



School of GeoSciences

Dissertation
for the degree of

**MSc in Earth Observation and
Geoinformation Management**

Oliver McFarlane

August 2024

Statement of Copyright and Originality

I declare that this dissertation represents my own work and that where the work of others has been used it has been duly accredited.

I further declare that the length of the components of this dissertation is 4,994 words for **Part I: Research Paper** and 2,886 words for **Part II: Technical Report**.

Copyright of this dissertation is retained by the author and the University of Edinburgh. Ideas contained in this dissertation remain the intellectual property of the author and their supervisors, except where explicitly otherwise referenced.

All rights reserved. The use of any part of this dissertation reproduced, transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise stored in any retrieval system without the prior written consent of the author and the University of Edinburgh (Institute of Geography) is not permitted.

I agree that this dissertation and associated electronic documents, web pages, data, files, and computer programs can be retained by the University of Edinburgh.

YES

I agree that, with the permission of my supervisor or the Programme Director, these materials will be made available for the purposes of preparing a publication.

YES

I agree that, with the permission of my supervisor or the Programme Director, these materials can be used within the University of Edinburgh for continued research or teaching.

YES



Oliver McFarlane

August 2024

Acknowledgements

I would like to express my profound gratitude to my supervisor, Prof Casey Ryan, for his patience, advice, and guidance throughout this dissertation. It has been an absolute pleasure to be a part of both LAND and SECO teams.

For simply making this dissertation possible, I would like to thank all contributors to the SEOSAW partnership. The work that goes into the organisation and maintenance of the plot network and its collaborators is truly impressive.

I also want to thank Dr Steve Hancock, for his insights and guidance over the course of both my undergraduate and postgraduate degrees, and for sharing his passion for all things GEDI.

Dr Sam Harrison has also been a fantastic Google Earth Engine guru and great all-around help with the conception of this dissertation.

Finally, a huge thank you to the many who read, listened, and fell asleep to the countless iterations of this dissertation.

Part I: Research Paper



Miombo Woodlands, Southern Angola (Huntley, 2023)

In-Situ Validation Reveals Poor Performance of Extrapolated GEDI Aboveground Biomass Estimates Across Miombo Landscapes

Abstract

The Global Ecosystem Dynamics Investigation (GEDI) collected a near-global sample of forest structural metrics between 2019 to 2023, which were used to estimate Aboveground Biomass Density (AGBD). GEDI AGBD estimates were derived from predictive models broadly stratified by plant functional type and continental region. However, the miombo region, encompassing 2.7 million km² of Southern Africa and characterised by heterogenous and complex vegetation, was critically under-represented in the calibration of GEDI AGBD estimates. Furthermore, sparse GEDI sampling prevents direct in-situ validation of GEDI AGBD estimates with field AGBD estimates. In lieu of airborne LiDAR data to simulate GEDI metrics, this study employed Random Forest machine learning to extrapolate GEDI AGBD estimates across two miombo landscapes and produce a time series from 2017 to 2023 to coincide with 48 field AGBD estimates from permanent sample plots. This study found moderate predictive relationships between GEDI AGBD estimates and EO predictor variables including phenology variables from Landsat and L-band backscatter from PALSAR-2, which yielded $R^2 = 0.29$, RMSE = 23.12 Mg/ha, Bias = +2.65 Mg/ha. However, spatial cross-validation revealed drastic differences between site-specific models, highlighting the importance of exercising caution to avoid misrepresenting the heterogeneity and complexity of mixed woodland-savanna ecosystems in large-scale models. Ultimately, in-situ validation of extrapolated GEDI AGBD estimates across both sites and census years yielded quasi-null predictive power with $R^2 = 0.09$, RMSE = 33.42 Mg/ha, and Bias = +19.75 Mg/ha. This study concludes that reliable GEDI AGBD estimates in the miombo region require locally calibrated models and highlights the critical importance of integrating in-situ field AGBD estimates across space and time for calibration and validation of large scale AGBD estimates.

Keywords: GEDI, Aboveground Biomass, In-Situ, Validation, Miombo, Woodland, Savanna.

1. Introduction

Carbon stocks and dynamics in African savanna ecosystems represent significant uncertainties within the terrestrial carbon cycle (Bouvet et al, 2018; Friedlingstein et al, 2023; Mitchard et al, 2013; Pötzschner et al, 2022; Tucker et al, 2023). The miombo region, spanning approximately 2.7 million km² of Southern Africa, is the most prominent example of complex mixed woodland-savanna ecosystems (Menaut et al, 1995; White, 1983). However, the ecological diversity and heterogeneity of the miombo region pose substantial challenges for estimating large-scale carbon stocks using Earth Observation (EO) instruments (Davies et al, 2023; Godlee et al, 2021; Hill et al, 2013; McNicol et al, 2018a; Rodríguez-Veiga et al, 2019; Ryan et al, 2013; 2017). As a result, the miombo region remains significantly underrepresented in large-scale carbon accounting and monitoring efforts (Demol et al, 2024; McNicol et al, 2018b; 2023; Ryan et al, 2016).

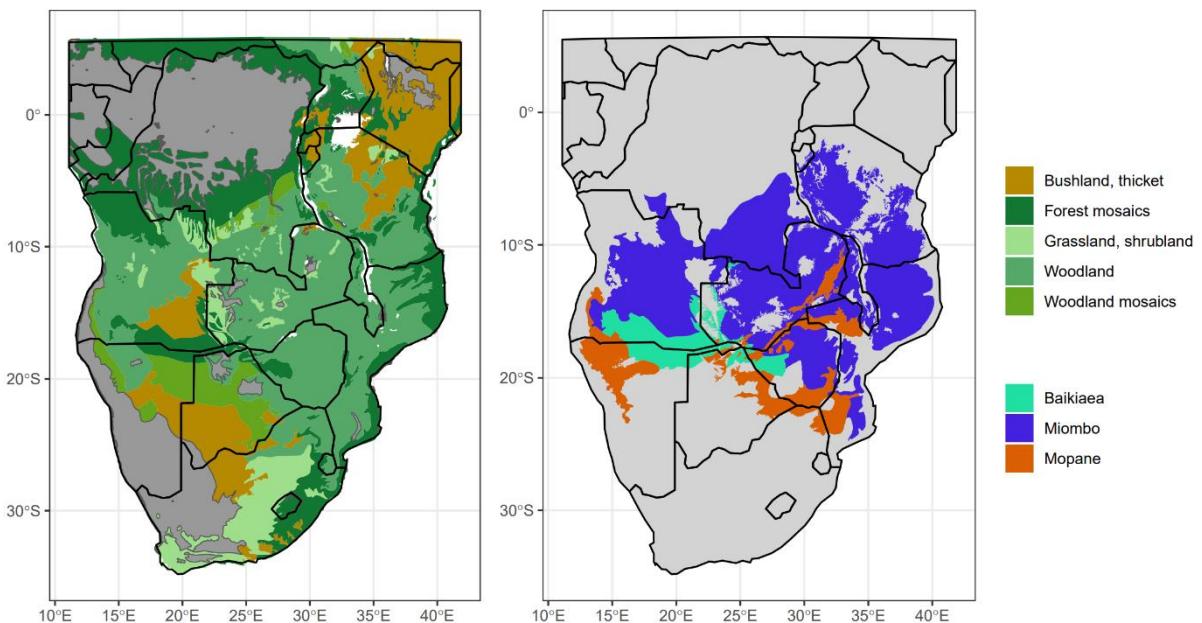


Figure 1.1: Distribution and Ecological Heterogeneity of Miombo Woodlands. Left: vegetation classification adapted from White (1983). Right: floristic classification adapted from Dinerstein et al. (2017). Source: Godlee (2021:41).

Aboveground Biomass Density (AGBD) is a highly valued Essential Climate Variable given it has a direct carbon conversion factor, and estimates can be obtained using EO instruments (FAO, 2009; IPCC, 2006:48). However, direct measurements of AGBD, the dry mass of live organic matter above soil, can only be obtained through the destructive harvesting of trees (Houghton, 2005). To avoid the expensive and laborious process, specific species or broad plant functional type (PFT) allometric relationships use other structural features, such

as stand height and diameter, to estimate AGB at larger scales (Chave et al, 2005; 2014; Clark & Kellner, 2012; Friedl et al, 2010; Mugasha et al, 2013; Poulter et al, 2011; Ryan et al, 2011; Ustin & Gamon, 2010).

In-situ field measurements of vegetation structure are resource intensive and practically demanding, inevitably resulting in sparse and sporadic sampling. Obtaining precise field measurements of stand or canopy height in more dense, heterogenous woodlands can be particularly challenging, leading to the favouring of stand diameter or another of the 400 other sought-after variables in National Forest Inventories (NFIs) (Gibbs et al, 2007; Goetz et al, 2015; McRoberts & Tomppo, 2007). However, the absence of height measurements can significantly increase the bias of AGB estimates derived from allometric relationships (Chave et al, 2014).

The potential of Light Detection and Ranging (LiDAR) EO instruments to provide precise vertical and structural measurements of vegetation, as well as allometry-derived AGBD estimates, has been demonstrated throughout the tropics (Avitabile et al, 2016; Baccini et al, 2012; 2017; Coops et al, 2021; Feldpausch et al, 2012; Harris et al, 2021; Saatchi et al, 2011; Zolkos et al, 2013). However, LiDAR AGBD estimates have only recently been evaluated at landscape scale in the miombo region (Li et al, *preprint*; Liang et al, 2023; Demol et al, 2024). Liang et al (2023) extrapolated spaceborne LiDAR-derived AGBD estimates across 9,107 km² in Mabalane, southern Mozambique, and predicted AGBD change over 10 years from 2008-2018. Demol et al (2024) used terrestrial and airborne LiDAR instruments to extrapolate AGB estimates across 500 km² of northern Mozambique. Li et al (*preprint*) used airborne and spaceborne LiDAR to extrapolate AGBD estimates across 833 km² of South Africa and used field AGB estimates to assess performance.

The Global Ecosystem Dynamics Investigation (GEDI), initially in operation from March 2019 to March 2023, collected a near-global sample of more than 26 billion precise footprint measurements of vegetation structure including tree stand (canopy) heights (Dubayah et al, 2020). The novel LiDAR instrument transmits pulses of laser energy towards the Earth's surface which interact with the structural components of vegetation, such as leaves and branches, before the ground topography. The returned energy is expressed in a waveform to represent proportions of returned intensity at relative (canopy) heights (RH) above the ground, see Figure 1.2 (Hofton & Blair, 2019). Collocated GEDI canopy heights (simulated with Airborne Laser Scanning data) and field AGBD estimates (derived from local allometric equations) were used to develop linear models stratified by 32 combinations of PFT and geographic world region (Duncanson et al, 2022; Friedl et al, 2010; Hancock et al, 2019; Kellner et al, 2023; Ustin & Gamon, 2010). The modelled L4A product contains a near-global sample of 25 m footprint AGB estimates, including generalised AGB estimates for the

dominant Deciduous Broadleaf Trees (DBT) and Grasses, Shrubs, and Woodlands (GSW) in Africa.

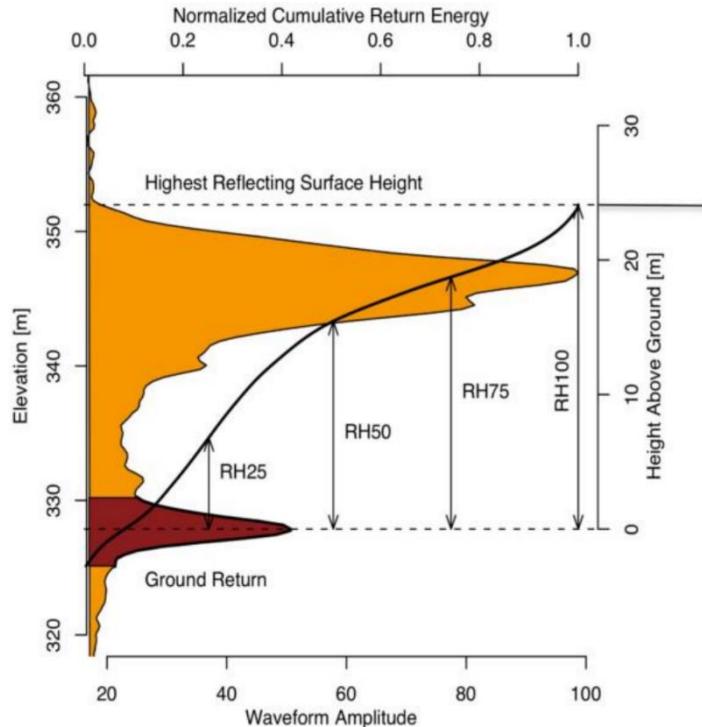


Figure 1.2: GEDI Waveform Return and Relative Height Metrics. Ground return is identified as the lowest identified mode (Hofton & Blair, 2019). Relative Height (RH) above ground is represented by percentiles of cumulative energy received. GEDI waveforms correspond to 25 m footprints Source: Dubayah et al. (2020).

However, there are plausible concerns that the generalised models developed for DBT and GSW in Africa, and their application to the miombo region, may be biased due to spatial dependency in their calibration (Ploton et al, 2020). The African DBT and GSW models were developed with 490 and 6 field plots respectively, in sites covering 0.36 sq. km of eastern Tanzania, see Figure 1.3. In the development of the African DBT model, the strongest predictors (RH metrics) of AGBD were RH98 and RH50, representing the 98th and 50th percentiles of cumulative return energy, which yielded $R^2 = 0.63$ and RMSE% = 57.24. However, due to the lack of calibration data (field AGBD estimates and airborne LiDAR) specific to African GSW, the development of a dedicated model was not possible. A global GSW model was used in its place, which relied on RH98 as the only predictor, although yielded $R^2 = 0.86$ and RMSE% = 55.39.

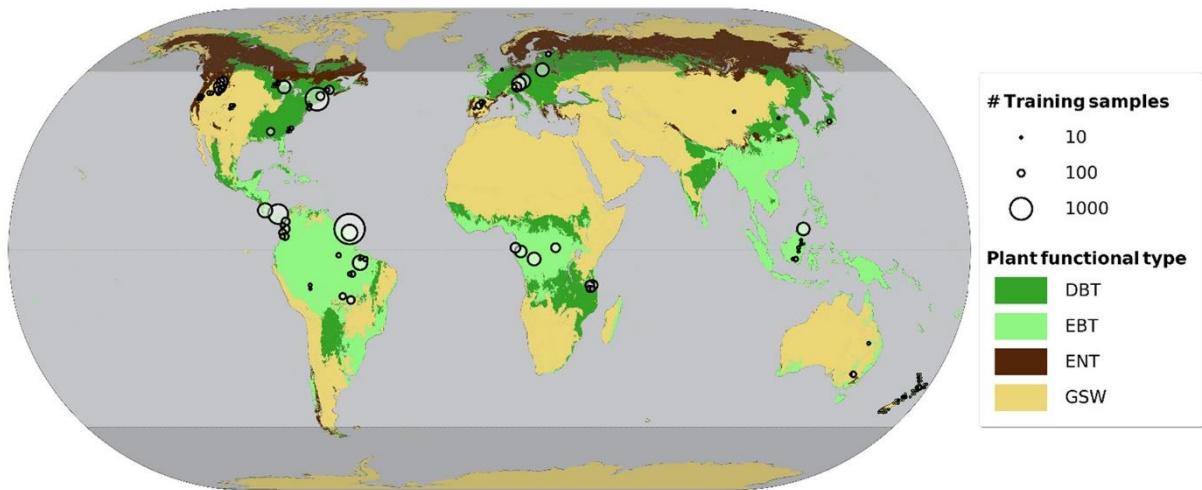


Figure 1.3: Plant Functional Types and Calibration Samples for GEDI AGB Estimates. Distribution of Deciduous Broadleaf Trees (DBT), Evergreen Broadleaf Trees (EBT), Evergreen Needleleaf Trees (ENT), and Grasses, Shrubs, and Woodlands (GSW) were derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover product (Friedl et al, 2010; Sulla-Menashe & Friedl, 2018). Models for GEDI AGBD were further stratified by geographic region: Africa, Australia-Oceania, Europe, North America, South (and Central) America, North Asia, and South Asia. Source: Duncanson et al. (2022).

Despite initial indications of reasonable predictive power for both African DBT and global GSW models, the heterogeneity and complexity of the miombo region is likely to hinder the reliability of GEDI AGBD estimates. For instance, Cushman et al. (2023) found significant sensitivities with up to 20% underestimation in RH metrics \geq RH50 and AGBD estimates during leaf defoliation European DBT. Furthermore, Li et al (2023) found strong sensitivities in RH98 metrics to leaf phenology across nine sites in South Africa, with worse performance in leaf off conditions ($R^2 = 0.43$, RMSE% = 40.90) compared with leaf on conditions ($R^2 = 0.61$, RMSE% = 29.80) attributed to denuded branches disrupting return signals. Seasonal artefacts in GEDI data have been partially controlled with the removal of footprints in leaf off conditions and/or delineating informed temporal constraints (Kellner et al, 2023; Li et al, *preprint*).

GEDI AGBD estimates are inherently sparse and their utility for landscape scale AGBD estimation depends on the reliability of their extrapolation across space and over time. However, the presence and magnitude of seasonal artefacts significantly impacts the sample size of GEDI footprints available for extrapolations. Furthermore, the propagation of errors through the process of extrapolating GEDI AGBD estimates (itself a modelled product), can undermine the reliability of landscape scale estimates of AGB stock and change (Ploton et al, 2020). To date, Li et al (*preprint*) provide the only assessment of GEDI AGBD estimates developed with the generalised global GSW model in relation to field AGB estimates derived from local allometric equations, which yielded $R^2 = 0.42$ and RMSE% = 79.50. In comparison,

Li et al. found alternative GEDI AGBD estimates derived from local allometric equations and GEDI RH metrics yielded an improved $R^2 = 0.71$ and RMSE% = 53.30.

These results present a preliminary indication that GEDI AGBD estimates are more biased across miombo landscapes than initial African DBT and GSW model performances suggest. Furthermore, due to the complex phenological conditions across the miombo region, using space-for-time substitution (assuming predictive power is consistent over time as it is across space) may lead to inaccurate predictions of AGBD stock and changes based on GEDI AGBD estimates (Liang et al, 2023). This study found no examples where in-situ field AGBD estimates were available over multiple years and used to validate extrapolations over both space and time in the miombo region.

1.1. Research Aims and Questions

The overall aim of this study was to assess the predictive performance of extrapolated GEDI AGBD estimates across space and over time using in-situ field AGBD estimates for validation in the miombo region of Southern Africa, and address three research questions, see Table 1.1.

Table 1.1: Research Questions.

Notation	Research Question
RQ1	How consistent is predictive performance when different EO predictor variables are used to extrapolate GEDI AGBD estimates across both sites?
RQ2	How consistent is predictive performance of extrapolated GEDI AGBD estimates with spatial (site) cross validation?
RQ3	How consistent is predictive performance of extrapolated GEDI AGBD estimates when validated with in-situ field AGBD estimates across both sites over time?

