Remote Sensing in Earth Systems Sciences
https://doi.org/10.1007/s41976-024-00170-8

RESEARCH

Determining Optimal SAR Parameters for Quantifying Above-Ground Grass Carbon Stock in Savannah Ecosystems Using a Tree-Based Algorithm

Reneilwe Maake[1,2] · Onisimo Mutanga[2] · Johannes George Chirima[1,4] · Mahlatse Kganyago[3]

Received: 5 June 2024 / Revised: 15 November 2024 / Accepted: 28 November 2024

Abstract
The quantification and monitoring of above-ground grass carbon stock (AGGCS) will inform emission reduction policies and aid in minimising the risks associated with future climate change. This study investigated the sensitivity of Synthetic Aperture Radar (SAR)-derived parameters to predict AGGCS in a savannah ecosystem in Kruger National Park, South Africa. Particularly, we investigated the capabilities of Sentinel-1 derived parameters, including backscatter coefficients, intensity ratios, normalised radar backscatter, arithmetic computations, and the XGBoost tree-based algorithm, to predict the AGGCS. We further tested if incorporating texture matrices (i.e. Gray Level Co-Occurrence Matrix) can enhance the predictive capability of the models. We found that the linear polarisation (i.e. VV) and the intensity ratio (i.e. VH/VV) achieved similar results ($R^2 = 0.38$, RMSE% = 31%, MAE = 6.87) and ($R^2 = 0.37$, RMSE = 31%, MAE = 8.80) respectively. The Radar Vegetation Index (RVI) performed marginally (1%) better ($R^2 = 0.39$, RMSE = 30% and MAE = 6.77) compared to the other variables. Nevertheless, the incorporation texture matrix into the model enhanced prediction capability by approximately 20% ($R^2 = 0.60$, RMSE% = 20%, MAE = 3.91). Furthermore, the most influential predictors for AGGCS estimation were RVI, VH_{cor} and VV_{cor} order of importance. These findings (R^2 values of 0.35–0.39) suggest that SAR data alone does not fully capture the variability in above-ground grass carbon stock, particularly in the complexly configured savannah ecosystems. Nevertheless, the results further suggest that the prediction accuracy of SAR-based above-ground grass carbon stock models can be enhanced with the incorporation of texture matrices.

Keywords Above-ground grass carbon stock · Savannah ecosystems · Sentinel-1 · XGBoost

1 Introduction

The savannah ecosystem is one of the many vital biomes that exist on Earth. It is characterised by a rich biodiversity dominated by grasslands co-existing with patches of woodlands and shrublands [1, 2]. The grass layer serves as a feeding resource for grazing animals, while the woody layer serves as a feeding stock for browsing animals. From a climate perspective, the savannah grass and woody layers account for an estimated 25% of total gross primary production (GPP), rendering this ecosystem a vital sink of carbon [3]. They help regulate the Earth's climate and maintain a balance of greenhouse gases in the atmosphere.

The woody vegetation layer is known for storing large quantities of carbon; however, the sustainability of this capability is threatened by extreme weather conditions such as drought, heatwaves, and increased fire activity [4]. Studies have uncovered a proliferation of mortality in woody plants, which is associated with extreme weather events [5], fire [6, 7], and wildlife interactions [8, 9]. In the face of climate change, the grass layer, due to its biological configuration, is emerging as a sustainable terrestrial carbon sink [4]. The grass layer is resilient to various

✉ Reneilwe Maake
maaker@arc.agric.za

1 Agricultural Research Council - Natural Resources and Engineering, Pretoria, South Africa

2 School of Agricultural, Earth and Environmental Sciences, Geography Department, University of KwaZulu-Natal, Pietermaritzburg, South Africa

3 Department of Geography, Environmental Management and Energy Studies, University of Johannesburg, Johannesburg, South Africa

4 Department of Geography, Geoinformatics & Meteorology, University of Pretoria, Pretoria, South Africa

Published online: 12 December 2024

disturbances, including grazing, fire activity, animal stampedes, and extreme weather conditions [10]. This implies that the grass layer can sustainably maintain a relatively stable carbon stock over time, an attribute that is vital as we approach the era of above-normal climate change.

As such, the quantification and monitoring of above-ground grass carbon stock (AGGCS) will inform emission reduction policies and aid in minimising the risks associated with future climate change. This will align with the urgency emphasised in Sustainable Development Goal 13, which calls for immediate action to address climate change and lessen its repercussions. A further benefit of quantifying AGGCS is the generation of baseline information for scientific research and environmental management. The Paris Agreement is one environmental management initiative mandating countries to disclose their greenhouse gas emissions along with the reduction efforts [11]. In addition, statistics on grass carbon stock are essential for meeting reporting demands and ensuring transparency and accountability in climate actions taken by multiple nations.

A myriad of grass carbon stock quantification missions rely on space-borne sensors due to their ability to capture data over large geographical and temporal footprints [12–17]. Primarily, optical sensors have been in use for decades, regardless of their limitations. For example, optical sensors rely on the visible and infrared bands, which are obstructed by clouds and smoke, and their interaction is limited to the canopy structure of the vegetation [18]. Consequently, optical sensors are unable to ensure a consistent distribution of data throughout multiple seasons.

