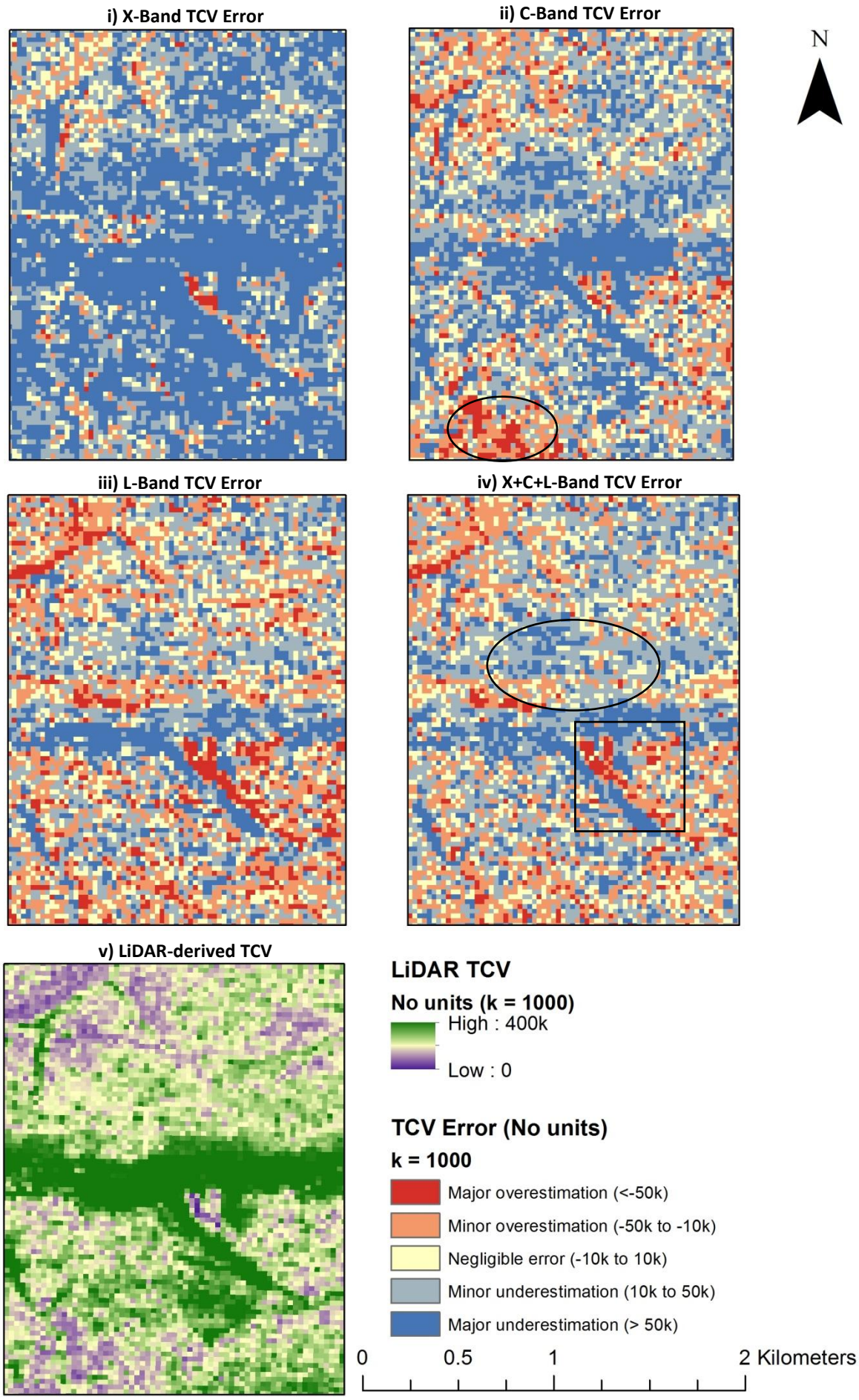


Figure 10

Click here to download Figure: Figure10_New.docx