2. Methods

2.1. Study Area

Two study sites were selected for this study, see Figure 2.1. Kilwa (TKW), within the Kilwa district in southern Tanzania, covers 17,194 km² of savanna and woodlands where the dominant genera are *Brachystegia* and *Jubbernardia*, both endemic to the miombo region (McNicol et al, 2018a). Gorongosa (MGR), encompassing the Gorongosa National Park in central Mozambique, covers 12,313 km² with dominant *Brachystegia* woodlands alongside various *Acacia* and *Combretum* (Ryan, 2009). The vegetation in both sites is predominantly deciduous, with defoliation the dry season (May–October) and leaf presentation during the wet season (November–April) (Menaut et al, 1995; Ryan et al, 2013; 2017).

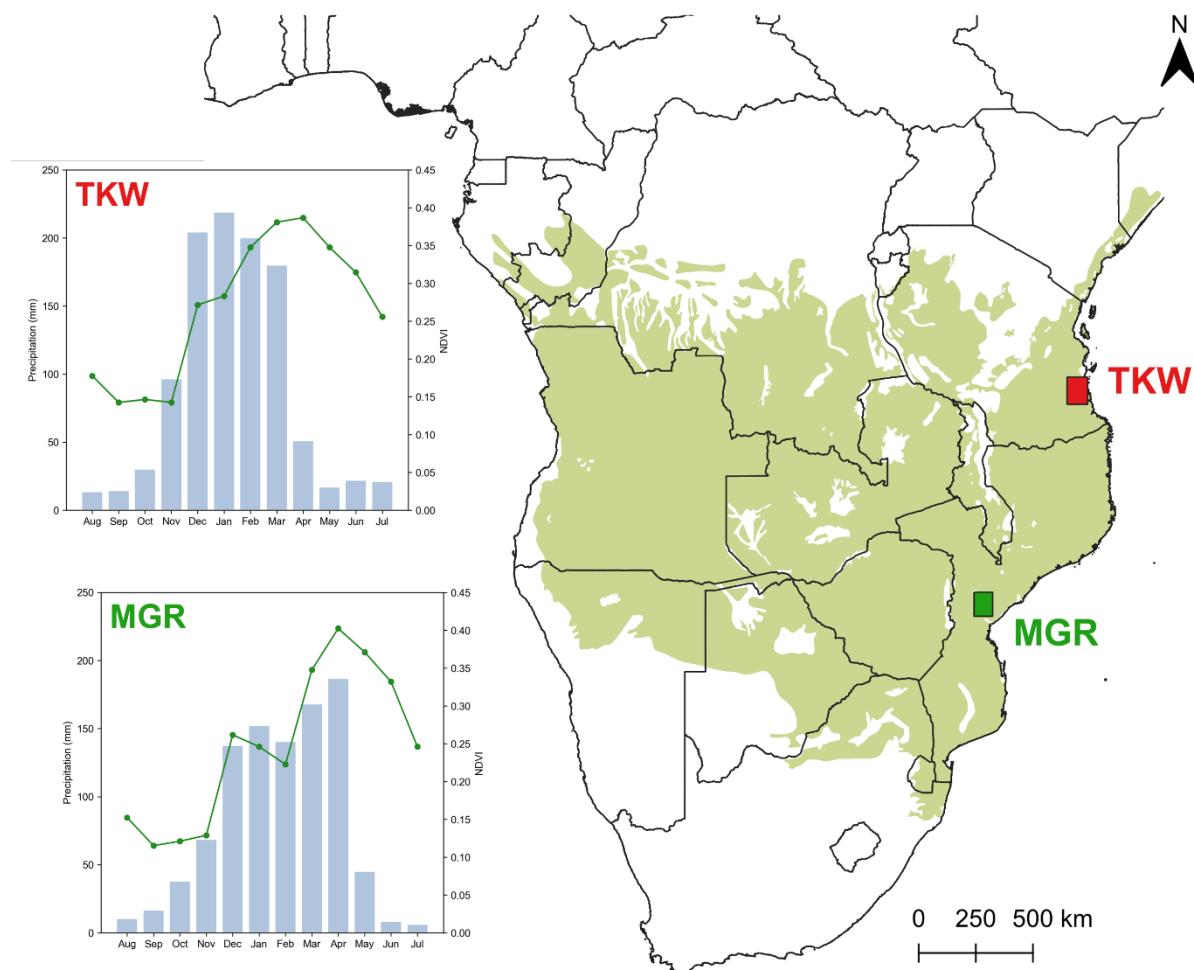


Figure 2.1: Study Sites. TKW in Kilwa, Tanzania and MGR in Gorongosa, Mozambique with the shaded miombo region (White, 1983). Aggregated ERA-5 Land precipitation (Muñoz-Sabater et al, 2021) and Normalised Difference Vegetation Index (NDVI) (Xue & Su, 2017) derived from Landsat-8 (Roy et al, 2014) shown for both sites, despite significant inter-annual variability (2015–2023). Authors own figure.

In both sites, considerable interannual variations in average annual rainfall (600-1300 mm) and sensitive fire regimes have strong influences on plant phenology and AGBD dynamics (Davies et al, 2023; Ryan, 2009; Ryan & Williams, 2011; Ryan et al, 2013; 2017; Wigley et al, 2024). Both sites include available in-situ field plots and are large enough to provide meaningful AGBD extrapolations using EO-based approaches.

2.2. Field AGB Estimates

This study utilises field AGBD estimates from the Socio-Ecological Observatory for Studying African Woodlands (SEOSAW) network. SEOSAW connects over 10,000 field plots and more than 100 researchers across 13 countries, with the goal of providing long term, in-situ ecological measurements across the under-represented miombo region (SEOSAW, 2021). Census observations of SEOSAW Permanent Sample Plots (PSPs) are intermittent. In TKW, 25 PSPs had been censused 2-5 times between 2010-2021 while in MGR, 15 PSPs had been censused 3-7 times from 2006-2021. All PSPs utilised in this study were all 1 ha (100 m²) to suitably match pixel resolution of EO products. In each census, plot-level AGB was estimated from direct measurements of at least 200 stems (diameter > 5 cm) and local allometric equations (Ryan et al, 2011; SEOSAW et al, 2021). The number of stem measurements made across all plot censuses to date is 38,261 in TKW and 28,053 in MGR, totalling 66,314. This study constrained the temporal range of field AGBD estimates, retaining two common censuses years (2017, 2021) in TKW and MGR. Combined, 48 field AGBD estimates were available across both sites (TKW = 34, MGR = 14).

2.3. GEDI AGB Estimates

The GEDI LiDAR instrument provides precise vegetation structural metrics in 25 m footprints across eight beam tracks (Dubayah et al, 2020). This study utilised the L2A Canopy Height and the L4A Footprint Aboveground Biomass Density products available through Google Earth Engine, from March 2019 to March 2023. GEDI data was segmented by site (TKW, MGR) and by austral year (August-July, denoted as year-end). In both TKW and MGR, GEDI AGBD estimates were predominantly derived from the African DBT model and the global GSW model. The model was assessed using collocated GEDI waveforms (simulated with airborne LiDAR) and field AGB estimates, themselves derived from local allometric equations (Colgan et al, 2013; Ene et al, 2017; Hancock et al, 2021; Mugasha et al, 2013). For both models, RH98 (DBT, GSW) and RH50 (DBT) were the strongest predictors of AGBD estimates. Useful GEDI footprints were selected based on several filtering conditions to ensure waveform fidelity and data quality, and minimise geolocation error and seasonal artefacts, see Table 2.1. Over the four years from August 2019 to (March) July 2023, 7,906 (TKW = 4,181; MGR = 3,615) high quality GEDI footprint AGBD estimates were retained.

Table 2.1: GEDI Footprint Filtering Conditions.

Variable	Description	Filter	References
L2A RH98	Relative Height (RH) above ground representing 98 th percentile of cumulative energy received.	> 3 m	Li et al. (2023)
L4A AGBD	Footprint AGBD estimates derived from RH metrics, field AGBD estimates, and allometric equations.	< 300 Mg/ha	SEOSAW (2021)
Beam Sensitivity	Signal to noise ratio representing the proportion of canopy cover through which GEDI waveform penetrates.	≥ 0.98	Leite et al. (2022) Li et al. (2023)
Quality Flags	Composite flag raised to signify GEDI waveform fidelity, and L2/L4 product validity for measuring surface and forest structure.	== 1	Hofton & Blair (2019) Kellner et al. (2023)
Degradate Flag	Composite flag indicating poor geolocation (positioning and/or pointing) performance.	== 0	Dubayah et al. (2021)
Geolocation Error	Absolute difference between GEDI L2B Canopy Cover Fraction and Landsat median NDVI, relative to mean difference.	$< 1 \sigma$	Liang et al. (2023) Roy et al. (2021)
Seasonal Artefacts	Fluctuations in GEDI waveform return reliability during phenological transitions and periods of defoliation.	November to April (minimised)	Menaut et al. (1995) Li et al. (2023) Li et al. (<i>preprint</i>) Part II: Technical Report

2.4. Predictor Variables

GEDI footprints are sparse and do not coincide with the in-situ field plots used in this study. To extrapolate GEDI AGBD estimates across space and time to align with field AGBD estimates, 54 spatially contiguous predictor variables were retrieved from four EO sources through Google Earth Engine. Topographic features from the Shuttle Radar Topography

Mission (SRTM) X-band radar instrument (Van Zyl, 2001); surface reflectance features from the Landsat Operational Land Imager (OLI) instrument (Roy et al, 2014); phenological features from Landsat-derived Normalised Difference Vegetation Index (NDVI) (Xue & Su, 2017); surface backscatter from the Sentinel-1 (hereafter Sentinel) C-band Synthetic Aperture Radar (SAR) instrument (Torres et al, 2012); and surface backscatter from the second Advanced Land Observing Satellite (ALOS-2) Phased Array L-band Synthetic Aperture Radar (PALSAR-2, hereafter Palsar) instrument (Rosenqvist et al, 2014). Topographic variables are based on the single year of SRTM observations in 2000 while all other variables were available to represent each austral year from 2017 to 2023.

2.5. Model Development

In the preliminary stage of model development, predictor variables with a paired absolute correlation coefficient greater than 0.9 were omitted, reducing the number of variables to 36 from the initial 54. Subsequently, various Random Forest (RF) machine learning algorithms (models) were developed to extrapolate GEDI AGB estimates over space and time using the predictor variables. RF models consolidate multiple decision tree outcomes into a single predictive relationship within large datasets (Belgiu & Drägut, 2016; Breiman, 2001). RF models are effective and commonly employed in large scale ecological modelling, including in efforts to extrapolate sparse GEDI data across space and time (Coops et al, 2021; Kacic et al, 2021; Li et al, *preprint*; Liang et al, 2023; Potapov et al, 2021; Simard et al, 2011).

To address RQ1, four RF models were developed with data from both sites (TKW and MGR), using a 10-fold random cross-validation approach where the available calibration data underwent 10 different, random splits of 70% for training and 30% for testing to produce 10 model iterations. All four models included SRTM variables but those otherwise sourced from either Landsat (RF_L), Sentinel (RF_S), Palsar (RF_P), or All (RF_A) predictor variables were grouped accordingly. Initial AGBD predictions from each RF model were tested with the random 30% withheld subset of available training data. RF model tests yielded three main performance statistics: coefficient of determination or variance explained (R^2), root mean square error (RMSE), and the absolute difference between predictions and ‘observations’ (Bias). All RF models were configured with 200 estimators or ‘trees’ after testing RF_A performance with 100-500 trees (in steps of 100) and finding a marginal improvement in performance.

To address RQ2, four RF models were developed with data segmented by site. Two RF models were developed using the same 10-fold random cross validation approach and a 70:30 split, focusing separately on TKW (RF_{TKW}) and MGR (RF_{MGR}). The other two RF models were developed with a spatial cross validation approach where all available calibration data from one site was used as training data and all available calibration data from the other site was

used as testing data: TKW-MGR ($RF_{TKW-MGR}$) and MGR-TKW ($RF_{MGR-TKW}$). A total of 62 RF models were developed.

2.6. Validation of GEDI AGB Estimates

To address RQ3, the best statistically performing RF model (from either RF_L , RF_S , RF_P , or RF_A) was used to extrapolate GEDI AGBD estimates across both sites and over seven years (2017-2023) encompassing the GEDI acquisition period (2019-2023) and in-situ field AGBD estimates from both census years (2017, 2021). Extrapolated GEDI AGBD estimates were then validated with the in-situ field AGBD estimates aggregated to each of the SEOSAW PSPs. Spatial matching of each field AGBD estimate to its coincident, extrapolated GEDI AGBD estimate was done by aligning an extrapolated GEDI AGBD estimate with the centroid of each PSP. The R^2 , RMSE, and Bias metrics of coincident, extrapolated GEDI AGBD estimates and in-situ field AGBD estimates was then used to statistically measure the overall reliability of GEDI AGBD estimates across space (both sites) and time (both census years).

3. Results

3.1. Effect of Predictor Variables on Model Performance

To address RQ1, four RF models were developed using the filtered sample of GEDI AGBD estimates and the 36 predictor variables grouped by EO source. Tested with a withheld subset of available calibration data, each of the 10 model iterations yielded R^2 , RMSE, and Bias statistics, see Table 3.1. By measure of R^2 , the best performing grouping of predictor variables were those derived from Landsat ($R^2 = 0.27$), which included surface reflectance and phenology metrics, and Palsar backscatter metrics ($R^2 = 0.27$). Sentinel backscatter metrics performed comparatively poorer ($R^2 = 0.21$). Model performance by RMSE shows a contrasting pattern where RF_P (23.82 Mg/ha) improves over RF_L (25.01 Mg/ha) despite similar R^2 metrics.

Notably, performance by bias and RMSE is inversely related to that by R^2 . This result indicates a trade-off between the model's ability to capture variance and the Bias in its predictions relative to the observed GEDI AGBD estimates. Here, the simpler model with fewer predictor variables, RF_S , is more stable in terms of Bias but has a lower R^2 . This suggests that RF_S is underfitting the data because it fails to capture the complexity of the underlying patterns in the GEDI AGBD estimates. Consequently, while it shows lower Bias, it does not explain much of the variance, leading to a comparatively lower R^2 value.

The range of R^2 performances across model iterations with a 10-fold random cross validation approach shows that the differences between model groups are slightly more apparent, see Figure 3.1. The smallest grouping of R^2 values across models was RF_L (0.06), followed by RF_P (0.14) and RF_S (0.16). Furthermore, the median R^2 values highlight a more apparent distinction between grouped models, with RF_L (0.25) showing more separation from RF_P (0.18) and RF_S (0.15).

The best performance of any model grouping, albeit marginal, was RF_A with $R^2 = 0.29$, RMSE = 23.12 Mg/ha, Bias = 2.65 Mg/ha. The inverse relationship between better statistical measures of variance explained and comparatively worse bias to other groups, suggests that RF_A is likely to suffer from overfitting. In addition, Figure 3.2 shows that RF_A overestimates lower GEDI AGBD estimates (0-50 Mg/ha) systematically and underestimates higher GEDI AGBD estimates (> 50 Mg/ha). However, the range of R^2 values is small (0.06) and the median R^2 (0.26) remains higher than other groups. The RF_A model iteration with the highest R^2 was retained for extrapolation of GEDI AGBD estimates over time and, ultimately, validation with in-situ field AGBD estimates.

Table 3.1: Best Performing Models by EO Source.

Performance	RF _A	RF _L	RF _P	RF _S
R ²	0.29	0.27	0.27	0.21
Bias (Mg/ha)	+ 2.65	+ 1.27	+ 1.63	+ 0.99
RMSE (Mg/ha)	23.12	25.01	23.82	24.50
RMSE%	55.11	58.55	56.90	56.75

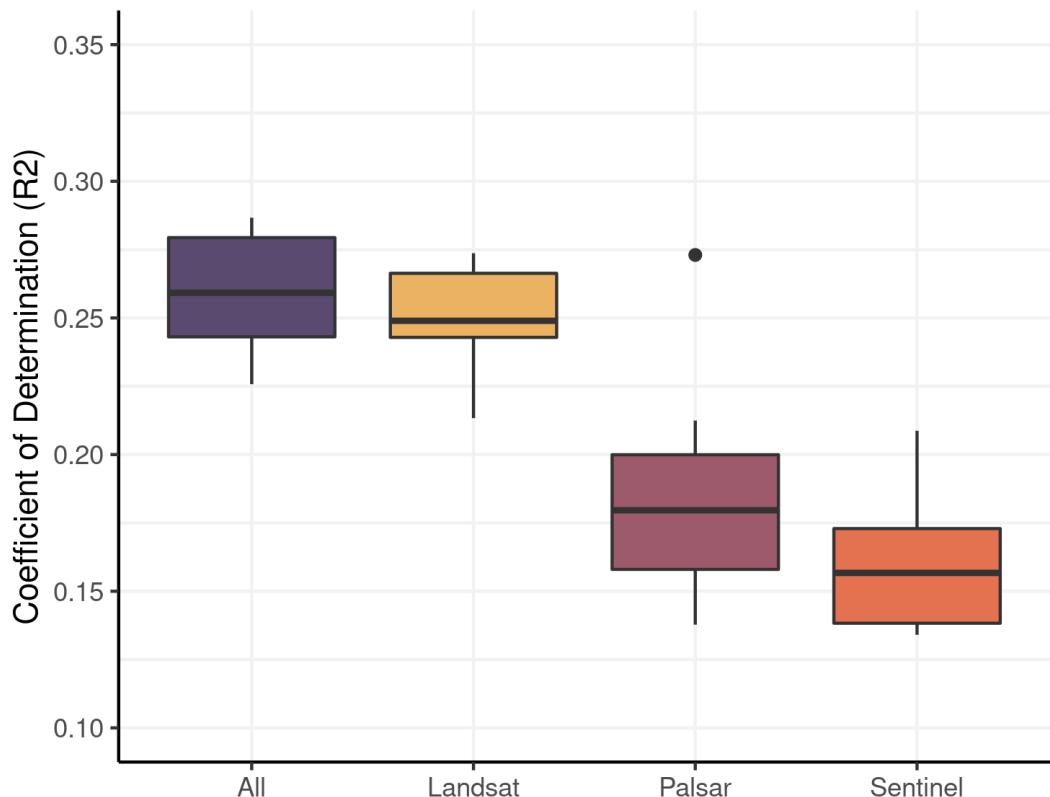


Figure 3.1: Range of Model Performances by EO Source. Boxplot showing performance of 10-fold random cross validation of model groups. Coefficient of determination used as measure of model performance (R²). Landsat models show greater differentiation to Palsar models despite similar best performing iterations. Authors own figure.