While SAR data has been in existence for decades, its application has been hindered by its price tag [18]. Sentinel-1 offers new opportunities for quantifying and monitoring above-ground carbon stock beyond the visible range of the electromagnetic spectrum, at no cost, and across smoky, hazy and overcast environmental conditions. Sentinel-1, constructed by the European Space Agency (ESA), is a Synthetic Aperture Radar (SAR) satellite mission [19]. The mission comprises a duo of identical sensors, namely Sentinel-1A and Sentinel-1B. These sensors can capture data under several weather conditions, an attribute that ensures a consistent flow of data, allowing for continuous quantification of carbon stock. Furthermore, Sentinel-1 can penetrate beyond the canopy structure of the vegetation, making it particularly valuable for capturing intra-canopy properties of vegetation [20]. SAR data affords several metrics which can be explored for quantifying above-ground carbon stock, some of which have been documented to be optimal for forests [21–23] and agricultural fields [24–27]. Studies showed that incorporating texture matrix with SAR data could improve biomass estimation [28–30]. However, fewer studies coupled SAR-derived parameters from sentinel-1 with texture matrix for estimating above-ground grass carbon stocks, particularly in savannah ecosystems. Employing a tree-based learning algorithm, the aim of this research is to identify the most effective SAR parameters, including texture for accurately estimating above-ground grass carbon stock in savannah ecosystems.

2 Materials and Methods

2.1 Test Site Description

Kruger National Park (KNP) is located in north-eastern South Africa, within the savannah ecosystem (Fig. 1). The park, particularly the southern part (24°97' to 25°45' S and 31°00' to 32°01' E), which is bounded by Sabi Park in the north, Marloth Park in the south, and Bushbuckridge in the west, served as an experimental site for this research. Established in 1926, the KNP is one of Africa's most renowned and iconic wildlife reserves and is approximately 20,000 km^2 in extent. The KNP has a subtropical climate with warm summers and moderate winters. The summer months, typically from October to March, bring higher temperatures and rainfall to the region. The average annual rainfall in the park ranges from 500 to 700 mm, mostly occurring during the summer months. The winters in this area are brief and dry [31] with an average annual temperature of 21 °C and maximum temperature of 25 °C [32]. The vegetation comprises dense woodlands and grassy plains [1], the latter being dominated by C4 type grass species [33]. The vegetation types present in the area include clay thorn bush, mixed bushveld, and sweet and sour lowveld bushveld [34]. The grass layer comprises various species such as *Themeda triandra, Panicum maximum, Cymbopogon* spp., and *Chlis gayana*. The geological substrate consists of granite and gabbro.

2.2 Field Above-Ground Grass Carbon Stock Sampling

A grass biomass sampling campaign was conducted from 7–14 February 2020 (Fig. 2). With the assistance of KNP's veld condition assessment (VCA) map [35], we randomly selected 25 sites from the VCA map and clustered 5 plots within/around these sites. This was done because the study site was geographically dispersed with a difficult terrain. The clustering method aids in minimising the considerable distances (3–12 km) between the VCA sites [36]. Using a Garmin Montana 650 Global Positioning System (GPS), we navigated to these centres of the selected sites and demarcated plots of 10 m × 10 m (0.01 ha) at each site. This plot size was chosen to ensure consistency with the spatial resolution of Sentinel-1. A total of 94 plots were sampled. Field measurements involved randomly throwing a quadrat of 0.5 m × 0.5 m twice within the demarcated