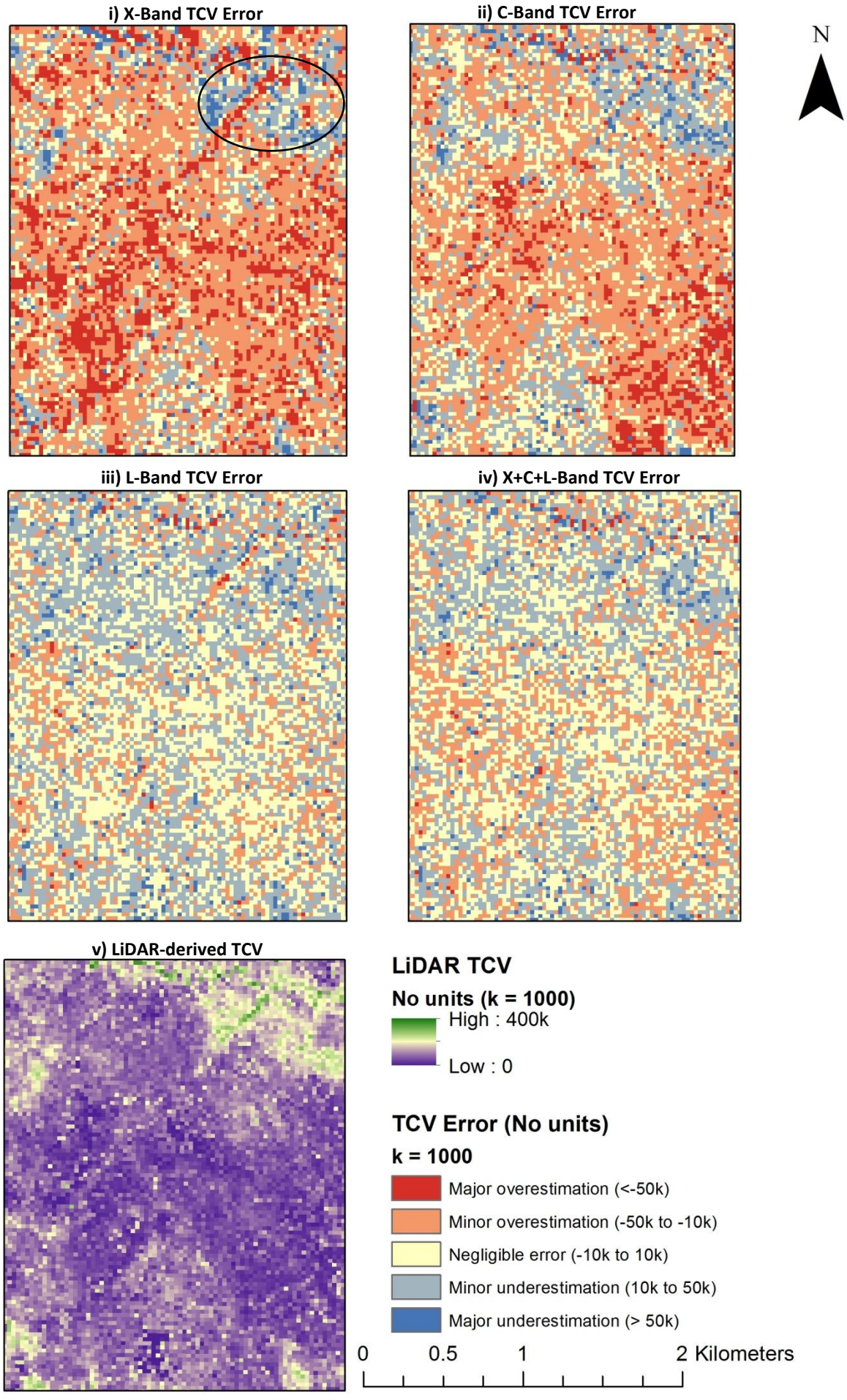
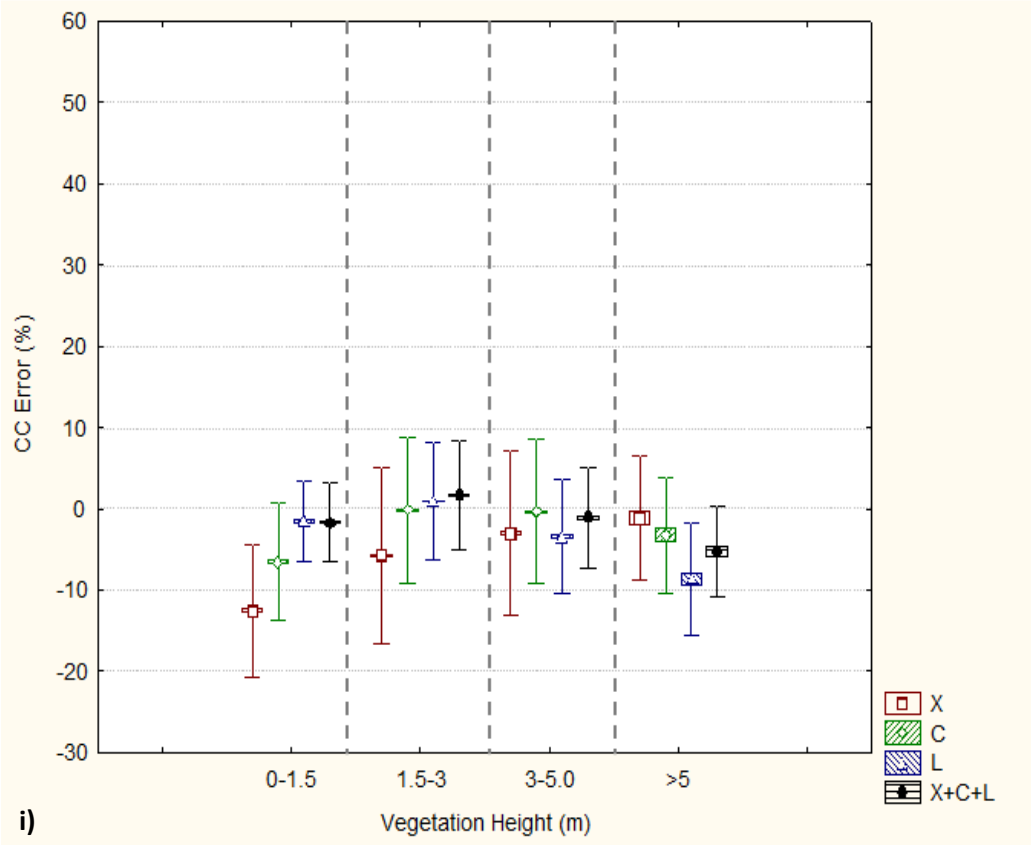