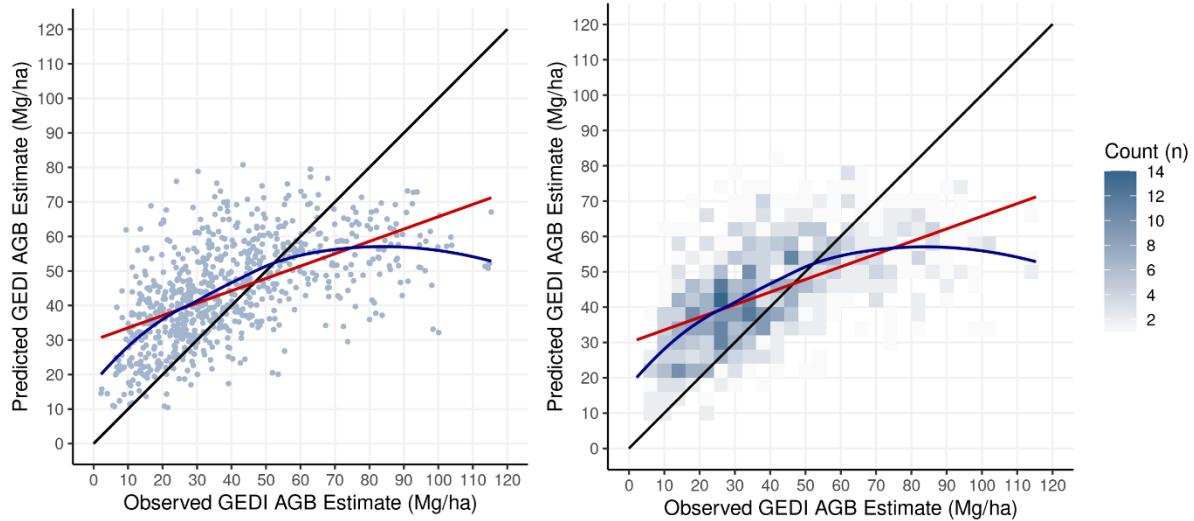


Figure 3.2: Best (RF_A) Model Prediction Compared to a Withheld Subset of GEDI AGBD Estimates from Both Sites. Left: Scatter plot of “observed” (withheld subset of GEDI AGBD estimates) and RF_A predicted GEDI AGBD estimates. Right: two-dimensional histogram of the scatter plot shown on left. One-to-one (black), linear (red), and non-linear (blue) lines of best fit shown for both plots, yielding $R^2 = 0.29$. Authors own figure.

3.2. Spatial Transferability of Model Performance

To address RQ2, four additional RF models were developed with all 36 predictor variables to assess the predictive performance of extrapolated GEDI AGBD estimates across space (between sites). The first two were tested on a random 30% subset of withheld data from the specific site, while the latter two were tested on all the available data from the other site to which the model was trained, see Table 3.2. RF_{TKW} and RF_{MGR} , trained and tested with a withheld subset from each specific site, show notable differences in performance. RF_{TKW} shows an improved R^2 (0.33) and RMSE (22.88 Mg/ha) to the best models grouped by EO source. Furthermore, RF_{MGR} performs even better in terms of R^2 (0.38) and RMSE (20.96 Mg/ha), exceeding that of any other model group or iteration.

However, while site-specific models perform comparatively well, spatial cross-validation yields dramatically poorer model performances with quasi-null predictive power, see Table 3.2. The most explicit case is $RF_{MGR-TKW}$ which demonstrates the lowest R^2 (0.07) and the highest Bias (14.64 Mg/ha) of any model. By comparison, $RF_{TKW-MGR}$ shows a low R^2 (0.12) but an improved and comparable Bias (-0.59 Mg/ha) to the other models and only slightly elevated RMSE (28.64 Mg/ha). The results from $RF_{MGR-TKW}$ indicate model underfitting, where the model fails to recognize and capture the complexities and patterns of the GEDI AGBD estimates in TKW. As a result, it performs poorly when applied to TKW data. Conversely, the $RF_{TKW-MGR}$ model, although it has slightly better R^2 and bias, still performs poorly overall. This suggests

that the $\text{RF}_{\text{TKW-MGR}}$ model also fails to accurately identify important patterns in the data, leading to its poor performance.

Moreover, the weighting of predictor variables in model development shows a strong distinction between sites, see Table 3.3. When trained with data within TKW, the model found stronger predictive relationships with Landsat metrics and topographic variables, while the model trained in MGR found a considerably greater role of Palsar metrics in the predictive relationship of GEDI AGBD estimates. The most dramatic difference is the Annual Median HV backscatter metric which was weighted 22.80% in $\text{RF}_{\text{MGR-TKW}}$ and only 6.19% in $\text{RF}_{\text{TKW-MGR}}$. The difference in variable importance indicates the inconsistencies between each site in terms of the vegetation, ecological features, and topographic profile, and highlighting the heterogeneity between miombo landscapes beyond that of the broader region.

Table 3.2: Model Performances with Spatial Cross Validation.

Performance	RF_{TKW}	$\text{RF}_{\text{TKW-MGR}}$	RF_{MGR}	$\text{RF}_{\text{MGR-TKW}}$
R^2	0.33	0.12	0.38	0.07
Bias (Mg/ha)	+ 1.45	- 0.59	+ 2.65	+ 14.64
RMSE (Mg/ha)	22.88	28.64	20.96	33.93
RMSE%	56.83	63.71	48.01	82.12

Table 3.3: Variable Importances with Spatial Cross Validation.

Importance	$\text{RF}_{\text{TKW-MGR}} (\%)$	$\text{RF}_{\text{MGR-TKW}} (\%)$
1	NDVI Median	11.51
2	Elevation	9.29
3	HV Median	6.19
4	Slope	4.26
5	VV Perc. 95 th	3.09
		HV Median
		22.80
		NDVI Median
		4.26
		Elevation
		3.71
		Band 5 Perc. 5 th
		3.59
		HH Perc. 5 th
		3.20

3.3. Model Validation with In-Situ Field AGB Estimates

To address RQ3, the best performing RF model (RF_A , $R^2 = 0.30$) was used to extrapolate GEDI AGBD estimates covering 4.87 km² (TKW = 2.61 km²; MGR = 2.26 km²) across 29,507 km² (TKW = 17,194 km²; MGR = 12,313 km²) over seven years (2017-2023). Utilising the predictor variables available for model calibration throughout the study period, a time series

of extrapolated GEDI AGBD estimates was generated for each site from 2017 to 2023, see Figure 3.3. The overall change in TKW was +0.56 Mg/ha from 2017 to 2023 with year-to-year changes ranging from -2.90 Mg/ha to +4.64 Mg/ha. In MGR, the overall change was -0.45 Mg/ha with year-to-year changes ranging from -4.66 Mg/ha to +5.81 Mg/ha. However, for both sites, no year-to-year changes or overall change exceeded RF_A RMSE (23.12 Mg/ha). The spatially contiguous extrapolation of GEDI AGBD estimates across both sites and over the seven years enabled the collocation of GEDI AGBD estimates with field (plot) AGBD census estimates in 2017 and 2021.

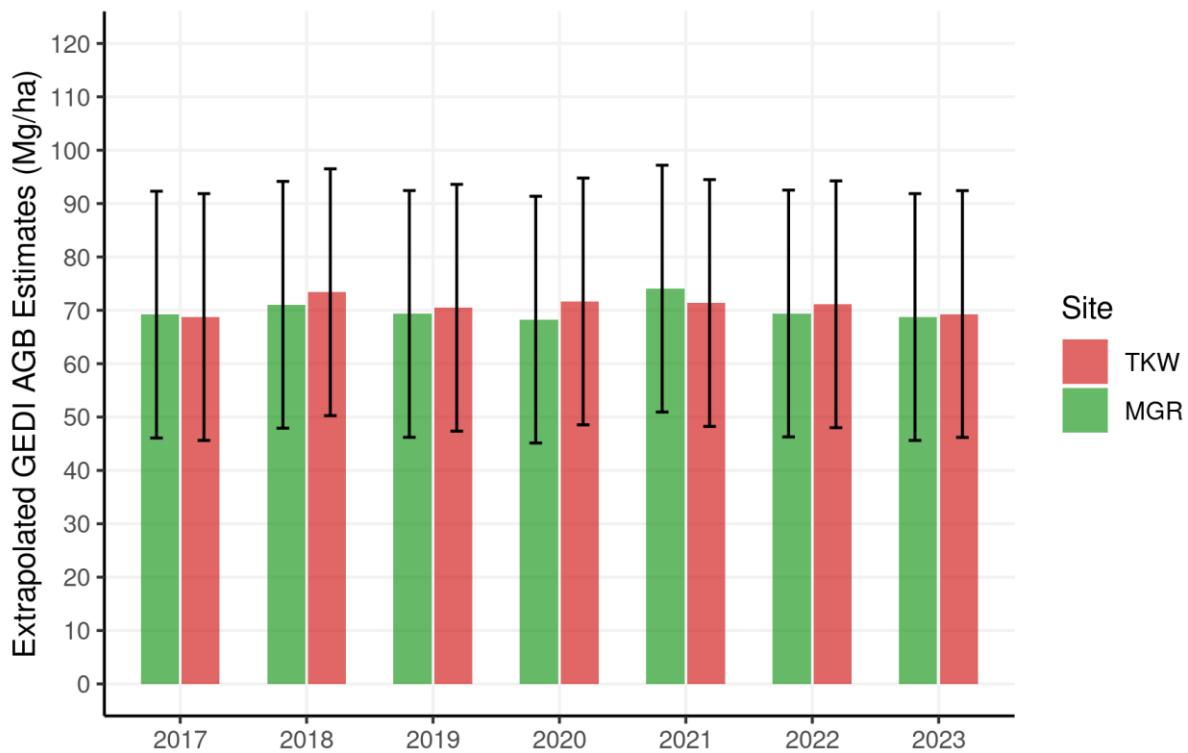


Figure 3.3: Extrapolated GEDI AGBD Estimates Time Series. Mean AGBD over entire site for each year with error bars denoting predictive model (RF_A) RMSE (23.12 Mg/ha), calculated by a withheld subset of GEDI AGBD estimates. Authors own figure.

The overall performance of extrapolated GEDI AGBD estimates was very poor with quasi-null predictive power, see Table 3.3 and Figure 3.4. The predictive power across sites was also very poor by measure of R^2 in TKW = 0.13 and in the case of the MGR, completely null with $R^2 = 0.00$. In every tested case, extrapolated GEDI AGBD estimates significantly overestimated AGBD with Bias ranging from +11.76 Mg/ha (MGR, 2017) to +23.17 Mg/ha (TKW, 2021). There were some improvements in performance by year and site, however, their significance should be weighted in context of the very small sample sizes. For example, MGR yielded $R^2 = 0.34$ in 2017 and $R^2 = -0.74$ in 2021, but only seven field AGBD estimates were used for validation. Similarly, TKW yielded $R^2 = 0.10$ in 2017 and $R^2 = 0.19$ in 2021 with 17

field AGBD estimates for validation. For all 48 available validation points, the qualifying relationship between extrapolated GEDI AGBD estimates and in-situ field AGBD estimates is reduced almost entirely to show very little predictive sensitivity, yielding $R^2 = 0.09$, RMSE = 33.42 Mg/ha, Bias = 19.75 Mg/ha.

Table 3.4: Performance of Extrapolated GEDI AGBD Estimates with In-Situ Validation.

Validation	TKW	MGR	Both
R^2	0.13	0.00	0.09
Bias (Mg/ha)	+ 22.09	+ 14.07	+ 19.75
RMSE (Mg/ha)	36.07	25.88	33.42
RMSE%	69.35	43.26	61.56

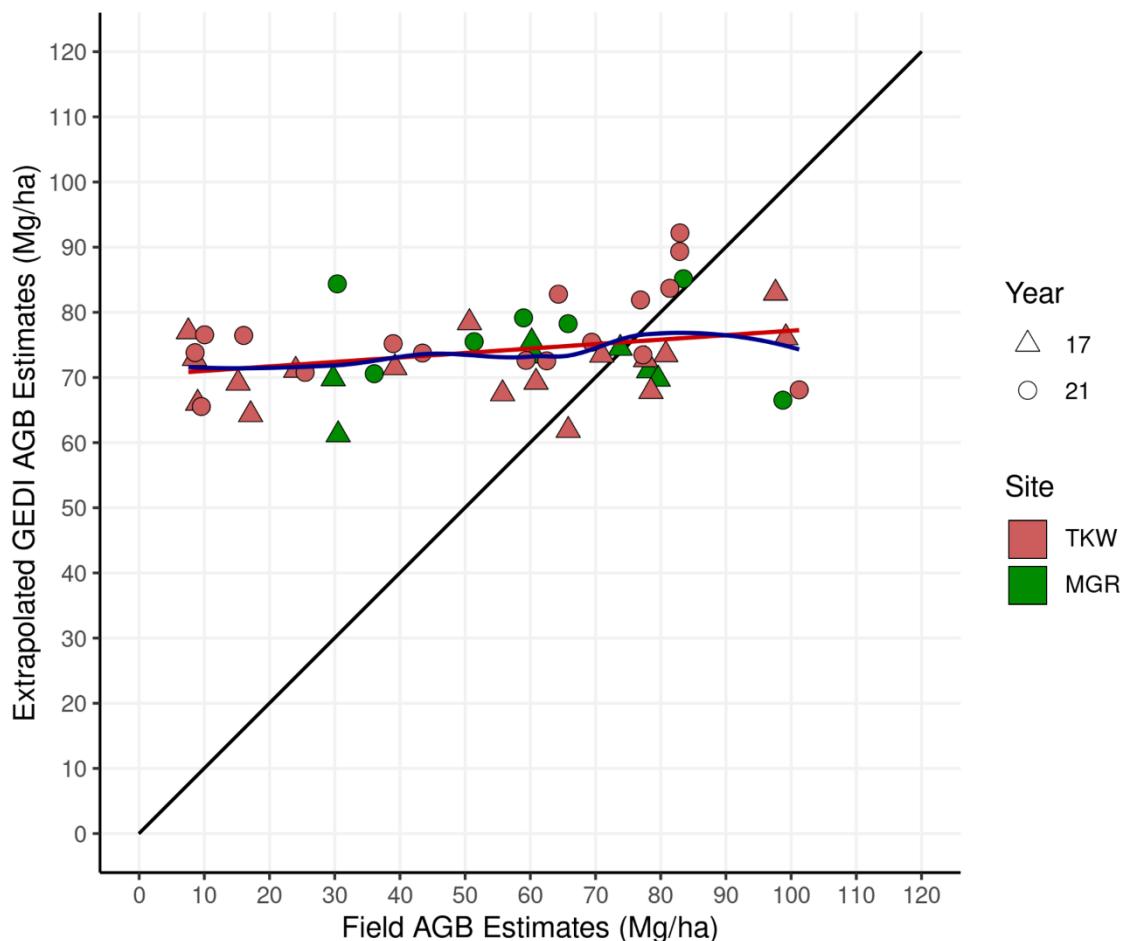


Figure 3.4: Extrapolated GEDI AGBD Estimates and Field AGBD Estimates. Points represent 48 field AGBD estimates from 24 plots across both sites. Colours denote site and symbol denotes census year. One-to-one (black), linear (red), and non-linear (blue) lines are also shown. Authors own figure.

4. Discussion

The GEDI L4A Footprint AGBD product represents a significant advancement in the global monitoring of carbon stocks and dynamics. However, recent efforts to monitor AGBD across the miombo region and over time have revealed a critical need for comprehensive validation of GEDI AGBD estimates in this systematically under-represented area (Demol et al, 2024; Li et al, 2023; *preprint*; Liang et al, 2023).

Accurate in-situ AGBD estimates are crucial for validating GEDI AGBD estimates (Chave et al, 2019; Duncanson et al, 2019; GFOI, 2020). However, the sparse spatial distribution of GEDI footprints and field AGBD estimates pose a challenge for direct collocation. To address this, two main approaches have been adopted: Simulation and Extrapolation. Simulation involves using airborne LiDAR to ‘simulate’ GEDI waveforms and extend the spatial coverage to match in-situ field plots (Hancock et al. 2019). For example, Li et al. (2023) used airborne LiDAR across 833 km² of northern South Africa to validate GEDI RH metrics (the predictors of AGB) and GEDI AGBD estimates themselves (*preprint*). Extrapolation, on the other hand, involves using other available EO data sources in conjunction with machine learning techniques to formulate predictive relationships of GEDI AGBD estimates across space and time. This does not rely on airborne LiDAR making more feasible for larger scale AGBD estimation. Liang et al. (2023) adopted this approach for 9,107 km² of central Mozambique, although without direct in-situ AGBD estimates. This study was the first to validate extrapolated GEDI AGBD estimates with in-situ field AGBD estimates available across sites, over time.

By addressing RQ1, this study intended to determine which EO variables yielded the strongest predictive power for GEDI AGBD estimates. The top performing models were calibrated with Palsar and Landsat variables, with RF_L and RF_P both yielding R² = 0.27, against a withheld subset of GEDI AGBD estimates. This aligns with the large body of literature demonstrating the sensitivity of L-band SAR backscatter to AGBD across the dry tropics and African savannas (Bouvet et al, 2018; McNicol et al, 2018b; 2023; Mitchard et al, 2009; Ryan et al, 2012; Santoro et al, 2021; Wessels et al, 2023), as well as multispectral surface reflectance and phenological metrics (Coops et al, 2021; Healey et al, 2020; Kacic et al, 2021; Liang et al, 2023; Potapov et al, 2021; Ryan, 2009; Ryan et al, 2013; 2017; Zolkos et al, 2013). Furthermore, the 10-fold cross validation approach showed greater differentiation between the model groupings with the median R² of RF_L = 0.25, RF_P = 0.18, RF_S = 0.15. However, the inclusion of all predictor variables (RF_A) yielded the strongest predictive performance against a withheld subset of GEDI AGBD estimates, with R² = 0.29, RMSE = 23.12 Mg/ha, and Bias = +2.65 Mg/ha. All models showed significant overestimation of lower AGBD (< 50 Mg/ha) and

underestimation of higher AGBD (> 50 Mg/ha) commonly observed with RF models (Ploton et al, 2020), and signal saturation occurring between 60-80 Mg/ha (Bouvet et al, 2018).

These results demonstrate the importance of utilising multiple EO data sources to effectively capture the heterogeneity and complexities of miombo landscapes. However, they also indicate that the predictive performance of the models calibrated in this study was significantly lower compared with other studies. Liang et al. (2023) used an extrapolation approach to calibrate a RF model for GEDI AGBD solely with Landsat variables in central Mozambique, yielding $R^2 = 0.64$, RMSE = 7.08 Mg/ha, Bias = +1.06 Mg/ha, relative to separate Landsat-derived metrics corroborated by in-situ field accounts. Li et al. (*preprint*), used a simulation approach for GEDI AGBD estimation, comparing the L4A product with in-situ field AGBD estimates which yielded $R^2 = 0.42$ and RMSE = 12.07 Mg/ha.

By addressing RQ2, this study examined the spatial transferability of extrapolated GEDI AGBD estimates across miombo landscapes with site cross validation. RF models calibrated and tested with data from each site yielded improved predictive performance with $R^2 = 0.33$ and $R^2 = 0.38$ for RF_{TKW} and RF_{MGR} , respectively. However, when RF models were tested with site cross validation, each model yielded quasi-null predictive power with $R^2 = 0.12$ ($RF_{TKW-MGR}$) and $R^2 = 0.07$ ($RF_{MGR-TKW}$). These results are dramatic and suggest that the underlying spatial autocorrelation prevents representation across different, unseen landscapes (Meyer et al, 2019; Ploton et al, 2020; Roberts et al, 2017). However, both site-specific models highlight the significant improvements possible for the accuracy of predictions with local calibration (Healey et al, 2020; Li et al, *preprint*). This study is the first to quantify the significance of the spatial transferability of extrapolated GEDI AGBD estimates in the miombo region, finding that the heterogeneity and complexity between miombo landscapes is stark and a significant hinderance to reliable extrapolations of GEDI AGBD estimates at larger scales.