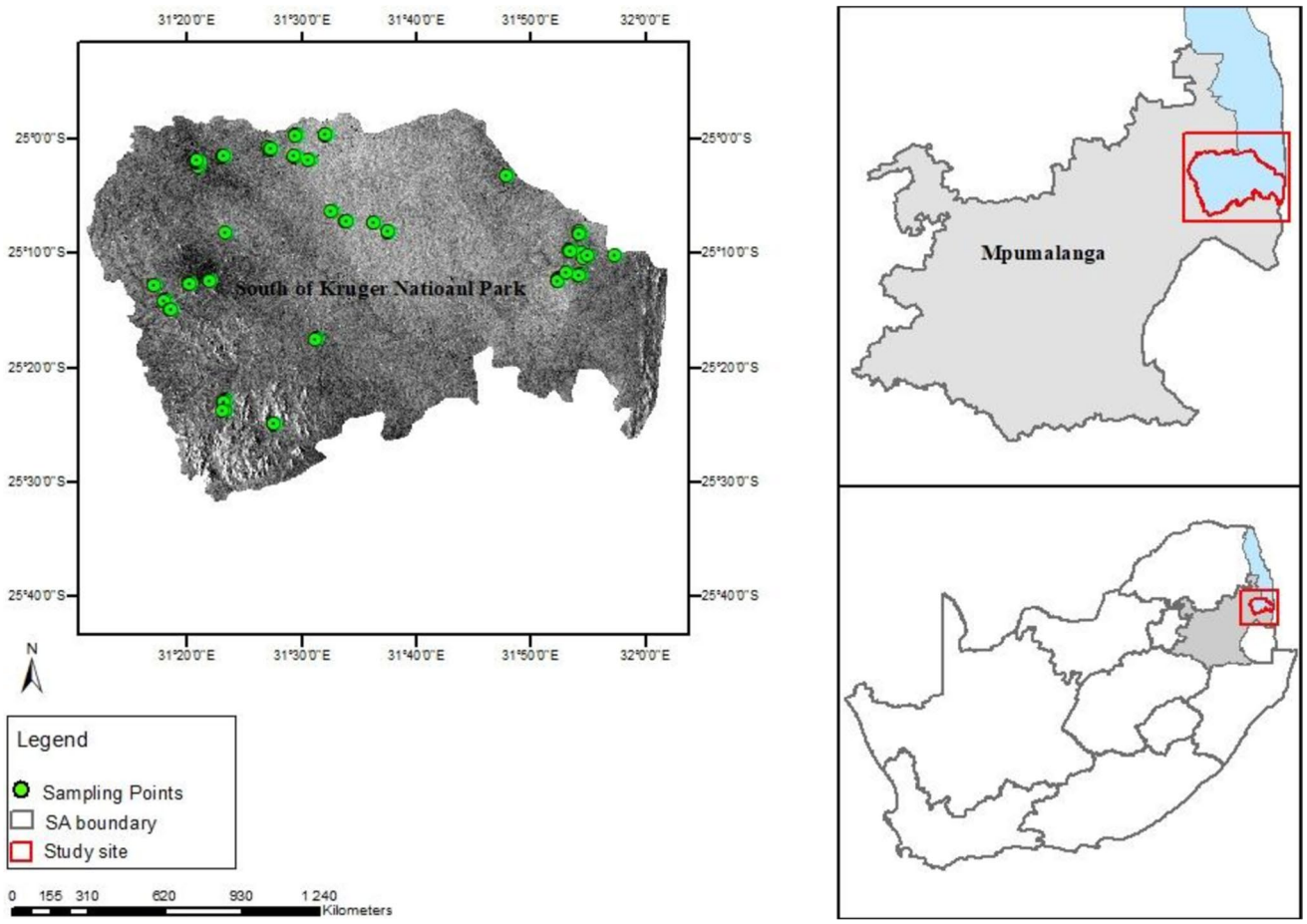

Fig. 1 Map showing the location of Kruger National Park and the test sites in the south. An inset of a Sentinel-1 satellite image is also shown

100 m^2 plot. All the grasses in each 0.5 m × 0.5 m quadrat were clipped to the ground surface, and the fresh biomass was weighed using a calibrated scale with a 0.5 g error. The samples were then taken to the laboratory where they were oven-dried at 65 °C [37] over 48 h until a constant dry biomass was obtained (Fig. 3). Then the two quadrats were averaged to represent the above-ground biomass at that plot. The oven-dried biomass was used as a dependent variable in the analysis.

2.3 Satellite Data

2.3.1 Sentinel-1 Pre-Processing

For this study, we acquired Level-1 Ground Range Detected (GRD) data from Sentinel-1 (S1) in C-band, which was captured in the Interferometric Wide (IW) Mode on February 10th, 2020. The image was extracted from the Copernicus Open Access Hub, accessible at https://scihub.copernicus.eu/dhus/. The image consists of the VV and VH channels captured on an ascending mode at a 10 m spatial resolution [38]. The pre-processing chain (Fig. 4) started with clipping the image to the dimensions of the study site to reduce the processing and computation time using the Sentinel Application Platform (SNAP) software. We then applied the orbit file to determine the satellite's position in space during the data acquisition. We also performed radiometric calibration to convert the raw digital numbers of the image into backscatter coefficients (i.e. sigma nought (σ0) values). Subsequently, the image was corrected for terrain using the Shuttle Radar Topography Mapper (SRTM) digital elevation model (DEM) of 1 s Grid (30-m) to account for the influence of terrain elevation on the backscatter values. Furthermore, we performed speckle filtering to reduce the noise on the image and enhance the image's quality for visualisation purposes. Lastly, we converted the backscatter coefficients from linear (i.e. sigma nought (σ0) values) to decibels (dB).

2.3.2 Deriving Sentinel-1 Predictor Variables

To test the sensitivity of SAR in estimating AGGCS, we extracted and calculated several parameters from the