Figure 11

Click here to download Figure: Figure11.docx



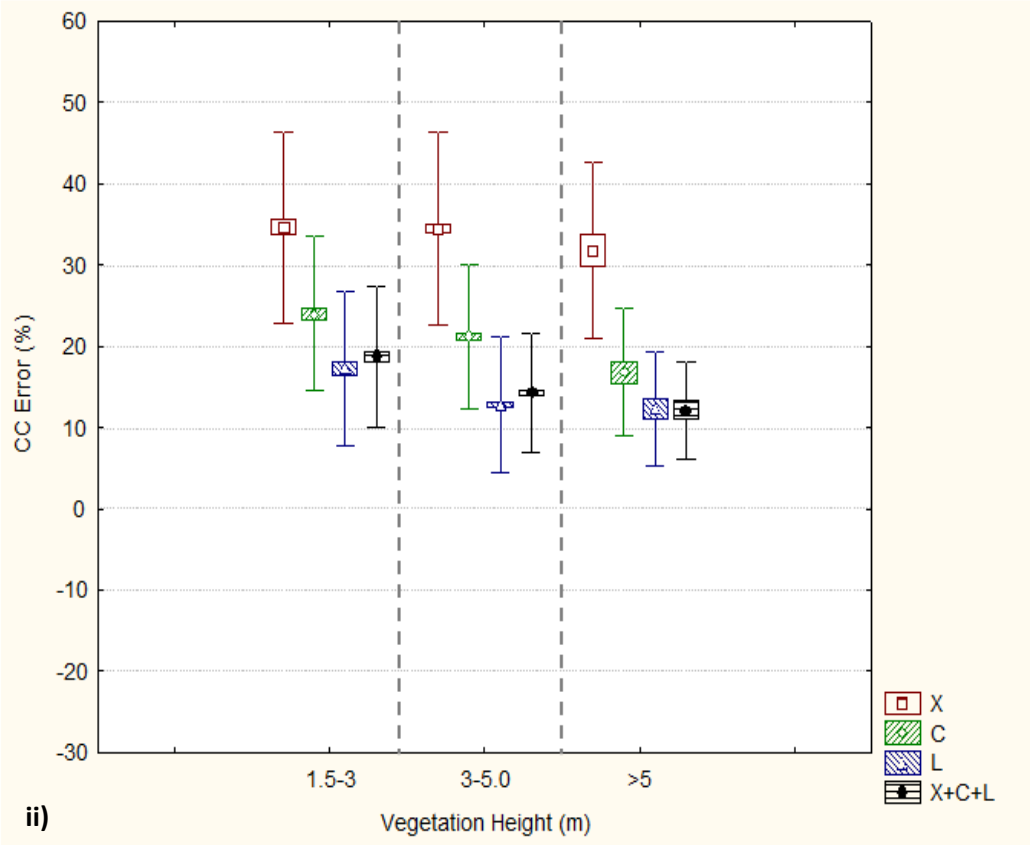


Table 1
Click here to download Table: Table1_New.docx

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

Imagery	Sensor	Mode	Incidence angle	Acquisition time	Season
1 2 3 4 5	TerraSAR-X X-band *ß*	StripMap Dual Polarized (HH & HV)	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	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 *σ*	Dual Polarized (HH & HV)	34.3°	14/08/2010 31/08/2010	Winter 2010
AGB (kg) Product CC (%) Product TCV Product	CAO LiDAR *Φ*	Discrete Footprint	Nadir	1/04/2012-24/05/2012	End summer 2012