By addressing RQ3, a time series of spatially contiguous (extrapolated) GEDI AGBD estimates was produced for both sites from 2017 to 2023. However, the relatively high model RMSE prevents any meaningful interpretations of AGBD change across either site, with changes remaining well within 23.12 Mg/ha. Furthermore, overall, in-situ validation reveals quasi-null predictive power, with $R^2 = 0.09$, RMSE = 33.42 Mg/ha, and Bias = +19.75 Mg/ha. While there is variability in validation metrics between sites and years, the relatively small sample size of field AGBD estimates hinder meaningful interpretations for individual censuses. Extrapolated GEDI AGBD estimates show very little sensitivity (61.20 - 92.19 Mg/ha) in relation to field AGBD estimates (7.54 - 101.24 Mg/ha), drastically overestimating lower AGBD and significantly underestimating higher AGBD (Ploton et al, 2020). The results from this in-situ validation demonstrate that extrapolated GEDI AGBD estimates have very little overall predictive power compared with field AGBD estimates in miombo landscapes, despite showing moderate performance compared with a withheld subset.

This study finds that the extrapolation of GEDI AGBD estimates across the miombo region is unreliable. However, several critical reflections should be considered. For example, this study has aimed to calibrate its predictive models with a comprehensive and informed combination of predictor variables from a variety of sources, although there may be additional, appropriate metrics and sources which can improve upon the performance found here. In addition, while RF models were used exclusively here, the adoption of deep learning techniques may improve upon the predictive relationships of GEDI AGBD estimates in the miombo region (Dong et al, *preprint*). Furthermore, the systematic geolocation error of GEDI footprints was filtered by measure of reasonable difference (1σ) between GEDI Canopy Cover Fraction and NDVI (Liang et al, 2023), however, this reduced the available GEDI sample significantly, perhaps more so than other alternative methods. Moreover, this study quantitatively examined the severity of spatial autocorrelation affecting model transferability but did not attempt to control it in subsequent predictions. Further studies would benefit from systematic controls of spatial autocorrelation (Meyer et al, 2019; Ploton et al, 2020; Roberts et al, 2017). Ultimately, this study emphasises the importance of utilising in-situ field AGB estimates for either calibration or validation of EO AGBD estimates and supports the proponents of locally calibrated GEDI AGBD estimates (Li et al, *preprint*).

5. Conclusion

This study quantitatively assessed the reliability of extrapolated GEDI AGBD estimates with in-situ field validation across two miombo landscapes over two census years: 2017 and 2021. Landsat NDVI-derived phenology variables and Palsar L-band backscatter variables were the most important predictors of GEDI AGBD estimates in RF models. However, the best overall predictive performance against a withheld subset of GEDI AGBD estimates included all predictor variables from Landsat, PALSAR-2, Sentinel-1, and SRTM. Despite moderate overall performance, spatial cross validation demonstrated improved performance with local calibration and testing, but revealed dramatically poorer performance when tested with different, unseen miombo landscapes. Furthermore, in-situ validation of extrapolated GEDI AGBD estimates yielded quasi null predictive power and consistent prediction bias across miombo landscapes and over time. Ultimately, this study highlights the importance of in-situ field AGBD estimates for calibration and or validation of EO AGBD estimates and cautions the space-for-time substitution in heterogenous and complex landscapes alike those in the miombo region.

6. References

- Avitabile, V.; Herold, M.; Heuvelink, G.; Lewis, S.; Phillips, O.; Asner, G.; Armston, J.; Ashton, P.; Banin, L.; Bayol, N. et al. (2016). An Integrated Pan-Tropical Biomass Map Using Multiple Reference Datasets. *Global Change Biology*, 22, p1406-1420.
- Baccini, A.; Goetz, S.; Walker, W.; Laporte, N.; Sun, M.; Sulla-Menashe, D.; Hackler, J.; Beck, P.; Dubayah, R.; Friedl, M. et al. (2012). Estimated Carbon Dioxide Emissions from Tropical Deforestation Improved by Carbon-Density Maps. *Nature Climate Change*, 2, p182-185.
- Baccini, A.; Walker, W.; Carvalho, L.; Farina, M.; Sulla-Menashe, D.; Houghton, R. (2017). Tropical Forests are a Net Carbon Source Based on Aboveground Measurements of Gain and Loss. *Science*, 358(6360), p230- 234.
- Belgiu, M. & Drăgut, L. (2016). Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, p24-31.
- Bouvet, A.; Mermoz, S.; Toan, T.; Villard, L.; Mathieu, R.; Naidoo, L.; Asner, G. (2018). An Above-Ground Biomass Map of African Savannahs and Woodlands at 25 m Resolution Derived from ALOS PALSAR. *Remote Sensing of Environment*, 206, p156-173
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, p5-32.
- Chave, J.; Andalo, C.; Brown, S.; Cairns, M.; Chamber, J.; Eamus, D.; Fölster, H.; Fromard, F.; Higuchi, N.; Kira, T. et al. (2005). Tree Allometry and Improved Estimation of Carbon Stocks and Balance in Tropical Forests. *Ecosystem Ecology*, 145, p87-99.
- Chave, J.; Davies, S.; Phillips, O.; Lewis, S.; Sist, P.; Schepaschenko, D.; Armston, J.; Baker, T.; Coomes, D.; Disney, M. et al. (2019). Ground Data are Essential for Biomass Remote Sensing Missions. *Surveys in Geophysics*, 40, p863-880.
- Chave, J.; Réjou-Méchain, M.; Burquez, A.; Chidumayo, E.; Colgan, M.; Delitti, W.; Duque, A.; Eid, T.; Fearnside, P.; Goodman, R. et al. (2014). Improved Allometric Models to Estimate the Aboveground Biomass of Tropical Trees. *Global Change Biology*, 20(10), p3177-3190.
- Clark, D. & Kellner, J. (2012). Tropical Forest Biomass Estimation and the Fallacy of Misplaced Concreteness. *Journal of Vegetation Science*, 23(6), p1191-1196.
- Colgan, S.; Asner, G.; Swemmer, T. (2013). Harvesting Tree Biomass At the Stand Level to Assess the Accuracy of Field and Airborne Biomass Estimation in Savannas. *Ecological Applications*, 23, p1170-1184.
- Coops, N.; Tompalski, P.; Goodbody, T.; Queinnec, M.; Luther, J.; Bolton, D.; White, J.; Wulder, M.; Lier, O.; Hermosilla, T. (2021). Modelling Lidar-Derived Estimates of Forest Attributes Over Space and Time: A Review of Approaches and Future Trends. *Remote Sensing of Environment*, 260, p1-16.
- Cushman, K.; Armston, J.; Dubayah, R.; Duncanson, L.; Hancock, S.; Janík, D.; Král, K.; Krucek, M.; Minor, D.; Tang, H. et al. (2023). Impact of Leaf Phenology on Estimates of Aboveground Biomass Density in a Deciduous Broadleaf Forest from Simulated GEDI Lidar. *Environmental Research Letters*, 18(6), p1-10.
- Davies, R.; Ryan, C.; Harrison, R.; Dexter, K.; Ahrends, A.; Beest, M.; Benitez, L.; Brade, T.; Carreiras, J.; Druce, D. et al. (2023). Precipitation Gradients Drive High Tree Species Turnover in the Woodlands of Eastern and Southern Africa. *Ecography*, 2023(10), p1-13.
- Demol, M.; Aguilar-Amuchastegui, N.; Bernotaite, G.; Disney, M.; Duncanson, L.; Elmendorp, E.; Espejo, A.; Furey, A.; Hancock, S.; Hansen, J. et al. (2024). Multi-Scale Lidar Measurements Suggest Miombo Woodlands Contain Substantially More Carbon Than Thought. *Nature Communications Earth & Environment*, 5(366), p1-11.
- Dinerstein, E.; Olseb, D.; Joshi, A.; Vynne, C.; Burgess, D.; Wikramanayake, E.; Hahn, N.; Palminteri, S.; Hedao, P.; Noss, R. et al. (2017). An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience*, 67(6), p534-545.

- Dong, W.; Mitchard, E.; Yu, H.; Hancock, S.; Ryan, C. (*preprint*). Forest Aboveground Biomass Estimation Using GEDI and Earth Observation Data Through Attention-Based Deep Learning. Available at arXiv: <https://arxiv.org/abs/2311.03067>.
- Dubayah, R.; Blair, J.; Beck, J.; Wirt, B.; Armston, J.; Hofton, M.; Luthcke, S.; Tang, H. (2021). Global Ecosystem Dynamics Investigation (GEDI) Level 2 User Guide, Version 2.0. Available: https://lpdaac.usgs.gov/documents/986/GEDI02_UserGuide_V2.pdf. Accessed: 14/07/2024.
- Dubayah, R.; Blair, J.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurt, G.; Kellner, J.; Luthcke, S. et al. (2020). The Global Ecosystem Dynamics Investigation: High Resolution Laser Ranging of the Earth's Forests and Topography. *Science of Remote Sensing*, 1, p1-16.
- Duncanson, L.; Armston, J.; Disney, M.; Avitabile, V.; Barbier, N.; Calders, K.; Carter, S.; Chave, S.; Herold, M.; Crowther, T. et al. (2019). The Importance of Consistent Global 56 Forest Aboveground Biomass Product Validation. *Surveys in Geophysics*, 40(4), p979- 999.
- Duncanson, L.; Kellner, J.; Armston, J.; Dubayah, R.; Minor, D.; Hancock, S.; Healey, S.; Patterson, P.; Saarela, S.; Marselis, S. et al. (2022). Aboveground Biomass Density Models for NASA's Global Ecosystem Dynamics Investigation (GEDI) Lidar Mission. *Remote Sensing of Environment*, 270, p1-20.
- Ene, L.; Næsset, E.; Gobakken, T.; Bollandsås, O.; Mauya, E.; Zahabu, E. (2017). Large-Scale Estimation of Change in Aboveground Biomass in Miombo Woodlands using Airborne Laser Scanning and National Forest Inventory Data. *Remote Sensing of Environment*, 188, p106-117.
- FAO [Food & Agricultural Organisation of the United Nations]. (2009). 'Global Terrestrial Observing System: Biomass'. In: *Assessment of the Status of the Development of the Standards for the Terrestrial Essential Climate Variables*. FAO: Rome, Italy.
- Feldpausch, T.; Lloyd, J.; Lewis, S.; Brienen, R.; Gloor, M.; Mendoza, A.; Lopez-Gonzalez, G.; Banin, L.; Salim, K.; Affum-Baffoe, K. et al. (2012). Tree Height Integrated into Pantropical Forest Biomass Estimates. *Biogeosciences*, 9, p3381-3403.
- Friedl, M.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. (2010). MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterisation of New Datasets. *Remote Sensing of Environment*, 114, p168-182.
- Friedlingstein, P.; O'Sullivan, M.; Jones, M.; Andrew, R.; Bakker, D.; Hauck, J.; Landschützer, P.; Le Quéré, C.; Luijkz, I.; Peters, G. et al. (2023). Global Carbon Budget 2023. *Earth System Science Data*, 15(12), p5301-5369.
- GFOI [Global Forest Observations Initiative]. (2020). *Integration of Remote-Sensing and Ground-Based Observations for Estimation of Emissions and Removals of Greenhouse Gases in Forests: Methods and Guidance*. (eds). Available: <https://www.reddcompass.org/mgd/resources/GFOI-MGD-3.1-en.pdf>. Accessed: 05/03/2024.
- Gibbs, H.; Brown, S.; Niles, J.; Foley, J. (2007). Monitoring and Estimating Tropical Forest Carbon Stocks: Making REDD a Reality. *Environmental Research Letters*, 2, p1-13.
- Godlee, J. (2021). *Biodiversity-Ecosystem Function Relationships in Southern African Woodlands*. PhD Thesis. University of Edinburgh.
- Godlee, J.; Ryan, C.; Bauman, D.; Bowers, S.; Carreiras, J.; Chisingui, A.; Cronsigt, J.; Druce D.; Finckh, M. et al. (2021). Structural Diversity and Tree Density Drives Variation in the Biodiversity-Ecosystem Function Relationship of Woodlands and Savannas. *New Phytologist*, 232(2), p579-594.
- Goetz, S.; Hansen, M.; Houghton, R.; Walker, W.; Laporte, N.; Busch, J. (2015). Measurement and Monitoring Needs, Capabilities and Potential for Addressing Reduced Emissions from Deforestation and Forest Degradation Under REDD+. *Environmental Research Letters*, 10(12), p1-24.
- Hancock, S.; Armston, J.; Hofton, M.; Sun, X.; Tang, H.; Duncanson, L.; Kellner, J.; Dubayah, R. (2019). The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. *Earth and Space Science*, 6(2), p294-310.

- Harris, N.; Gibbs, D.; Baccini, A.; Birdsey, R.; Bruun, S.; Farina, M.; Fatoyinbo, L.; Hansen, M.; Herold, M.; Houghton, R. et al. (2021). Global Maps of Twenty-First Century Forest Carbon Fluxes. *Nature Climate Change*, 11, p234-240.
- Healey, S.; Yang, Z.; Gorelick, N.; Ilushchenko, S. (2020). Highly Local Model Calibration with a New GEDI LiDAR Asset on Google Earth Engine Reduces Landsat Forest Height Signal Saturation. *Remote Sensing*, 12(17), p1-10.
- Hill, T.; Williams, M.; Bloom, A.; Mitchard, E.; Ryan, C. (2013). Are Inventory Based and Remotely Sensed Above-Ground Biomass Estimates Consistent? *PLOS One*, 8(9), p1-8.
- Hofton, M. & Blair, B. (2019). *Algorithm Theoretical Basis Document (ATBD) for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products Version 1*. Available: https://lpdaac.usgs.gov/documents/581/GEDI_WF_ATBD_v1.0.pdf. Accessed: 07/07/2024.
- Houghton, R. (2005). Aboveground Forest Biomass and the Global Carbon Balance. *Global Change Biology*, 11(6), p945-958.
- Huntley, B. (2023). *The Mesic Savanna Biome*. In: 'Ecology of Angola'. Springer: Cham, Germany.
- IPCC [Intergovernmental Panel on Climate Change]. (2006). 'Agriculture, Forestry, and Other Land Use'. In: Eggleston, S. et al. (eds) *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. IGES: Hayama, Japan.
- Kacic, P.; Hirner, A.; Ponte, E. (2021). Fusing Sentinel-1 and -2 to Model GEDI-Derived Vegetation Structure Characteristics in GEE for the Paraguayan Chaco. *Remote Sensing*, 13(24), p1-17.
- Kellner, J.; Armston, J.; Duncanson, L. (2023). Algorithm Theoretical Basis Document for GEDI Footprint Aboveground Biomass Density. *Earth and Space Science*, 10(4), p1-20.
- Leite, R.; Silva, C.; Broadbent, E.; Amaral, C.; Liesenberg, V.; Almeida, D.; Mohan, M.; Godinho, S.; Cardil, A.; Hamamura, C. et al. (2022). Large Scale Multi-Layer Fuel Load Characterisation in Tropical Savanna using GEDI Spaceborne Lidar Data. *Remote Sensing of Environment*, 268, p1-19.
- Li, X.; Wessels, K.; Armston, J.; Duncanson, L.; Urbazaev, M.; Naidoo, L.; Mathieu, R.; Main, R. (preprint). Evaluation of GEDI Footprint Level Biomass Models in Southern African Savannas using Airborne Lidar and Field Measurements. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4716466.
- Li, X.; Wessels, K.; Armston, J.; Hancock, S.; Mathieu, R.; Main, R.; Naidoo, L.; Erasmus, B.; Scholes, B. (2023). First Validation of GEDI Canopy Heights in African Savannas. *Remote Sensing of Environment*, 285, p1-17.
- Liang, M.; Duncanson, L.; Silva, J.; Sedano, F. (2023). Quantifying Aboveground Biomass Dynamics from Charcoal Degradation in Mozambique Using GEDI Lidar and Landsat. *Remote Sensing of Environment*, 284, p1-15.
- McNicol, I.; Keane, A.; Burgess, N.; Bowers, S.; Mitchard, E.; Ryan, C. (2023). Protected Areas Reduce Deforestation and Degradation and Enhance Woody Growth Across African Woodlands. *Nature Communications Earth & Environment*, 4(392), p1-14.
- McNicol, I.; Ryan, C.; Dexter, K.; Ball, S.; Williams, M. (2018a). Aboveground Carbon Storage and Its Links to Stand Structure, Tree Diversity and Floristic Composition in South-Eastern Tanzania. *Ecosystems*, 21, p740-754.
- McNicol, I.; Ryan, C.; Mitchard, E. (2018b). Carbon Losses from Deforestation and Widespread Degradation Offset by Extensive Growth in African Woodlands. *Nature Communications*, 9(3045), p1-11.
- McRoberts, R. & Tomppo, E. (2007). Remote Sensing Support for National Forest Inventories. *Remote Sensing of Environment*, 110(4), p412-419.
- Menaut, J.; Lepage, M.; Abbadie, L. (1995). 'Savannas, Woodlands, and Dry Forests in Africa'. In: Mooney, H. et al. *Seasonally Dry Tropical Forests*. Cambridge University Press: Cambridge, England. p64-96.
- Meyer, H.; Reudenbach, C.; Wöllauer, S.; Nauss, T. (2019). Importance of Spatial Predictor Variable Selection in Machine Learning Applications – Moving from Data Reproduction to Spatial Prediction. *Ecological Modelling*, 411, p1-11.