Fig. 2 Photos of the data collection process

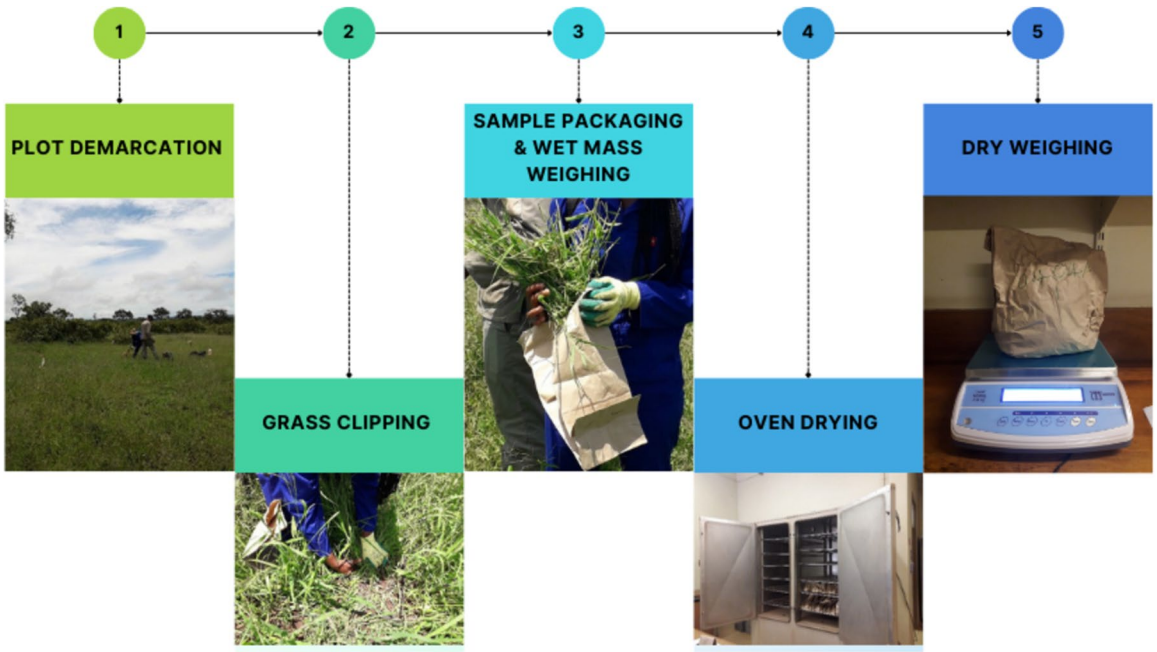

Fig. 3 A schematic depiction of the approach adopted for field data collection

Sentinel-1 image, including 2 raw backscatter coefficients, 2 polarisation intensity ratios, 1 radar indices (radar vegetation index), and 2 intensity arithmetic computations, 20 texture matrices as well as 2 geographic coordinates (Latitude and Longitude), as shown in Table 1. This amounted to 29 variables.

Table 1 Variables used in estimating above-ground grass carbon stock

Predictor variable	Expression/Description
Polarisation backscatter coefficients	VH (Vertical Transmit/horizontal receive VV (Vertical Transmit/vertical receive)
Polarisation Intensity ratios	VH/VV VV-VH/VV + VH
Radar Indices (RI)	RVI (Radar Vegetation Index)
Arithmetic Intensity Computations	VH-VV VV + VH
Texture Matrix	VH/VV_{Con} (Contrast) VH/VV_{Dis} (Dissimilarity) VH/VV_{Hom} (Homogeneity) VH/VV_{ASM} (Agular Second Moment) VH/VV_{Ene} (Energy) VH/VV_{Max} (Maximum Probability) VH/VV_{Ent} (Entropy) VH/VV_{Mea} (GLCM Mean) VH/VV_{Var} (GLCM Variance) VH/VV_{Cor} (GLCM Correlation)

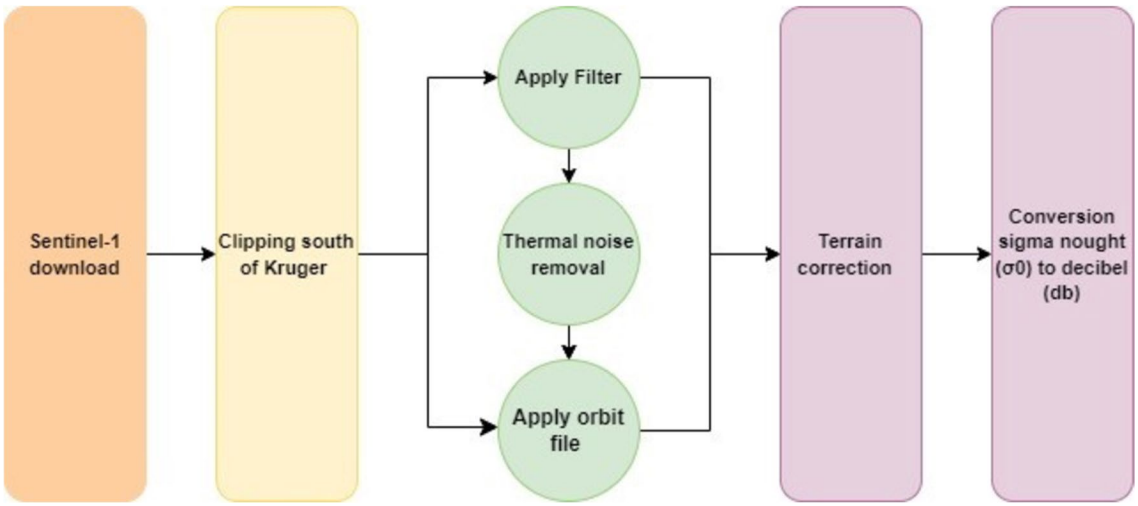

Fig. 4 A depiction of the Sentinel-1 pre-processing workflow adopted in this study