ß: http://www.geoimage.com.au/satellite/TerraSar ; ¥: http://www.asc-csa.gc.ca/eng/satellites/radarsat/radarsat-tableau.asp ; σ: http://www.eorc.jaxa.jp/ALOS/en/about/palsar.htm ; *Φ*: Asner et al., (2012)

Table 2

Click here to download Table: Table2_New.docx

Table 2: Original, modified and final SAR pixel size changes during multi-looking and pre-processing steps

SAR Dataset	Original Pixel Size [m] (Range X Azimuth)	Multi-Looking factors (no. Looks for Range X Azimuth)	Modified Pixel Size [m] (after multi-looking)	Final Pixel Size [m] (map geometry)Φ
ALOS PALSAR FBD	9.37 X 3.23	2 X 8	18.74 X 25.84	12.5 X 12.5
RADARSAT-2 SLC	4.70 X 5.10	1 X 1	4.70 X 5.10	5 X 5
TerraSAR-X StripMap MGD	2.75 X 2.75	4 X 4	11 X 11	12.5 X 12.5

Φ Resolutions used in the modelling stage but all were resampled to 12.5m for mapping

Table 3

Click here to download Table: Table3.docx

Table 3: Woody Canopy Cover (CC), Total Canopy Volume (TCV) and Above Ground Biomass (AGB) parameter modelling accuracy assessment (validation) results obtained from the Random Forest algorithm according to seven SAR frequency scenarios

	CC (%)		*TCV (unitless per hectare)*		*AGB (tonnes per hectare)*	
SAR Frequency	**R^2**	**RMSE (SEP %)**	**R^2**	**RMSE (SEP %)**	**R^2**	**RMSE (SEP %)**
X-band only	0.34	18.12 (50.87)	0.35	35534.50 (33.79)	0.32	10.88 (59.82)
C-band only	0.61	13.20 (38.50)	0.66	24731.06 (24.07)	0.60	7.81 (43.66)
L-band only	0.77	10.59 (29.64)	0.79	19902.79 (18.88)	0.78	6.05 (32.90)
X+C-band	0.69	11.71 (33.94)	0.72	22243.64 (21.59)	0.67	7.19 (40.33)
X+L-band	0.80	9.90 (27.78)	0.82	18609.04 (17.70)	0.81	5.70 (31.35)
C+L-band	0.81	9.23 (26.94)	0.83	17236.50 (16.77)	0.81	5.45 (30.44)
X+C+L-band	0.83	8.76 (25.40)	0.85	16443.57 (15.96)	0.83	5.20 (29.18)

Datasets split into 35% Training and 65% Validation for modelling

Table 4

Click here to download Table: Table4_New.docx

Table 4: Total woody Canopy Cover (CC), Total Canopy Volume (TCV) and Above Ground Biomass (AGB) % error across the entire LiDAR-SAR coverage for the four main SAR frequency scenarios (Number of observations = 17559)

CC 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
TCV Error Classes	**X-band Error**	**C-band Error**	**L-band Error**	**X+C+L-band Error**
Major overestimation (<-50k)	7.54	1.69	0.40	0.35
Minor overestimation (-50k to -10k)	28.58	22.96	22.32	18.57
Negligible error (-10k to 10k)	**4.64**	**8.26**	**15.56**	**16.62**
Minor underestimation (10k to 50k)	32.41	58.43	57.12	60.31
Major underestimation (>50k)	26.82	8.66	4.60	4.14
AGB Error Classes	**X-band Error**	**C-band Error**	**L-band Error**	**X+C+L-band Error**
Major overestimation (<-15t/ha)	4.53	1.95	0.79	0.65
Minor overestimation (-15t/ha to -5t/ha)	27.46	18.85	15.47	13.16
Negligible error (-5t/ha to 5t/ha)	**13.29**	**22.05**	**36.42**	**36.05**
Minor underestimation (5t/ha to 15t/ha)	25.07	41.00	37.24	39.70
Major underestimation (>15t/ha)	29.65	16.15	10.08	10.43