- Mitchard, E.; Saatchi, S.; Baccini, A.; Asner, G.; Goetz, S.; Harris, N.; Brown, S. (2013). Uncertainty in the Spatial Distribution of Tropical Forest Biomass: A Comparison of Pan-Tropical Maps. *Carbon Balance and Management*, 8(10), p1-13.
- Mugasha, W.; Eid, T.; Bollandsås, O.; Malimbwi, R.; Chamshama, S.; Zahabu, E.; Katani, J. (2013). Allometric Models for Prediction of Aboveground and Belowground Biomass of Trees in the Miombo Woodlands of Tanzania. *Forest Ecology and Management*, 310, p87-101.
- Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H. et al. (2021). ERA-5 Land: A State-of-the-Art Global Reanalysis Dataset for Land Applications. *Earth System Science Data*, 13(9), p4349-4383.
- Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V.; Dormann, C.; Cornu, G.; Viennois, G.; Bayol, N. et al. (2020). Spatial Validation Reveals Poor Predictive Performance of Large-Scale Ecological Mapping Models. *Nature Communications*, 11(4050), p1-11.
- Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C. et al. (2021). Mapping Global Forest Canopy Height Through the Integration of GEDI and Landsat Data. *Remote Sensing of Environment*, 253, p1-11.
- Pötzschner, F.; Baumann, M.; Gasparri, N.; Conti, G.; Loto, D.; Piquer-Rodríguez, M.; Kuemmerle, T. (2022). Ecoregion-Wide, Multi-Sensor Biomass Mapping Highlights a Major Underestimation of Dry Forests Carbon Stocks. *Remote Sensing of Environment*, 269, p1-12.
- Poulter, B.; Ciais, P.; Hodson, E.; Lischke, H.; Maignan, F.; Plummer, S.; Zimmermann, N. (2011). Plant Functional Type Mapping for Earth System Models. *Geoscientific Model Development*, 4, p993-1010.
- Roberts, D.; Bahn, V.; Cuiti, S.; Boyce, M.; Elith, J.; Guillera-Arriota, G.; Hauenstein, S.; Lahoz-Monfort, J.; Schröder, B.; Thuiller, W. et al. (2017). Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography*, 40(8), p913-929.
- Rodríguez-Veiga, P.; Quegan, S.; Carreiras, J.; Persson, H.; Fransson, J.; Hoscilo, A.; Ziolkowski, D.; Stereńczak, K.; Lohberger, S.; Stängel, M. et al. (2019). Forest Biomass Retrieval Approaches from Earth Observation in Different Biomes. *International Journal of Applied Observation and Geoinformation*, 77, p53-68.
- Rosenqvist, A.; Shimada, M.; Suzuki, S.; Ohgushi, F.; Tadano, T.; Watanabe, M.; Tsuzuku, K.; Watanabe, T.; Kamijo, S.; Aoki, E. (2014). Operation Performance of the ALOS Global Systematic Acquisition Strategy and Observation Plans for ALOS-2 PALSAR-2. *Remote Sensing of Environment*, 155, p3-12.
- Roy, D.; Kashongwe, H.; Armston, J. (2021). The Impact of Geolocation Uncertainty on GEDI Tropical Forest Canopy Height Estimation and Change Monitoring. *Science of Remote Sensing*, 4, p1-19.
- Roy, D.; Wulder, M.; Loveland, T.; Woodcock, C.; Allen, R.; Anderson, M.; Helder, D.; Irons, J.; Johnson, D.; Kennedy, R. et al. (2014). Landsat-8: Science and Product Vision for Terrestrial Global Change Research. *Remote Sensing of Environment*, 145, p154-172.
- Ryan, C. & Williams, M. (2011). How Does Fire Intensity and Frequency Affect Miombo Woodland Tree Populations and Biomass? *Ecological Applications*, 21, p48-60.
- Ryan, C. (2009). *Carbon Cycling, Fire and Phenology in a Tropical Savanna Woodland in Nhambita, Mozambique*. PhD Thesis. University of Edinburgh.
- Ryan, C.; Hill, T.; Woollen, E.; Ghee, C.; Mitchard, E.; Cassells, G.; Grace, J.; Woodhouse, I.; Williams, M. (2012). Quantifying Small-Scale Deforestation and Forest Degradation in African Woodlands Using Radar Imagery. *Global Change Biology*, 18, p243-257.
- Ryan, C.; Pritchard, R.; McNicol, I.; Owen, M.; Fisher, J.; Lehmann, C. (2016). Ecosystem Services from Southern African Woodlands and Their Future Under Global Change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, p1-16.
- Ryan, C.; Williams, M.; Grace, J. (2011). Above- and Belowground Carbon Stocks in a Miombo Woodland Landscape of Mozambique. *Biotropica*, 43(4), p423-432.

- Ryan, C.; Williams, M.; Grace, J.; Woollen, E.; Lehmann, C. (2017). Pre-Rain Green-Up is Ubiquitous Across Southern Tropical Africa: Implications for Temporal Niche Separation and Model Representation. *New Phytologist*, 213(2), p625-233.
- Ryan, C.; Williams, M.; Hill, T.; Grace, J.; Woodhouse, I. (2013). Assessing the Phenology of Southern Tropical Africa: A Comparison of Hemispherical Photography, Scatterometry, and Optical/NIR Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 52, p1-10.
- Saatchi, S.; Harris, N.; Brown, S.; Lefsky, M.; Mitchard, E.; Salas, W.; Zutta, B.; Buermann, W.; Lewis, S.; Hagen, S. et al. (2011). Benchmark Map of Forest Carbon Stocks in Tropical Regions Across Three Continents. *Proceedings of the National Academy of Sciences*, 108, p9899-9904.
- Santoro, M.; Cartus, O.; Carvalhais, N.; Rozendaal, D.; Avitabile, V.; Araza, A.; Bruin, S.; Herold, M.; Quegan, S.; Rodríguez-Veiga, P. et al. (2021). The Global Forest Above-Ground Biomass Pool for 2010 Estimated from High-Resolution Satellite Observations. *Earth System Science Data*, 13(8), p3927-3950.
- SEOSAW. (2021). A Network to Understand the Changing Socio-Ecology of the Southern African Woodlands (SEOSAW): Challenges, Benefits, and Methods. *Plants, People, Planet*, 3(3), p249-267.
- Simard, M.; Pinto, N.; Fisher, J.; Baccini, A. (2011). Mapping Forest Canopy Height Globally with Spaceborne Lidar. *Biogeosciences*, 116(4), p1- 12.
- Sulla-Menashe, D. & Friel, M. (2018). User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product. Available: https://lpdaac.usgs.gov/documents/101/MCD12_User_Guide_V6.pdf. Accessed 14/07/2024.
- Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Flouri, N.; Brown, M. et al. (2012). GMES Sentinel-1 Mission. *Remote Sensing of Environment*, 120, p9-24.
- Tucker, C.; Brandt, M.; Hiernaux, P.; Kariyaa, A.; Rasmussen, K.; Small, J.; Igel, C.; Reiner, F.; Melocik, K.; Meyer, J. et al. (2023). Sub-Continental-Scale Carbon Stocks of Individual Trees in African Drylands. *Nature*, 615, p80-86.
- Ustin, S. & Gamon, J. (2010). Remote Sensing of Plant Functional Types. *New Phytologist*, 186(4), p795-816.
- Van Zyl, J. (2001). The Shuttle Radar Topography Mission (SRTM): A Breakthrough in Remote Sensing of Topography. *Acta Astronautica*, 48, p559-565.
- White, F. (1983). *The Vegetation of Africa: A Descriptive Memoir to Accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa*. UNESCO: Paris, France.
- Wigley, B.; Coetsee, C.; February, E.; Dobelmann, S.; Higgins, S. (2024). Will Trees or Grasses Profit from Changing Rainfall Regimes in Savannas? *New Phytologist*, 241(6), p1-16.
- Xue, J. & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, 2017, p1-18.
- Zolkos, S.; Goetz, S.; Dubayah, R. (2013). A Meta-Analysis of Terrestrial Aboveground Biomass Estimation using Lidar Remote Sensing. *Remote Sensing of Environment*, 128, p289-298.

Part II: Technical Report



Miombo Woodlands, Northern Zambia (Walker, 2024)

Contents

List of Figures	ii
List of Tables	ii
List of Formulas	ii
1. Introduction	1
1.1. Research Aims and Questions	1
1.2. Computing Requirements	1
1.3. Code Repositories	2
2. Data	3
2.1. SEOSAW Field AGB Estimates	3
2.2. GEDI Footprint AGB Estimates	3
2.3. GEDI Quality Filtering	5
2.4. GEDI Sensitivity to Phenology	7
2.5. Predictor Variables	10
3. Model Implementation	15
3.1. Parameter Configuration	15
3.2. Model Performances by EO Source	15
3.3. Model Performances by Site Cross Validation	17
4. In-Situ Validation	19
5. References	22
6. Directory Structure and File Locations	25
7. Appendix	27

List of Figures

2.1. SEOSAW PSP Locations.	3
2.2. Examination of Seasonal Artefacts in GEDI AGB Estimates.	9
2.3. Predictor Variable Cross-Correlation Matrix.	13
3.1. Model Performances by EO Source: Bias.	16
3.2. Model Performances by EO Source: RMSE.	16
3.3. Comparison of TKW Model Performances with Spatial Cross Validation.	17
3.4. Comparison of MGR Model Performances with Spatial Cross Validation.	18
4.1. Distribution of Extrapolated GEDI AGB Estimates and Field AGB Estimates.	19
4.2. In-Situ Validation of Extrapolated GEDI AGB Estimates Across Site, Over Census Years.	20

List of Tables

1.1. Research Questions.	1
1.2. Summary of Main Code Scripts	2
2.1. Calibration Data for GEDI AGB Estimates in the Miombo Region.	4
2.2. Plant Functional Types for GEDI AGB Model Development.	4
2.3. Filtered GEDI Footprints Available for Model Calibration.	6
2.4. Complete List of Predictor Variables for GEDI AGB Extrapolation.	10
2.5. Best Performance of Predictor Variables with Linear Regression.	12
2.6. Retained Predictor Variables.	14
3.1. Model Performance by Number of Trees.	15
4.1. Comparison of RF Model Performance and Validation.	21

List of Formulas

2.1. GEDI Geolocation Filtered Condition.	6
2.2. PALSAR Backscatter Coefficient Conversion.	6

1. Introduction

This Technical Report accompanies **Part I: Research Paper** of this dissertation titled “*In-Situ Validation Reveals Poor Performance of Extrapolated GEDI Aboveground Biomass Estimates Across Miombo Landscapes*”. This acts as supplementary material for the research paper, providing more detailed explanations and justifications of the decisions, processes, and outputs of the study.

1.1. Research Aims and Questions

The context for and framing of this Technical Report is based on the Research Aims and Questions in **Part I: Research Paper**. The overall aim of this study was to quantitatively assess the predictive performance of Earth Observation (EO)-extrapolated Global Ecosystem Dynamics Investigation (GEDI) Aboveground Biomass (AGB) estimates across space and examine the reliability of GEDI AGB estimates over time using in-situ field AGB estimates. The three research questions formulated to address this aim are repeated here in Table 1.1.

Table 1.1: Research Questions.

Notation	Research Question
RQ1	How consistent is predictive performance when different EO predictor variables are used to extrapolate GEDI AGB estimates across both sites?
RQ2	How consistent is predictive performance where extrapolated GEDI AGB estimates with spatial (site) cross validation?
RQ3	How consistent is bias of extrapolated GEDI AGB estimates relative to in-situ field AGB estimates across both sites over time?

1.2. Computing Requirements

The work presented in both **Part I: Research Paper** and here in **Part II: Technical Report** was predominantly undertaken on the University of Edinburgh School of GeoSciences Linux servers, *baltic* and *stream*. The *baltic* server was used to perform tests and non-intensive tasks while the more intense tasks such as running each of the 62 Random Forest (RF) machine learning algorithms was performed on the *stream* server. Readers are advised that at the time of publication, the *stream* server is set to be replaced by *slurm* in the “Forth” High-Performance Compute cluster by September 2024.

All the EO data used in this study were accessed using Google Earth Engine (GEE), a cloud-computing platform with a vast archive of EO and geospatial datasets. GEE was predominantly used for data acquisition and processing as is expanded upon further in “2. Data”. Preliminary tasks and some data visualisation throughout this study were performed in RStudio with R version 4.1.2. The main model development and validation tasks were performed using Python version 3.10.12. Map visualisation was completed using QGIS version 3.18.

1.3. Code Repositories

Several scripts were developed for this study with JavaScript for GEE, R for RStudio, and Python. A complete repository of scripts used in this study are categorised and hosted in a public GitHub repository (Wilson et al, 2017), accessible [here](#). In this repository, a README.md file is provided, outlining the capability of each script along with the built-in customisation with command line arguments, as well as explanatory docu-strings and comments. Table 1.2 presents a summary of the main scripts developed for this research.

Table 1.2: Summary of Main Code Scripts.

Script	Description
libraries.py	Contains the libraries and methods used in all python scripts.
gediSeasonality.R	Examines and visualises seasonal artefacts in GEDI L4A Product.
extractData.py	Reads predictor variable data intersecting with GEDI footprints.
alignData.py	Merges input data and performs geolocation filtered condition.
filterFeatures.py	Applies paired correlation coefficient filtered condition.
assessData.py	Performs initial linear regression and visualises available dataset.
testModelTrees.py	Examines Random Forest performance with various configurations.
trainModel.py	Calibrates and tests grouped Random Forest model conditions.
testModelSite.py	Calibrates and tests site-specific Random Forest model conditions.
predictBiomass.py	Extrapolates GEDI AGB estimates across sites and years.
validateBiomass.py	Examines validity of extrapolated estimates with in-situ data.

2. Data

2.1. SEOSAW Field AGB Estimates

For a complete explanation of how each plot census and AGB estimates are conducted from SEOSAW, please consult the SEOSAW Dataset Manual available [here](#). MGR and TKW are some of the most long-standing Permanent Sample Plots (PSPs) in the SEOSAW network, with censuses dating from 2006 to 2021, see Figure 2.1. Each PSP in MGR and TKW are 1 ha (100 m^2) which is notably advantageous for the calibration and or validation of EO products and predictions. The temporal constraint placed on this study, that being to ‘only’ examine seven years (2017–2023) rather than extending to 2006, was made to address RQ1. Not all EO predictor variables were available for the extended time frame. There remain opportunities to expand the scope of this analysis and validation to more SEOSAW sites and over longer time frames.

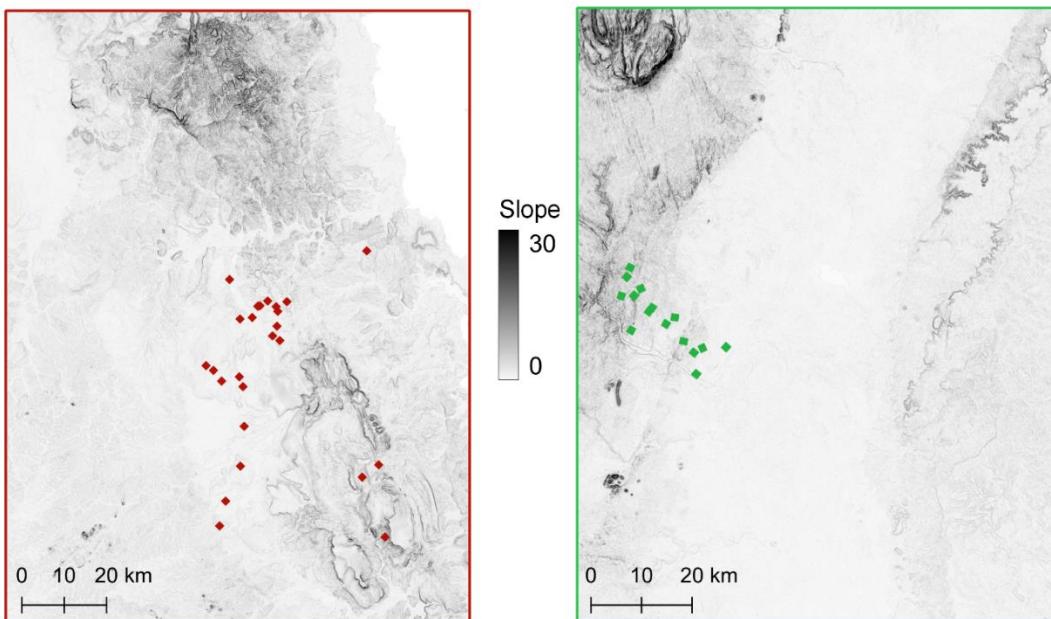


Figure 2.1: SEOSAW PSP Locations. Left: TKW with all 25 PSPs. Right: MGR with all 15 PSPs. PSPs have been arbitrarily inflated. Basemap of terrain slope degrees. Source SEOSAW (2021).

2.2. GEDI Footprint AGB Estimates

The GEDI Footprint AGB estimates are a modelled product. To derive a near global sample of AGB estimates, the GEDI team compiled a global database (Forest Structure and Biomass Database, FSBD) of spatially and temporally coincident airborne LiDAR data and in-situ field plot AGB estimates (Duncanson et al, 2022). Field AGB estimates, for plots approximately the size of GEDI footprints (25 m), were derived from local allometric models, which estimate AGB as a function of structural components of woody vegetation, see Table

2.1. Where field AGB estimates were available with coincident airborne LiDAR data, a GEDI waveform simulator was used to process the airborne LiDAR data and produce GEDI-like waveforms and derived Relative Height (RH) metrics (Hancock et al, 2019). Then, an Ordinary Least Squares (OLS) regression technique was adopted to model field AGB estimates as a function of simulated GEDI RH metrics. GEDI AGB models were stratified by geographic region and Plant Functional Type (PFT).

Table 2.1: Calibration Data for GEDI AGB Estimates in the Miombo Region.

PFT(s)	Country	Site Name(s)	Site Area (ha)	Site Plots (n)	Allometric Model(s)	Relevant Papers
DBT,	Tanzania	Liwale	36	513	Mugasha et al. (2013)	Ene et al. (2017)
EBT,						
GSW						

PFT is an ecological classification based on plant functional traits (such as phenological cycle and growth form) as opposed to their taxonomic groups (such as species or genus) (Lavorel et al, 1997; Tilman et al, 1997; Ustin & Gamon, 2010). The stratification of GEDI models was based on four broad PFTs from the global land and phenology product derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) multispectral instrument, see Table 2.2 (Friedl et al, 2010; Salomonson et al, 1989; Sulla-Menashe & Friedl, 2018). The GEDI AGB models for footprints in the miombo region were predominantly African Deciduous Broadleaf Trees (DBT). The lack of available calibration data representing African Grasses, Shrubs, and Woodlands (GSW) resulted in the adoption of a global GSW model. The African DBT model expressed AGB as a function of RH98 and RH50 (relative heights at which 98th and 50th percentiles of cumulative returned energy), yielding $R^2 = 0.63$ and RMSE% = 57.24. The global GSW expressed AGB as a function of RH50, yielding $R^2 = 0.86$ and RMSE% = 55.39.

Table 2.2: Plant Functional Types for GEDI AGB Model Development.

Plant Functional Type	Notation	Description
Deciduous Broadleaf Trees	DBT	Dominated by trees which shed their ‘broad’ or flat leaves. Tree cover > 10% and Height > 2 m.
Evergreen Broadleaf Trees	EBT	Dominated by trees with ‘broad’ or flat leaves that are not shed. Tree cover > 10% and Height > 2 m.

Evergreen Needleleaf Trees	ENT	Dominated by trees with needles rather than leaves that are not shed. Tree cover > 10% and Height > 2 m.
Grasses, Shrubs, and Woodlands	GSW	Dominated by a combination of herbaceous annuals (non-woody vegetation) that are not cultivated and short woody stands. Tree cover < 10% and Height < 2 m.

2.3. GEDI Quality Filtering

Not all GEDI footprint data can be treated equally. The GEDI instrument consists of four ‘power’ beams (10 mJ) and four ‘coverage’ beams (4.5 mJ) which are split from the three GEDI lasers (Dubayah et al, 2020). Beam sensitivity is a measure of the signal to noise ratio and the proportion of canopy cover through which waveforms penetrate. Liang et al. (2023) evaluated the reliability of the GEDI footprint AGB estimates in Mozambique and found that beam sensitivity of 0.98 or above captured more complex structural properties.

The GEDI team denote the most useful footprints with quality flags at both L2 and L4 product processing stages alongside a degrade flag which indicating poor positioning or pointing performance. In addition, Li et al. (2023) evaluated GEDI RH metrics in African savannas and found RH98 where height exceeded 3 metres were most reliable.