2.4 Statistical Analysis

2.4.1 Regression Algorithms

To predict AGGCS using Sentinel-1 derived variables, we employed Extreme Gradient Boosting (XGBoost) to regress field-measured grass carbon stock against Sentinel-1 derived variables. This is an enhanced version of techniques such as Gradient Boosting Machines (GBM) and AdaBoosts, configured with additional features. It is popular for its scalability, efficiency, and ability to deliver high performance on a wide range of regression and classification tasks. XGBoost adheres to the structure of gradient boosting where the model is built stage by stage, and each stage focuses on correcting the errors of the previous stage. [39]. Misclassified points receive increased weight during each iteration, directing subsequent steps to focus on rectifying these misclassifications[39]. It typically uses decision trees as base learners. These shallow trees help prevent overfitting and keep the model computationally efficient [39, 40]. Furthermore, each subsequent tree focuses on rectifying the errors introduced by its predecessors. This is achieved by training each new tree on the residuals, representing the discrepancies between the predicted and actual values of the preceding predictions [39]. It uses a weighted sum of the predictions from individual trees, where the weights are determined during training based on the performance of each tree. The optimal parameter identification process is based on a grid search method where the hyperparameters are selected based on the

lowest Root Mean Squared Error (RMSE) from a tenfold Cross Validation (CV) resampling strategy. More information about how XGBoost works can be found in [39]. The decision to choose this algorithm was informed by its performance in previous studies [41] as well as its ability to handle sparse data which is the case with our samples.

We first regressed field-measured grass carbon stocks against individual Sentinel-1 derived variables to establish their individual sensitivity to predicting AGGCS. We then incorporated texture matrices to assess their combined effect on the prediction accuracy. Incorporating the texture matrices increased the dimension of the feature space which tends to suffer from multicollinearity. Shrestha [42] defines multicollinearity as a condition wherein multiple independent variables in a regression analysis are highly correlated with each other. To deal with multicollinearity, we performed a pairwise Pearson correlation matrix to assess the relationship between SAR parameters and texture matrices. As in Chen et al. [43], we selected a threshold of $R \geq 0.8$) as an elimination criterion to select the optimal variables to develop the models. The benefits of removing redundant features from a model are demonstrated in Luo et al. [44]. Their results indicated that eliminating redundant parameters improved the accuracy of their AGB prediction models compared to when no feature parameter selection is applied.

2.4.2 Model Evaluation

To evaluate how the XGBoost algorithm performs, we utilised the tenfold CV method. This validation method enhances the reliability of assessing a model's performance by fitting the model several times using different training and testing sets each time [45], therefore helping to reduce the impact of variability in the data [46]. Furthermore, this method is valuable when working with limited samples. The predictor variables listed in Table 1 were utilised in the regression analysis. The analysis was performed at multiple experimental levels (see Table 3), based on different configurations of the predictor variables to find the optimal Sentinel-1-derived variable suitable for estimating above-ground grass biomass in savannah ecosystems.

To ensure the accuracy and reliability of the model, we computed the coefficient of determination (R^2), the root mean squared error (RMSE %) metrics, and the Mean Absolute Error (MAE) [47] to evaluate the performance of the XGBoost. R^2 ranges between 0 and 1. An R^2 value of 1 indicates that the model perfectly explains the variance in the dependent variable, while a value of 0 indicates that the model does not explain any of the variance. These metrics were calculated based on the Eqs. (1)–(4) as indicated below.

$$R^2 = 1 - \frac{\sum(Y_{measured} - Y_{predicted})^2}{\sum(Y_{measured} - Y_{predicted})^2} \quad (1)$$

$$RMSE = \frac{\sqrt{\sum_{i-1}^{N} - (Y_{measured} - Y_{predicted})^2}}{N} \quad (2)$$

$$RMSE\% = \frac{\sqrt{\frac{1}{n}\sum_{i-N}^{N} - (Y_{measured} - Y_{predicted})^2}}{\overline{Y}} \times 100 \quad (3)$$

$$MAE = \sum_{n}^{1} \frac{\left|Y_{measured} - Y_{predicted}\right|}{n} \quad (4)$$

where Y_{measured} is the observed AGGCS, $Y_{\text{predicted}}$ is the predicted AGGCS, N is number of observed samples, i is the included predictor variable, and $\overline{Y}$ is the mean of the observed AGGCS. The R^2, RMSE, and RMSE% equations are adopted from [48], while the MAE is adopted from [49].

3 Results

3.1 Exploratory Analysis of Collected Above-Ground Grass Carbon Stock Samples

Descriptive statistics of the above-ground grass carbon stock in the south of Kruger National Park are presented in Table 2. The AGGCS observed ranged between 4.5 and 58.2 g per square metre (g m^{-2}), the maximum was 58.2 g m^{-2} and the average was 28.7 g m^{-2}. The standard deviation, a measure of the variability or spread of the data points around the mean, was 11.89 g m^{-2}.