The systematic geolocation error of GEDI footprints (10-20 m) was addressed in the same manner as Liang et al. (2023) who retained footprints where the absolute difference between GEDI L2B Canopy Cover Fraction and Landsat Annual Median NDVI was within one standard deviation of the mean difference, see Formula 2.1. An additional geolocation filter was tested in this study where the same formula was used to express the differences between GEDI L4A AGB estimates and resampled ALOS-2 PALSAR-2 ScanSAR L-band backscatter expressed in decibels (Rosenqvist et al, 2014). All Palsar-derived metrics were converted from Digital Numbers (DN) to backscatter in decibels (dB), see Formula 2.2 (Rosenqvist et al, 2007). However, the geolocation test only retained a smaller threshold of GEDI AGB estimates (< 40 Mg/ha). Given the range and distribution of in-situ field AGB estimates exceeded the retained GEDI AGB estimates, the geolocation filter adopted by Liang et al. was used which retained a larger range and higher threshold of GEDI AGB estimates (< 150 Mg/ha). The geolocation filter condition was the largest contributor to footprint omission.

Formula 2.1: GEDI Geolocation Filtered Condition.

$$| CCF_{GEDI} - \text{NDVI}_{Landsat} | \leq \mu\Delta + \sigma\Delta$$

where:

CCF_{GEDI} = GEDI L2B Canopy Cover Fraction

$\text{NDVI}_{Landsat}$ = Landsat Annual Median Normalised Difference Vegetation Index

$\mu\Delta$ = Mean difference between CCF_{GEDI} and $\text{NDVI}_{Landsat}$

$\sigma\Delta$ = Standard deviation of the difference between CCF_{GEDI} and $\text{NDVI}_{Landsat}$

Formula 2.2: PALSAR Backscatter Coefficient Conversion.

$$\gamma_0 = 10 \cdot \log_{10}(\text{DN}^2) - 83.0 \text{ dB}$$

where:

γ_0 = Backscatter coefficient

DN = Digital Number

Biologically implausible GEDI AGB estimates ($> 300 \text{ Mg/ha}$) were omitted and a temporal constraint was placed to retain observations within April to November, where phenological transitions and defoliation were minimised. A more substantial assessment of GEDI sensitivity to phenology is discussed in the next section. The final, filtered calibration sample of GEDI AGB estimates amounted to 7,906 across both sites, see Table 2.3.

Table 2.3: Filtered GEDI Footprints Available for Model Calibration.

Site	Year	Initial	Filtered
TKW	2020	137,400	119
	2021	266,706	1,943
	2022	178,974	1,389
	2023	167,870	840
	Total	750,950	4,181

MGR	2020	76,759	306
	2021	44,256	751
	2022	132,783	2,119
	2023	120,428	439
	Total	374,226	3,615
Both	2020	214,159	425
	2021	310,962	2,694
	2022	311,757	3,508
	2023	288,298	1,279
	Total	1,125,176	7,906

2.4. GEDI Sensitivity to Phenology

It has been established that GEDI waveform returns and by extension GEDI derived products, are significantly less reliable when interacting with DBTs during defoliation, or “leaf-off” conditions (Cushman et al, 2023; Duncanson et al, 2022; Kellner et al, 2023; Li et al, 2023; Liang et al, 2023). As a result, GEDI products include a leaf-off flag which denotes if footprint returns were received during defoliation (= 1) or leaf presentation (= 0) of DBTs. The leaf-off flag is determined by the 1 km gridded Global Land Surface Phenology (GLSP) product, derived from the National Polar-Orbiting Partnership (NPP) Visible Infrared Imaging Radiometer Suite (VIIRS) instrument data (Zhang et al, 2018). The removal of leaf-off flags significantly affects the available sample of GEDI footprints in the DBT strata. Globally, from April 2019 to November 2021, 55% of the footprints were acquired during leaf-off conditions (Kellner et al, 2023). Filtering by GEDI L4 quality flag includes the omission of leaf-off observations.

However, there are valid concerns as to whether the leaf-off flag can reliably isolate the most accurate GEDI observations in highly heterogenous regions with complex phenological cycles, such as the miombo woodlands of southern Africa (Davies et al, 2023; Godlee, 2021; Godlee et al, 2021; Menaut et al, 1995; McNicol et al, 2018a; Ryan et al, 2013; 2017; Ryan & Williams, 2011; Wigley et al, 2024). Part of the concern is that leaf-off conditions are aggregated to a 1 km gridded product, while GEDI footprints are significantly smaller at ~25 m. In a study where GEDI AGB estimates were extrapolated across an area in Mozambique, Liang et al. (2023) did not apply any additional temporal constraints to the available sample beyond the filtering conditions detailed in the previous section. By contrast, Li et al. (*preprint*)

restricted GEDI AGB estimates between April and November of each year of the GEDI acquisition period. This decision acted to further ensure that GEDI observations acquired during defoliation or phenological transition periods were not included.

To appropriately consider and constrain the reliability of GEDI observations given their sensitivity to phenology in both sites, the temporal autocorrelation and autoregression of monthly aggregated (filtered) GEDI AGB estimates was examined, see Figure 2.2. The top row shows monthly aggregated GEDI AGB estimates, and illustrates visible albeit subtle fluctuations in both sites, with the highest values typically in the known months of leaf presentation between November and April (Menaut et al, 1995). This pattern is more visible in TKW in part because of the lack of available (filtered) GEDI AGB estimates in several months of each year in MGR.

The middle row shows the output from a temporal autocorrelation function (ACF) conducted in RStudio; a sequence of values representing the autocorrelation or dependency of each value of the complete time series of GEDI observations over various time lags (expressed in months). The ACF plots identify statistically significant temporal dependency over time lags of 2 months in TKW and 1, 2, 4 and 5 months in MGR. The results suggest the GEDI AGB estimates in TKW are less seasonally constrained and more dependent on within-season dynamics while those in MGR are more typical of season-season variation.

To express the evidence of seasonal artefacts in units of AGB, an autoregression (AR) model was conducted in RStudio, from which the deviation from the AR modelled time series can be quantified as AR residuals. The bottom row of plots shows the AR residuals and statistically significance thresholds for both sites over the three complete austral years of the GEDI acquisition period. The result shows that there are considerable statistically significant AR residuals in both TKW and MGR. The most significant residuals tend to follow the phenological transition months (August-October and January-April). The findings from this examination suggest that there are seasonal artefacts in GEDI AGB estimates which are important to consider beyond those captured by the leaf-off flag. As a result, this study constrained GEDI AGB estimates to November to April of each austral year alike Li et al. (*preprint*).

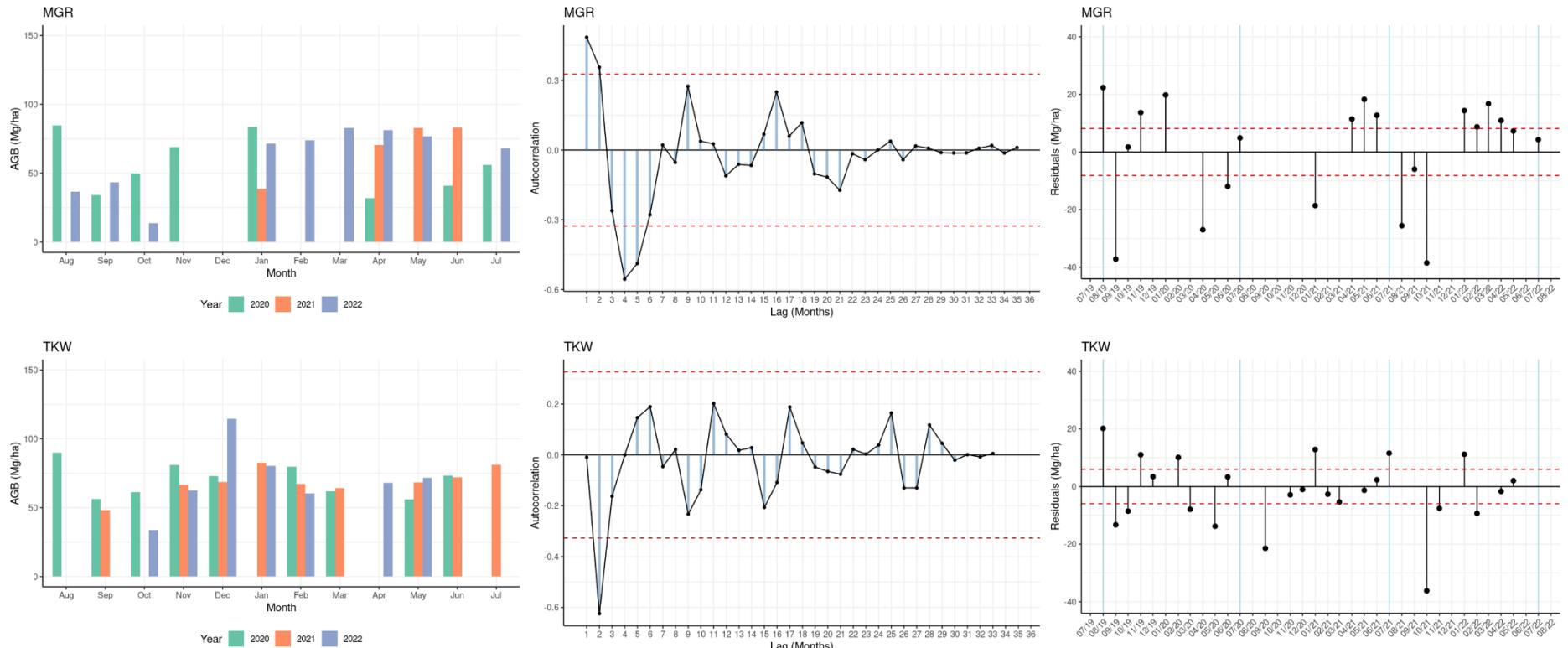


Figure 2.2: Examination of Seasonal Artefacts in GEDI AGB Estimates. Left: TKW. Right: MGR. Top row shows monthly aggregated GEDI AGB estimates for the three complete austral years in the GEDI acquisition period (2020-2022). The middle row shows the temporal autocorrelation (dependency) of GEDI AGB estimates over various time lags (months). The bottom row shows the residuals derived from an autoregression (AR) time series model of GEDI AGB estimates. Authors own figure.

2.5. Predictor Variables

The predictor variables for this study were selected based on a wide array of literature and established relationships between Light Detection and Ranging (LiDAR) instrument data (Coops et al, 2021; Zolkos et al, 2013), passive optical multispectral data and derived vegetation indices (Baccini et al, 2011; Liang et al, 2023; Potapov et al, 2021; Saatchi et al, 2011; Simard et al, 2011), and Synthetic Aperture Radar (SAR) instrument data (Bouvet et al, 2018; McNicol et al, 2018b; 2023; Mitchard et al, 2009; Santoro et al, 2021; Ryan et al, 2011; Wessels et al, 2023). The relationship between topography and distribution of AGB and vegetation has also been demonstrated (Bell, 1982; Woollen et al, 2012). A complete list of the 54 predictor variables and their grouping by EO source is shown in Table 2.4.

Table 2.4: Complete List of Predictor Variables for GEDI AGB Extrapolation.

Source	Variable	Description
Shuttle Radar	Elevation	Height above sea level in metres.
Topography	Slope	Terrain slope in degrees.
Mission (SRTM) 2000	Multi-Scale Topographic Index (mTPI)	Positive-Negative values distinguishing ridges from values relative to mean elevation in metres.
Landsat-8 <i>2013-present</i>	Surface Reflectance Bands 2-7: Median Standard Deviation 95 th Percentile 5 th Percentile	Atmospherically corrected surface reflectance expressed in eight bands from Visible, Near InfraRed, and Short-Wave InfraRed spectrums (0.452-2.294 μm) with less than 10% cloud cover (30% in 2021).
	Normalised Difference Vegetation Index (NDVI): Annual Median Term (1-3) Median 95:5 th Gradient Wet 95 th Percentile Wet 5 th Percentile Dry 95 th Percentile Dry 5 th Percentile	Index derived from Bands 4 (Visible, Red) and 5 (Near InfraRed). Less than 30% cloud cover. Exceptions as Annual Median & Term Medians calculated with less than 50% cloud cover. Wet months defined as those where ERA-5 Land total monthly precipitation > 30 mm.

Sentinel-1 <i>2014-present</i>	C-Band Backscatter - VV Polarisations: Median Standard Deviation 95 th Percentile 5 th Percentile	Ground Range Detected 5.6 cm wavelength backscatter with single co-polarisation: vertical transmit and vertical receive.
	C-Band Backscatter - VH Polarisations: Median Standard Deviation 95 th Percentile 5 th Percentile	Ground Range Detected 5.6 cm wavelength backscatter with dual band cross-polarisation: vertical transmit and horizontal receive.
	L-Band Backscatter - HH Polarisations: Median Standard Deviation 95 th Percentile 5 th Percentile	Normalised, orthorectified, and radiometrically corrected 23 cm wavelength backscatter in ScanSAR mode, manually converted to decibels (Rosenqvist et al, 2007), with single co-polarisation: horizontal transmit and horizontal receive.
	L-Band Backscatter - HV Polarisations: Median Standard Deviation 95 th Percentile 5 th Percentile	Normalised, orthorectified, and radiometrically corrected 23 cm wavelength backscatter in ScanSAR mode, manually converted to decibels (Rosenqvist et al, 2007), with dual band cross polarisation: horizontal transmit and vertical receive.
	HH:HV Ratio	Ratio of HH to HV backscatter.
	HH:HV Index	Normalised difference index of HH and HV backscatter.

Once acquired, each predictor variable was initially tested for its relationship with GEDI AGB estimates using linear regression. This test provided a baseline indication of the strength of predictor variable association. The best performing predictor variables were Palsar-derived and Landsat-derived Median NDVI metrics. The best five performing variables yielded $R^2 = 0.10-0.14$. The outputs of the linear regression for those five variables are shown in Table 2.5.

Table 2.5: Best Performance of Predictor Variables with Linear Regression.

Variable	Coefficient	Intercept	R ²	Rank (R ²)
HV Median	5.13	122.34	0.14	1
HV 5 th Percentile	4.39	126.24	0.13	2
NDVI Median	172.98	- 9.21	0.12	3
HV 95 th Percentile	4.82	102.21	0.11	4
HH Median	5.96	101.05	0.10	5

To reduce the number of predictor variables and improve computing efficiency in later stages of model execution, a cross-correlation matrix was produced, see Figure 2.3. Then, predictor variables with paired correlation coefficients greater than 0.9 were removed (Liang et al, 2023), resulting in a final 36 retained predictor variables, see Table 2.6.

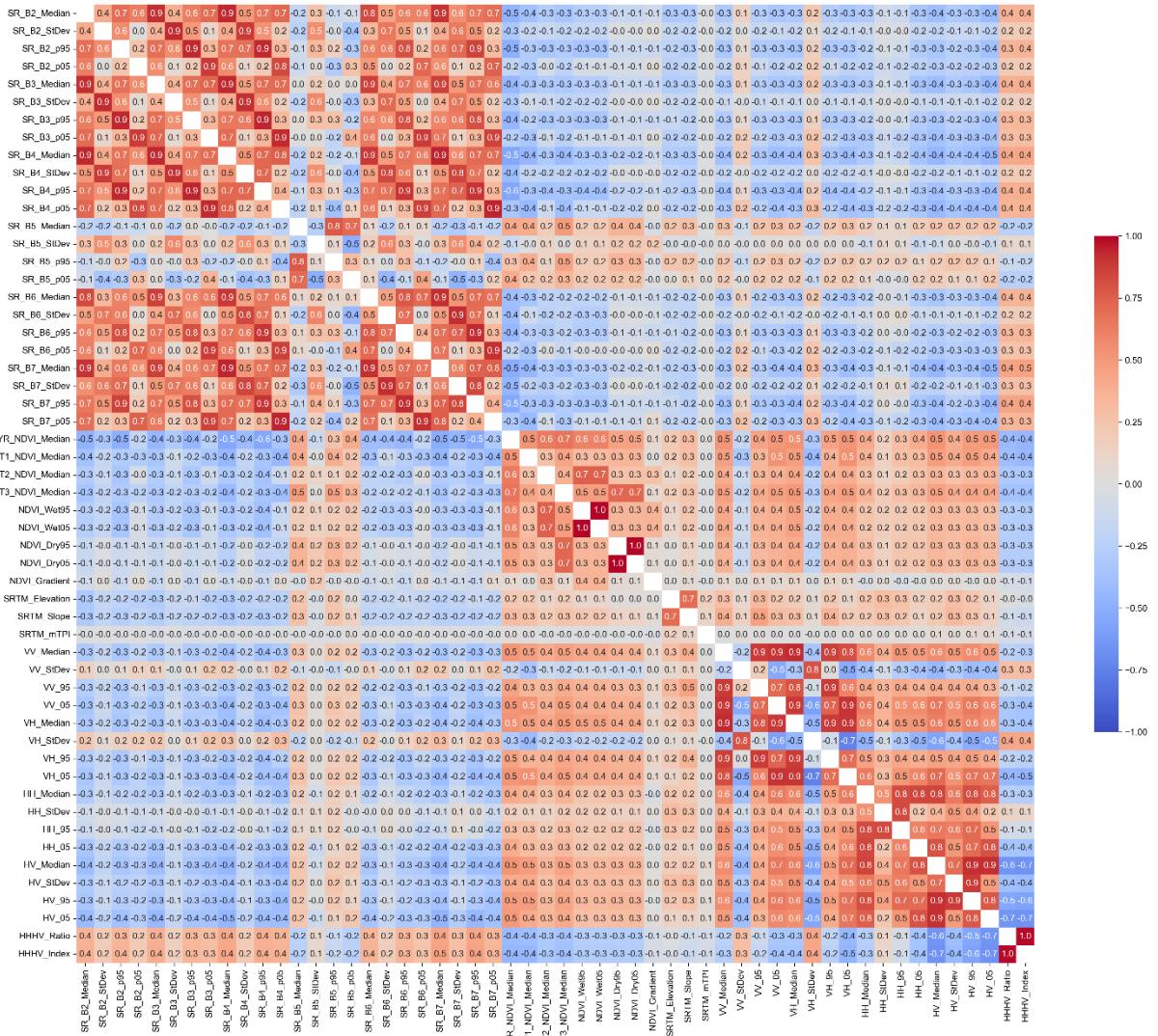


Figure 2.3: Predictor Variable Correlation Matrix. Paired correlation coefficients are shown from -1.0 (blue) to +1.0 (red), with the highest values found with the same grouping of variables by source. Authors own figure.

Table 2.6: Retained Predictor Variables.

SRTM	Landsat (Reflectance)	Landsat (Phenology)	Sentinel	Palsar
Elevation	SR_B2_Med	NDVI_Year	VV_Med	HH_Med
Slope	SR_B2_p95	NDVI_Term1	VV_p95	HH_p95
mTPI	SR_B2_p05	NDVI_Term2	VV_StDev	HH_p05
	SR_B2_StDev	NDVI_Term3	VH_p95	HH_StDev
	SR_B3_p05	NDVI_Gradient	VH_p05	HV_Med
	SR_B5_Med	NDVI_Dry_p95	VH_StDev	HV_StDev
	SR_B5_p95	NDVI_Wet_p95		HHHV_Ratio
	SR_B5_p05			
	SR_B5_StDev			
	SR_B6_Med			
	SR_B6_p95			
	SR_B6_p05			
	SR_B6_StDev			

3. Model Implementation

3.1. Parameter Configuration

To optimise the configuration of each RF model iteration, five tests were run with a different number of estimators or “trees”. The number of trees in a RF model influences how well the model can capture the complexity of the GEDI AGB estimates while controlling computational efficiency of the model algorithm. Increasing the number of trees theoretically reduces the likelihood and severity of model overfitting because RF models produce an average of each tree outcome. However, the computational demand increases significantly as more trees are configured in the model. The advantage of the reasonably sized sample of filtered GEDI AGB estimates ($n = 7,906$) made it possible to test five different model configurations with the entire sample. Five RF_A model iterations were run with intervals of 100 trees, see Table 3.1. This test showed that there was a slight improvement in performance by measure of R^2 when trees were greater than or equal to 200, but thereafter there was no evidence to suggest the number of trees improved model performance. Therefore, 200 trees were used in all subsequent models.