3.2 Predictor Selection for Above-Ground Grass Carbon Stock Prediction

A matrix showing the correlation between of Sentinel-1 derived parameters and texture matrices is presented in Fig. 5. As indicated on Fig. 5 of the 29 variables, 25 were highly correlated ($R = \geq 0.8$). The variables that were not

Table 2 Descriptive statistics of measured above-ground grass carbon stock in Kruger National Park

Variables	No of samples	Min	Max	Mean	StDev[1]
Dry grass carbon stock (g m^{-2})	94	4.5	58.2	28.73	11.89

[1]StDev = Standard Deviation

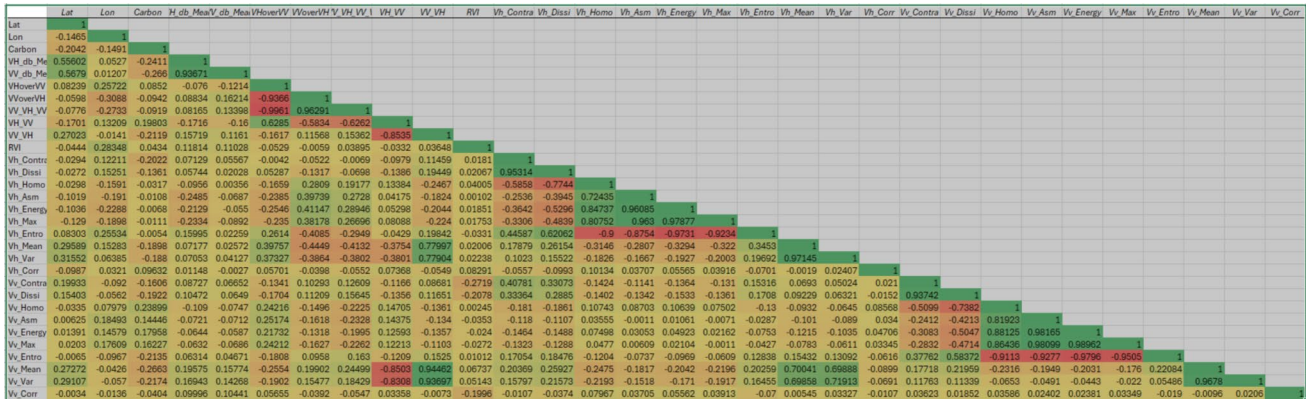

Fig. 5 A correlation matrix of Sentinel-1 derived parameters and texture matrices

Table 3 Summary of combinations for predicting above-ground grass carbon stock in Kruger National Park

Experiment	Matrices	R^2	RMSE% (g m^{-2})	MAE
1	Linear polarisation (LI)			
	Intensity (VH)	0.38	33.22	6.86
	Intensity (VV)	0.38	31	6.87
2	Radar Vegetation Index (RVI)			
	RVI	0.39	30	6.77
3	Intensity ratios (IR)			
	VH/VV	0.37	31.03	6.9
	VV-VH/VV + VH	0.37	31.11	6.9
4	Arithmetic Computations (AC)			
	VH-VV	0.36	30.84	6.73
	VV + VH	0.37	31.31	6.84
5		All variables		
	All optimal S1 variables	0.36	30.84	7.02
6	Texture Matrices			
	VH	0. 40	30	5.32
	VV	0. 39	30	5.21
		Optimal S1 variable + Texture		
7		0.60	20	3.97

highly correlated included: RVI, VH_{Cor}, and VV_{Cor}. These variables were used to predict the AGGCS.

3.3 Predictive Performance of Parameters

The predictive capabilities of Sentinel-1 derived parameters coupled with texture matrices and the Extreme Gradient Boosting algorithm from the seven levels of analysis are summarised in Table 3. The linear polarisation (i.e. VV) and the intensity ratio (i.e. VH/VV) achieved similar results ($R^2 = 0.38$, RMSE% = 31, MAE = 6.87) and ($R^2 = 0.37$, RMSE% = 31, MAE = 6.9) respectively. The Radar Vegetation Index (RVI) performed marginally (≤ 1%) better ($R^2 = 0.39$, MAE = 6.77, RMSE% = 30) compared to the other SAR predictors. Arithmetic computations between VH and VV intensities also yielded moderate R^2 values between 0.36 and 0.37. Meanwhile, combining all SAR variables results in a slightly lower R^2 of 0.36 compared to the individual experiments. Furthermore, an experiment based on texture matrices alone performed within the same range as SAR parameters for both VH and VV ($R^2 = 0.39$–40, MAE = 5.32–5.21, RMSE% = 30) respectively. Nonetheless, integrating texture matrix with SAR parameters improved the modelling accuracy ($R^2 = 0.60$, RMSE% = 20, MAE = 3.97) (Fig. 6).