Table 3.1: Model Performance by Number of Trees.

Performance	100	200	300	400	500
R^2	0.25	0.27	0.26	0.27	0.26
Bias (Mg/ha)	+ 0.66	+ 0.50	+ 0.60	+ 0.57	+ 0.61
RMSE (Mg/ha)	24.89	24.62	24.69	24.66	24.69
RMSE%	57.50	56.87	57.03	56.95	57.03

3.2. Model Performances by EO Source

The main measures of model performance used in this study were R^2 , Bias, and RMSE. In **Part I: Research Paper**, Figure 3.1 shows the range of model group performances by measure of R^2 . The distribution of Bias and RMSE are shown here in Figure 3.1 and 3.2 to provide further illustration of the differences between each model group, concurring with the overall result and explanation in **Part I: Research Paper**, where the model with all 36 variables used for calibration outperforms any of the models trained with a specific grouping of predictor variables.

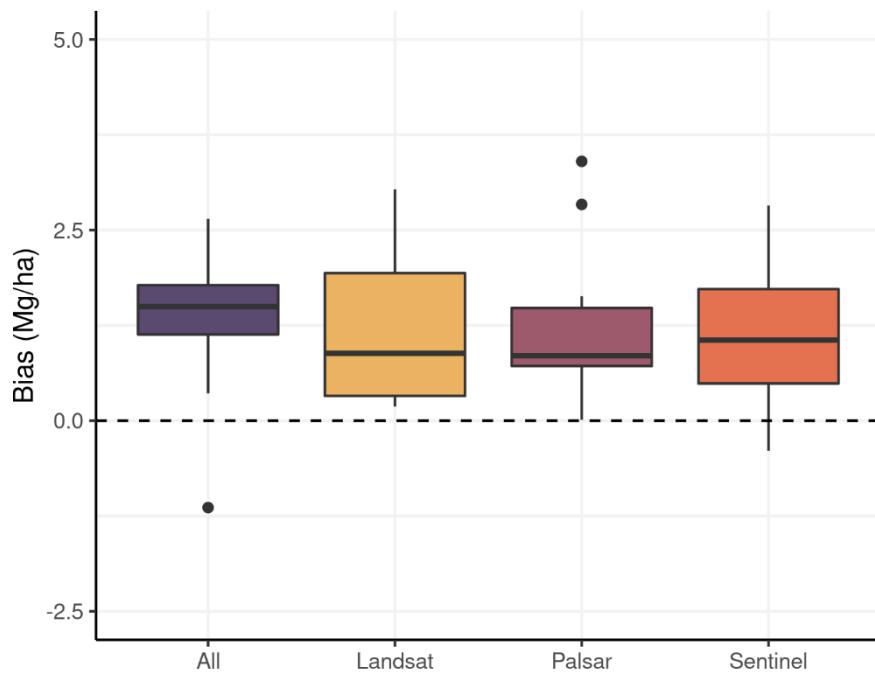


Figure 3.1: Range of Model Performances by EO Source: Bias. Bias is defined as the absolute difference of predicted GEDI AGB estimates and the random withheld subset of GEDI AGB estimates. Results concur with the pattern shown in model performance by R^2 . Authors own figure.

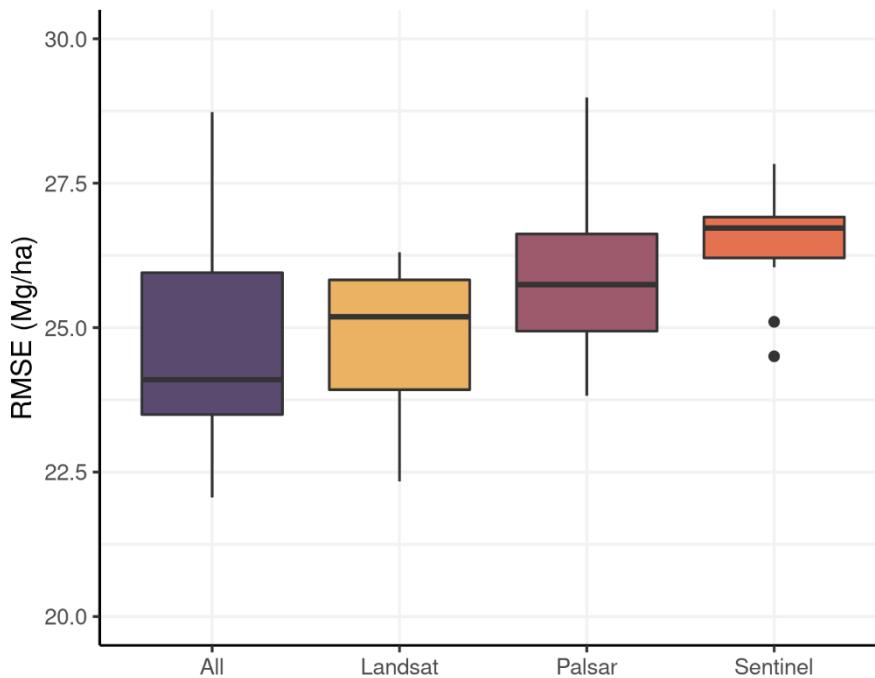


Figure 3.2: Range of Model Performances by EO Source: RMSE. RMSE is a measure of difference between predicted and observed values as with bias, though quantifies the magnitude of errors in the same units. While the lowest RMSE is shown with the inclusion of all predictor variables, improving performance more so than any one group of variables. Authors own figure.

3.3. Model Performances by Site Cross Validation

Following the assessment of model performance when site-site conditions were in place, poor spatial transferability of models became apparent, see Figure 3.3 and 3.4. Where models were calibrated and tested with subsets of data from one site, models show equal or improved performance metrics. However, where models are calibrated with all available data from one site and tested with the data from another, models show quasi-null predictive power. This analysis demonstrates the significance of spatial autocorrelation within miombo landscapes but also the stark differences between miombo landscapes, see Table 3.2, as the heterogeneity and complexity of surface properties prevent meaningful spatial transferability of predictive models.

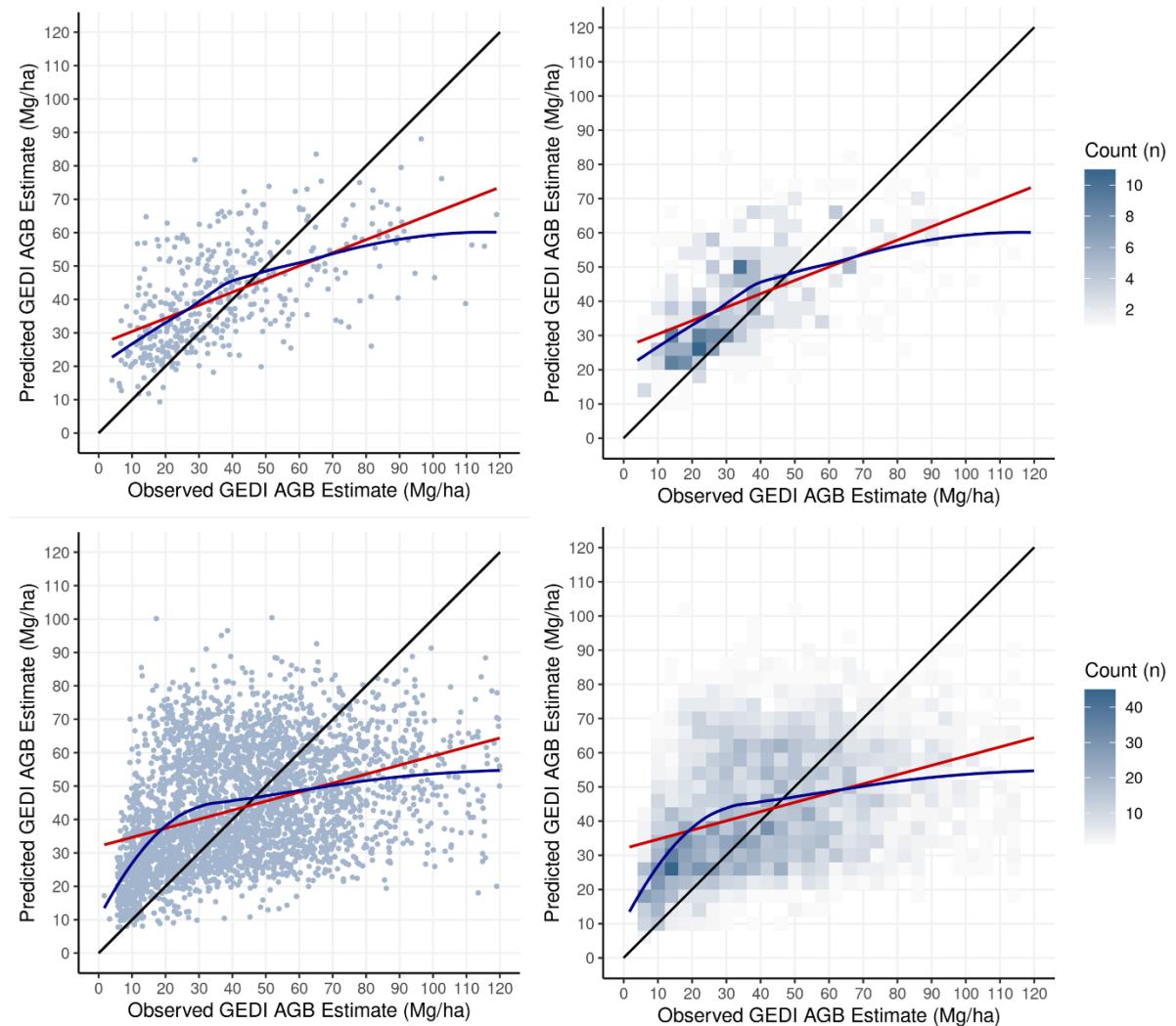


Figure 3.3: Comparison of TKW Model Performances with Spatial Cross Validation. Top row: scatter plot and two-dimensional histogram of RF_{TKW} predicted GEDI AGB estimates against a withhold subset of GEDI AGB estimates within TKW ($R^2 = 0.33$). Bottom row: scatter plot and two-dimensional histogram of RF predicted GEDI AGB estimates against GEDI AGB estimates in MGR, $\text{RF}_{\text{TKW-MGR}}$ ($R^2 = 0.12$).

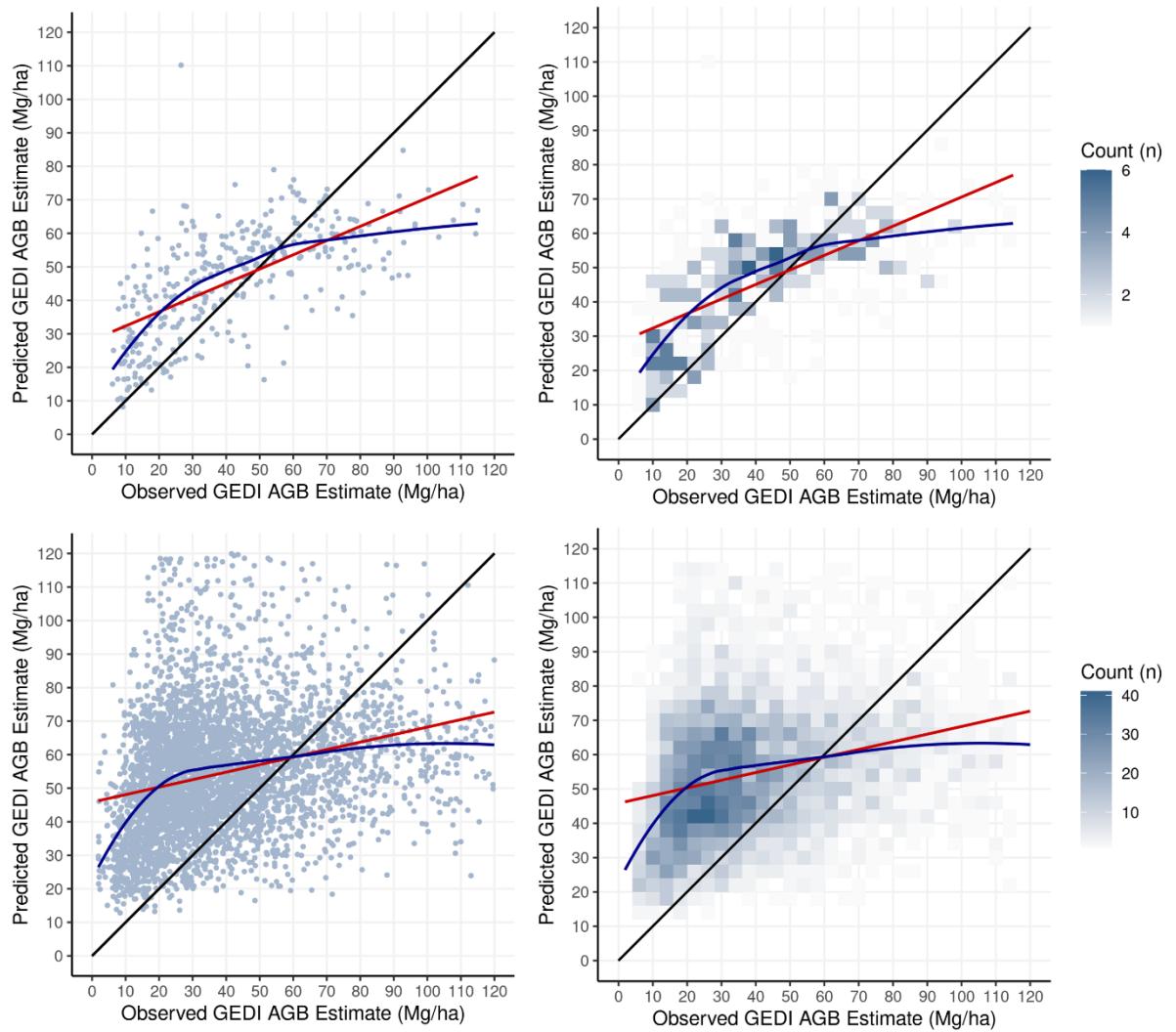


Figure 3.4: Comparison of MGR Model Performances with Spatial Cross Validation. Top row: scatter plot and two-dimensional histogram of RF_{MGR} predicted GEDI AGB estimates against a withhold subset of GEDI AGB estimates within MGR ($R^2 = 0.38$). RF_{MGR} yielded the strongest predictive performance of any RF model trained in this study. Bottom row: scatter plot and two-dimensional histogram of RF predicted GEDI AGB estimates against GEDI AGB estimates in MGR, $\text{RF}_{\text{MGR-TKW}}$ ($R^2 = 0.07$).

4. In-Situ Validation

Once RFA had been used to extrapolate GEDI AGB estimates across both sites and over the seven years, the intersections with field AGB estimates in 2017 and 2021 were isolated, see Figure 4.1. As examined in **Part I: Research Paper**, the extrapolated GEDI AGB estimates showed very little sensitivity to the field AGB estimates across both sites and over both census years. The poor sensitivity in the GEDI AGB estimates is reflected in poor performance by measure of R^2 , RMSE, and Bias across both sites and years, see Figure 4.2. Notably, MGR for 2021 yielded an arbitrarily worse performance with an $R^2 = -0.74$. A final comparison of performance metrics for the RFA model with a withheld subset of GEDI AGB estimates, and with in-situ field AGB estimates is shown in Table 4.1.

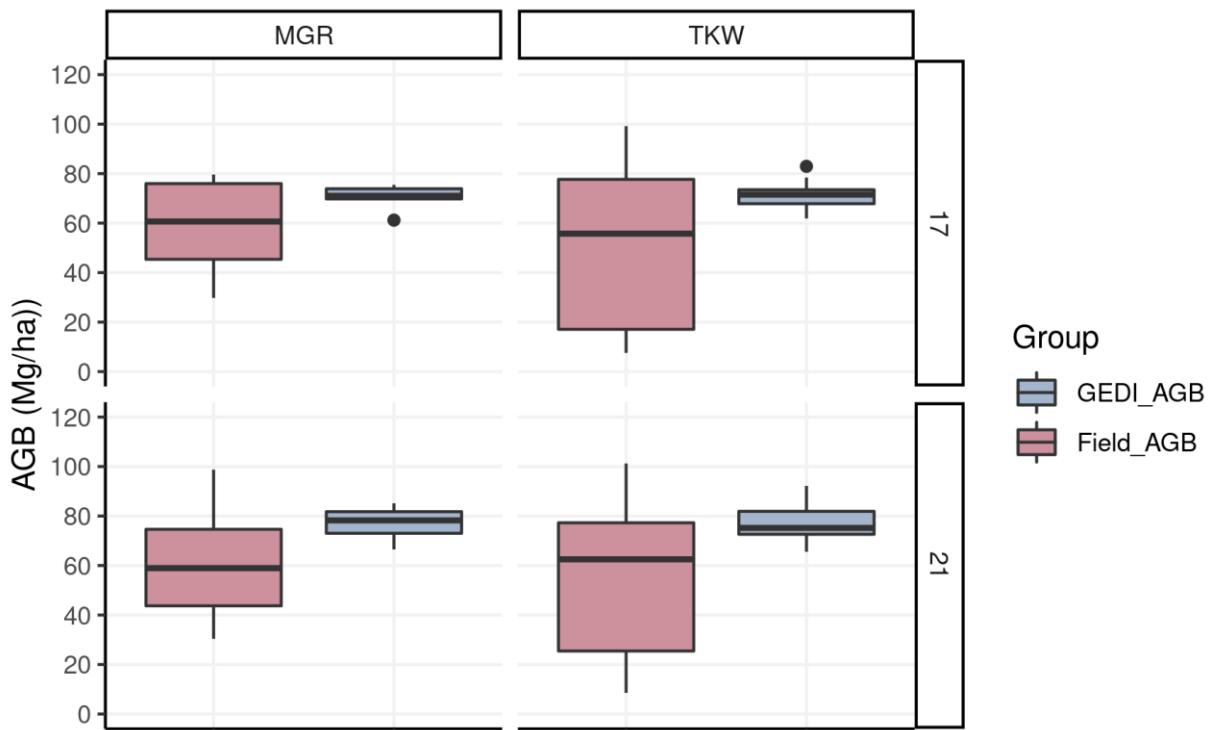


Figure 4.1: Distribution of Extrapolated GEDI AGB Estimates and Field AGB Estimates. Across both sites and over both census years, GEDI AGB estimates show very little sensitivity compared with the distribution of field AGB estimates. Authors own figure.