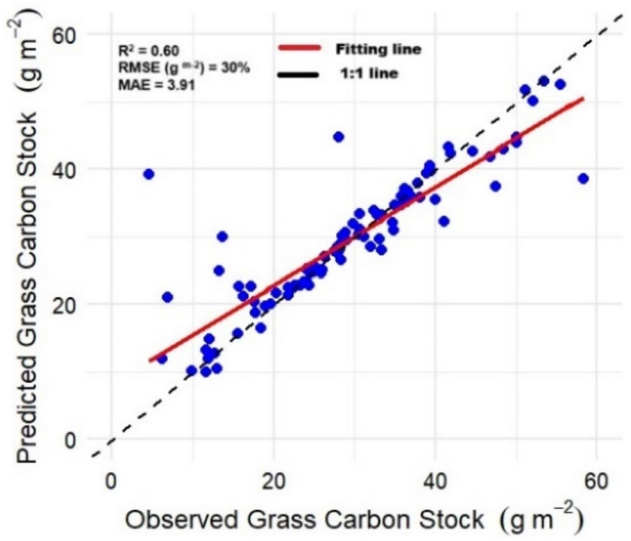

Fig. 6 Scatterplot showing the relationship between observed and predicted above-ground grass carbon stocks based on the XGBoost model integrating texture matrices

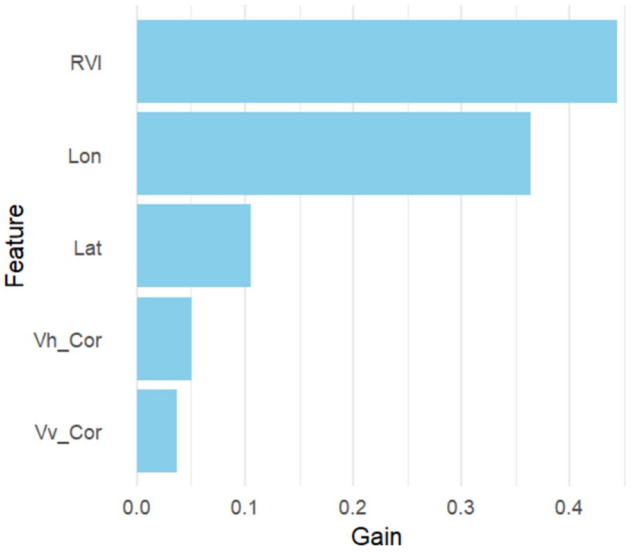

Fig. 7 Optimal parameters selected for predicting above-ground grass carbon stock

3.4 Important Variable Selection

Figure 7 illustrates the influential variables selected for predicting AGGCS as measured by their Gain. The Gain metric represents the contribution of each feature to the model's predictive power, with higher values indicating greater importance in making accurate predictions. Results from the correlation matrix show that of all the 30 variables calculated, only five had a correlation below the set, threshold (i.e. ≥ 0.8). The RVI was the most influential from the SAR parameters whereas the VH_{cor} and VV_{cor} were the least influential from the texture matrices. Overall, the combined input of this parameters optimally predicted AGGCS (Fig. 6).

3.5 Comparison of Predicted vs Observed Above-Ground Grass Carbon Stock

Figure 8 compares the number of sample points (frequency) of the measured and predicted AGGCS. The predicted AGGCS for most classes (i.e. < 25 g m^{-2}, 31–35 g m^{-2}, 36–40 g m^{-2}, and > 40 g m^{-2}) are close to the observed values. An exception is the 26–30 g m^{-2} range which shows a huge underestimate compared to the observed values. This overestimation is more pronounced in the lowest AGGCS class range. The model seems to have obtained a relatively lower error in predicting AGGCS in the 36–40 g m^{-2} class range.

4 Discussion

Precise estimation of climate-regulating ecosystem services in pristine savannah ecosystems such as Kruger National Park is essential for monitoring their contribution to climate regulation and assessing their potential as carbon sinks. This study derived several SAR parameters coupled with texture matrix for quantifying above-ground grass carbon stock in savannah ecosystems using a tree-based algorithm.

4.1 Predictive Performance of Parameters

In this study, we incorporated a range of SAR-derived parameters, raw backscatter coefficients (VH, VV) including polarisation intensity ratios (VH/VV, VV-VH/VV + VH), radar indices (RVI), and intensity arithmetic computations (VH-VV, VV + VH), with the expectation that they would optimally predict AGGCS. Particularly, the VH/VV Ratio and RVI was anticipated to capture the variability of AGGCS because it separates the cross-polarized (VH) and co-polarized (VV) backscatter signals, which can reflect vegetation density and structure. High biomass areas often exhibit stronger VH signals due to increased volume scattering [50] while VV signals can indicate structural orientation in vegetation [51]. The RVI considers various polarization channels (e.g. VH, VV); therefore, combining information from both co-polarized and cross-polarized returns captures additional insights regarding vegetation structure, density, and volume, as well as water content [25] all of which are closely linked to biomass. However, the predictive power of VH/VV was