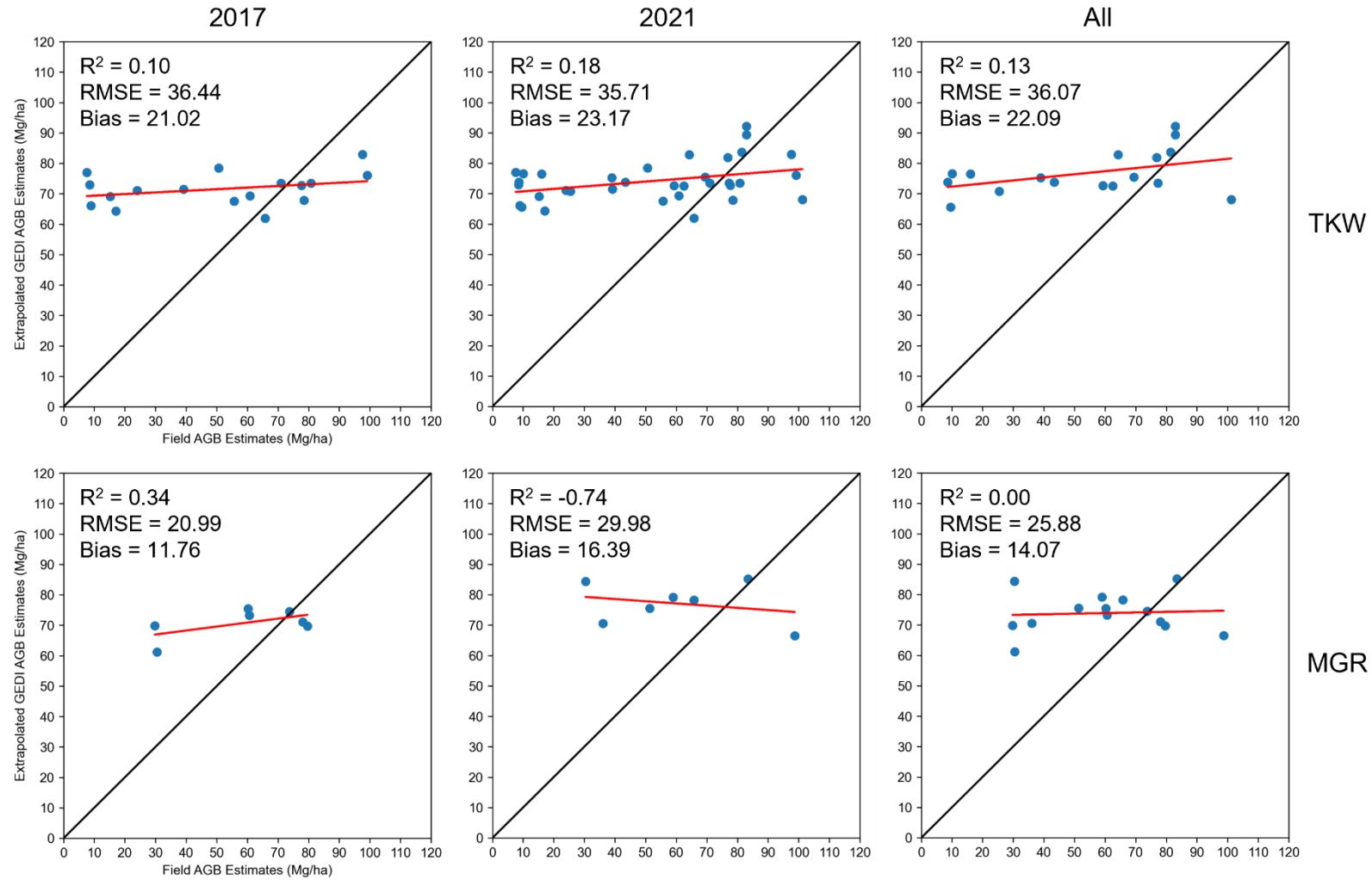


Figure 4.2: In-Situ Validation of Extrapolated GEDI AGB Estimates Across Sites, Over Census Years. Generally poor performance in all examined cases, though in one example (MGR, 2021), extrapolated GEDI AGB estimates are arbitrarily worse compared to field AGB estimates from SEOSAW PSPs. Combined, the overall performance of extrapolated GEDI AGB estimates was very poor with $R^2 = 0.09$, RMSE = 33.42 Mg/ha, and Bias = +19.75 Mg/ha. Authors own figure.

Table 4.1: Comparison of RF Model Performance and Validation.

Performance	RF _A In-Situ	RF _A Withheld	*Li et al. (<i>preprint</i>)	**Liang et al. (2023)
	Validation	Subset		
R ²	0.09	0.29	0.42	0.64
Bias (Mg/ha)	+ 19.75	+ 2.65	n/a	+ 1.06
RMSE (Mg/ha)	33.42	23.12	12.07	7.08
RMSE%	61.56	55.11	79.50	53.30

* Li et al. simulated GEDI AGB estimates with airborne LiDAR data which this study did not do, but used direct in-situ field AGB estimates for validation. Absolute bias was not presented in their preprint article.

** Liang et al. extrapolated GEDI AGB estimates alike this study, but no direct in-situ field AGB estimates were available for validation, instead performance is based on in-situ accounts of charcoal site production corroborating Landsat NDVI metrics.

5. References

- Bell, R. (1982). *The Effect of Soil Nutrient Availability on Community Structure in African Ecosystems*. In: Huntley, B. & Walker, B. (eds.) *Ecology of Tropical Savannas*. Springer: Berlin, Germany. p193-216.
- Bouvet, A.; Mermoz, S.; Toan, T.; Villard, L.; Mathieu, R.; Naidoo, L.; Asner, G. (2018). An Above-Ground Biomass Map of African Savannahs and Woodlands at 25 m Resolution Derived from ALOS PALSAR. *Remote Sensing of Environment*, 206, p156-173.
- Coops, N.; Tompalski, P.; Goodbody, T.; Queinnec, M.; Luther, J.; Bolton, D.; White, J.; Wulder, M.; Lier, O.; Hermosilla, T. (2021). Modelling Lidar-Derived Estimates of Forest Attributes Over Space and Time: A Review of Approaches and Future Trends. *Remote Sensing of Environment*, 260, p1-16.
- Cushman, K.; Armston, J.; Dubayah, R.; Duncanson, L.; Hancock, S.; Janík, D.; Král, K.; Krůček, M.; Minor, D.; Tang, H. et al. (2023). Impact of Leaf Phenology on Estimates of Aboveground Biomass Density in a Deciduous Broadleaf Forest from Simulated GEDI Lidar. *Environmental Research Letters*, 18(6), p1-10.
- Davies, R.; Ryan, C.; Harrison, R.; Dexter, K.; Ahrends, A.; Beest, M.; Benitez, L.; Brade, T.; Carreiras, J.; Druce, D. et al. (2023). Precipitation Gradients Drive High Tree Species Turnover in the Woodlands of Eastern and Southern Africa. *Ecography*, 2023(10), p1-13.
- Dubayah, R.; Blair, J.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurt, G.; Kellner, J.; Luthcke, S. et al. (2020). The Global Ecosystem Dynamics Investigation: High Resolution Laser Ranging of the Earth's Forests and Topography. *Science of Remote Sensing*, 1, p1-16.
- Duncanson, L.; Kellner, J.; Armston, J.; Dubayah, R.; Minor, D.; Hancock, S.; Healey, S.; Patterson, P.; Saarela, S.; Marselis, S. et al. (2022). Aboveground Biomass Density Models for NASA's Global Ecosystem Dynamics Investigation (GEDI) Lidar Mission. *Remote Sensing of Environment*, 270, p1-20.
- Ene, L.; Næsset, E.; Gobakken, T.; Bollandsås, O.; Mauya, E.; Zahabu, E. (2017). Large-Scale Estimation of Change in Aboveground Biomass in Miombo Woodlands using Airborne Laser Scanning and National Forest Inventory Data. *Remote Sensing of Environment*, 188, p106-117.
- Friedl, M.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. (2010). MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterisation of New Datasets. *Remote Sensing of Environment*, 114, p168-182.
- Godlee, J. (2021). *Biodiversity-Ecosystem Function Relationships in Southern African Woodlands*. PhD Thesis. University of Edinburgh.
- Godlee, J.; Ryan, C.; Bauman, D.; Bowers, S.; Carreiras, J.; Chisingui, A.; Cronsigt, J.; Druce, D.; Finckh, M. et al. (2021). Structural Diversity and Tree Density Drives Variation in the Biodiversity-Ecosystem Function Relationship of Woodlands and Savannas. *New Phytologist*, 232(2), p579-594.
- Hancock, S.; Armston, J.; Hofton, M.; Sun, X.; Tang, H.; Duncanson, L.; Kellner, J.; Dubayah, R. (2019). The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. *Earth and Space Science*, 6(2), p294-310.
- Kellner, J.; Armston, J.; Duncanson, L. (2023). Algorithm Theoretical Basis Document for GEDI Footprint Aboveground Biomass Density. *Earth and Space Science*, 10(4), p1-20.
- Lavorel, S.; McIntyre, S.; Landsberg, J.; Fores, T. (1997). Plant Functional Classifications: From General Groups to Specific Groups Based on Response to Disturbance. *Trends in Ecology & Evolution*, 12, p474-478.
- Li, X.; Wessels, K.; Armston, J.; Duncanson, L.; Urbazaev, M.; Naidoo, L.; Mathieu, R.; Main, R. (preprint). Evaluation of GEDI Footprint Level Biomass Models in Southern African Savannas using Airborne Lidar and Field Measurements. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4716466.

- Li, X.; Wessels, K.; Armston, J.; Hancock, S.; Mathieu, R.; Main, R.; Naidoo, L.; Erasmus, B.; Scholes, B. (2023). First Validation of GEDI Canopy Heights in African Savannas. *Remote Sensing of Environment*, 285, p1-17.
- Liang, M.; Duncanson, L.; Silva, J.; Sedano, F. (2023). Quantifying Aboveground Biomass Dynamics from Charcoal Degradation in Mozambique Using GEDI Lidar and Landsat. *Remote Sensing of Environment*, 284, p1-15.
- McNicol, I.; Keane, A.; Burgess, N.; Bowers, S.; Mitchard, E.; Ryan, C. (2023). Protected Areas Reduce Deforestation and Degradation and Enhance Woody Growth Across African Woodlands. *Nature Communications Earth & Environment*, 4(392), p1-14.
- McNicol, I.; Ryan, C.; Dexter, K.; Ball, S.; Williams, M. (2018a). Aboveground Carbon Storage and Its Links to Stand Structure, Tree Diversity and Floristic Composition in South-Eastern Tanzania. *Ecosystems*, 21, p740-754.
- McNicol, I.; Ryan, C.; Mitchard, E. (2018b). Carbon Losses from Deforestation and Widespread Degradation Offset by Extensive Growth in African Woodlands. *Nature Communications*, 9(3045), p1-11.
- Menaut, J.; Lepage, M.; Abbadie, L. (1995). 'Savannas, Woodlands, and Dry Forests in Africa'. In: Mooney, H. et al. *Seasonally Dry Tropical Forests*. Cambridge University Press: Cambridge, England. p64-96.
- Mitchard, E.; Saatchi, S.; Woodhouse, I.; Nangendo, G.; Ribeiro, N.; Williams, M.; Ryan, C.; Lewis, S.; Feldpausch, T.; Mier, P. (2009). Using Satellite Radar Backscatter to Predict Above-Ground Woody Biomass: A Consistent Relationship Across Four Different African Landscapes. *Geophysical Research Letters*, 36(23), p1-6.
- Mugasha, W.; Eid, T.; Bollandsås, O.; Malimbwi, R.; Chamshama, S.; Zahabu, E.; Katani, J. (2013). Allometric Models for Prediction of Aboveground and Belowground Biomass of Trees in the Miombo Woodlands of Tanzania. *Forest Ecology and Management*, 310, p87-101.
- Rosenqvist, A.; Shimada, M.; Ito, N.; Watanabe, M. (2007). ALOS PALSAR: A Pathfinder Mission for Global-Scale Monitoring of the Environment. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11), p3307-3316.
- Rosenqvist, A.; Shimada, M.; Suzuki, S.; Ohgushi, F.; Tadano, T.; Watanabe, M.; Tsuzuku, K.; Watanabe, T.; Kamijo, S.; Aoki, E. (2014). Operation Performance of the ALOS Global Systematic Acquisition Strategy and Observation Plans for ALOS-2 PALSAR-2. *Remote Sensing of Environment*, 155, p3-12.
- Ryan, C. & Williams, M. (2011). How Does Fire Intensity and Frequency Affect Miombo Woodland Tree Populations and Biomass? *Ecological Applications*, 21, p48-60.
- Ryan, C.; Hill, T.; Woollen, E.; Ghee, C.; Mitchard, E.; Cassells, G.; Grace, J.; Woodhouse, I.; Williams, M. (2011). Quantifying Small-Scale Deforestation and Forest Degradation in African Woodlands using Radar Imagery. *Global Change Biology*, 18, p243-257.
- Ryan, C.; Williams, M.; Grace, J.; Woollen, E.; Lehmann, C. (2017). Pre-Rain Green-Up is Ubiquitous Across Southern Tropical Africa: Implications for Temporal Niche Separation and Model Representation. *New Phytologist*, 213(2), p625-233.
- Ryan, C.; Williams, M.; Hill, T.; Grace, J.; Woodhouse, I. (2013). Assessing the Phenology of Southern Tropical Africa: A Comparison of Hemispherical Photography, Scatterometry, and Optical/NIR Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 52, p1-10.
- Salomonson, V.; Barnes, W.; Maymon, P.; Montgomery, H.; Ostrow, H. (1989). MODIS: Advanced Facility Instrument for Studies of the Earth as a System. *IEEE Transactions on Geoscience and Remote Sensing*, 27(2), p145-153.
- Santoro, M.; Cartus, O.; Carvalhais, N.; Rozendaal, D.; Avitabile, V.; Araza, A.; Bruin, S.; Herold, M.; Quegan, S.; Rodríguez-Veiga, P. et al. (2021). The Global Forest Above-Ground Biomass Pool for 2010 Estimated from High-Resolution Satellite Observations. *Earth System Science Data*, 13(8), p3927-3950.

- SEOSAW. (2021). A Network to Understand the Changing Socio-Ecology of the Southern African Woodlands (SEOSAW): Challenges, Benefits, and Methods. *Plants, People, Planet*, 3(3), p249-267.
- Sulla-Menashe, D. & Friel, M. (2018). User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product. Available: https://lpdaac.usgs.gov/documents/101/MCD12_User_Guide_V6.pdf. Accessed 14/07/2024.
- Tilman, D.; Knops, J.; Wedin, D.; Reich, P.; Ritchie, M.; Siemann, E. (1997). The Influence of Functional Diversity and Composition on Ecosystem Processes. *Science*, 277, p1300-1302.
- Ustin, S. & Gamon, J. (2010). Remote Sensing of Plant Functional Types. *New Phytologist*, 186(4), p795-816.
- Walker, J. (2024). *The Miombo Woodlands in Northern Zambia are the Site of a Number of Large-Scale Biofuel Investments*. Available: <https://www.cifor.org/es/knowledge/photo/36191335385/>. Accessed: 03/08/2024.
- Wessels, K.; Li, X.; Bouvet, A.; Mathieu, R.; Main, R.; Naidoo, L.; Erasmus, B.; Asner, G. (2023). Quantifying the Sensitivity of L-Band SAR to a Decade of Vegetation Structure Changes in Savannas. *Remote Sensing of Environment*, 284, p1-21.
- Wilson, G.; Bryan, J.; Cranston, K.; Kitzes, J.; Nederbragt, L.; Teal, T. (2017). Good Practices in Scientific Computing. *PLOS Computational Biology*, 13(6), p1-20.
- Woollen, E.; Ryan, C.; Williams, M. (2012). Carbon Stocks in an African Woodland Landscape: Spatial Distributions and Scales of Variation. *Ecosystems*, 15, p804-818.
- Zhang, X.; Liu, L.; Liu, Y.; Jayavelu, S.; Wang, J.; Moon, M.; Henebry, G.; Friedl, M.; Schaaf, C. (2018). Generation and Evaluation of the VIIRS Land Surface Phenology Product. *Remote Sensing of Environment*, 216, p212-229.
- Zolkos, S.; Goetz, S.; Dubayah, R. (2013). A Meta-Analysis of Terrestrial Aboveground Biomass Estimation using Lidar Remote Sensing. *Remote Sensing of Environment*, 128, p289-298.

6. Directory Structure and File Locations

The following is a directory structure wherein the data, scripts, and outputs utilised and or produced for this dissertation are stored. The **pred_vars/** folder contains the bulk of the data wherein all predictor variables for all years (2017-2023) are stored. The cumulative size of data exceeds the capacity of the allocated Scratch drive (200 GB), accessed below through a linked folder **scratch/**, resulting in partial storage of remaining data in the s1949330 user OneDrive (1 TB). Data was transferred manually and by years (2017-2019, 2020-2023) from between OneDrive and Scratch. The local **src/** folder is connected to the public GitHub repository accessible [here](#).

```
/home/s1949330/scratch/diss_data/
```

```
├── figures/
│   ├── part1_figures/
│   ├── part2_figures/
│   └── study_sites/
├── gedi/
│   ├── TKW/
│   ├── MGR/
│   └── gedi_data.csv
├── model/
│   ├── All/
│   │   └── predict/
│   │       └── validate/
│   ├── Landsat/
│   ├── Palsar/
│   ├── Sentinel/
│   ├── TKW/
│   ├── TKW-MGR/
│   ├── MGR/
│   ├── MGR-TKW/
│   └── GROUP_MODEL_STATS.csv
└── pred_vars/
    ├── TKW/
    └── MGR/
```

```
|   └── input_init/
|   └── input_merge/
|   └── input_final/
├── seosaw/
|   ├── TKW/
|   ├── MGR/
|   └── SEOSAW_AGB.csv
└── shapes/
    ├── TKW/
    ├── MGR/
    ├── SEOSAW/
    ├── national_borders/
    └── study_areas.qgz
└── src/
    ├── gee/
    ├── main/
    ├── vis/
    └── README.md
```

7. Appendix

To meet the mandatory requirements of this dissertation, the main scripts (as per Table 1.2) and supplementary scripts (for data retrieval and output visualisation) produced for this dissertation made accessible [here](#). As these scripts are not designed for constrained formatting, they are not attached in this document. Readers are recommended to view and examine all code scripts via the public GitHub repository. However, a copy of the **libraries.py** script is shown below, denoting the main python packages and libraries used throughout this study.

libraries.py

```
# Libraries for Python Scripts

import numpy as np
from pyproj import Proj, transform, Transformer
from osgeo import gdal, osr
from math import floor
from glob import glob
from tqdm import tqdm
import argparse
from pprint import pprint
import os
import pandas as pd
from itertools import zip_longest
from sklearn.model_selection import train_test_split, KFold
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score,
explained_variance_score
from sklearn.linear_model import LinearRegression
import joblib
import random
from math import sqrt
from scipy import stats
import scipy.ndimage as ndi
import statsmodels.api as sm
import time
import psutil
import matplotlib.pyplot as plt
import csv
import rasterio as rio
from rasterio.enums import Resampling
from rasterio.warp import calculate_default_transform, reproject
import seaborn as sns
from scipy.interpolate import make_interp_spline
import math
```