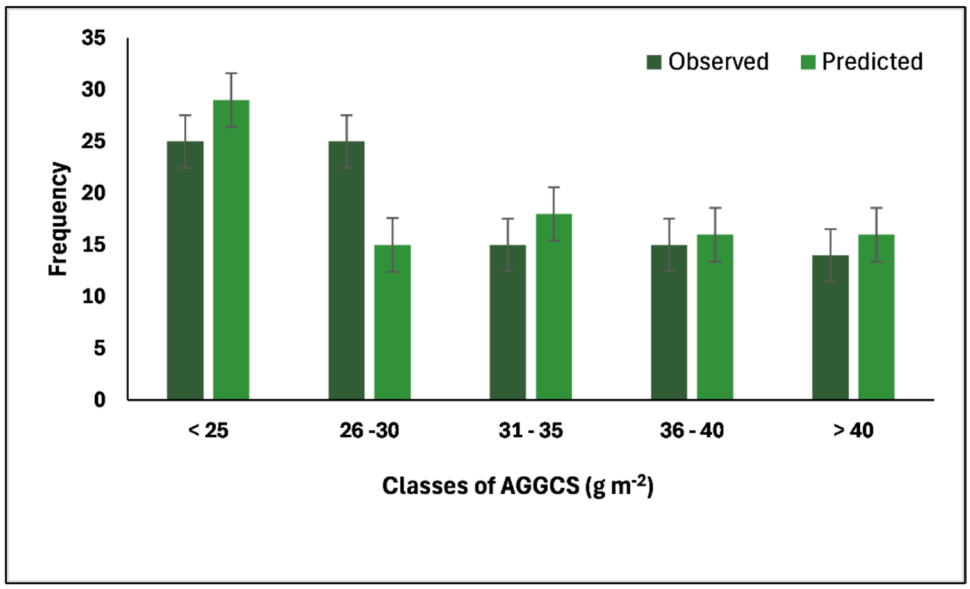

Fig. 8 Histogram of measured and predicted above-ground grass carbon stock. Frequency refers to the number of observations per AGGCS range for each category

limited, potentially due to the sparse vegetation in savannah environments, which reduces volume scattering and minimizes the distinction between VH and VV signals. Furthermore, the sparse vegetation is often associated with bare soil background in savannahs [52], which may affect backscatter signals where a larger portion of the signal return will emanate from soil that grasses. These traits differ from forested ecosystems, where high soil moisture and organic content often amplify the vegetation signal in SAR data [50].

Similar results were obtained in previous studies, for example [53, 54] reported very weak correlation ($R^2 = 0.01–0.05$) when biomass reached a high level. They attributed this weak correlation to rainfall. In this study, AGGCS measurements were done during the wet season. The conditions at the test site at the time of data collection were such that the grasses could have been wet. Generally, wet vegetation tends to exhibit higher backscatter due to increased attenuation and scattering effects. Furthermore, during the rainy season, soil moisture levels typically increase significantly due to precipitation, and this affects the radar signal as it is reported to be sensitive to surface moisture [55]. Water on leaves after rainfall can cause substantial change in SAR backscattering coefficient [56]. Nasirzadehdizaji et al. [57] also reported weaker correlations ($R^2 = -0.49$) for agricultural fields where the density of the crops was above 70%. Further evidence is provided in Rapiya et al. [58] who observed an unsatisfactory correlation between above-ground biomass (AGB) in natural rangeland using SAR data and optical data. They reported that fluctuation in AGB is most pronounced during periods of peak productivity, with the highest levels observed in late summer, followed by early summer, and gradually decreasing during winter (the cold, dry season). We suggest that future studies factor in these topo-climatic parameters with SAR parameters to better characterise the variability of the AGGCS across different landscapes. However, most available Rainfall data (Climate Hazards Group Infrared Precipitation with Station data (CHIRPS) [59] are too coarse (i.e. 1km resolution) for small-scale mapping. Studies can explore locally available in situ weather station and soil moisture data for mapping at higher spatial resolutions, 10 m in the case of our study.

Beriaux et al. [60] discovered that the correlation between SAR polarimetric parameters and biomass was relatively weak due to the complex interactions between radar signals and forest structure. In this study, AGGCS measurements were done at the peak period where grass growth was at maximum. Compared to dormant grasses, grasses at peak growing stages tend to have taller and denser canopies, which may have an impact on SAR signal penetration depth and scattering processes, and ultimately the accuracy of the AGGCS estimation. Further evidence is provided in Rapiya et al. [58] who observed an unsatisfactory correlation between above-ground biomass (AGB) in natural rangeland using SAR data and optical data. They reported that fluctuation in AGB is most pronounced during periods of peak productivity, with the highest levels observed in late summer, followed by early summer, and gradually decreasing during winter. The transitioning of grasses from one stage to another (i.e. germination, growth, maturity and senescence) may result in variations in vegetation structure. Therefore, acquiring SAR data at different stages of grass growth may capture these temporal variations and provide valuable information for biomass estimation. In addition, savannah ecosystems are complex in nature, consisting of a mixture of grasslands and scattered trees that coexist. The vertical structure and density of grasses and trees may significantly affect SAR signals.

Wang et al. [61] also reported a moderate correlation between SAR polarimetric parameters and biomass and attributed these findings to factors such as soil moisture. In this study, SAR and field data acquisitions took place in the rainy season. The field conditions at the study site at the time of data collection were such that the soil could have been wet. During the rainy season, soil moisture levels typically increase significantly due to precipitation, and this may have an effect on the radar signals, as it is reported to be sensitive to surface moisture [62]. We suggest that future studies factor in these soil moisture with SAR parameters to improve the predictive capabilities of the AGGCS models. However, most available Soil moisture products such as the Soil Moisture Active Passive (SMAP) [63], Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) Cawse-Nicholson and Anderson [64] and the International Soil Moisture Network [65] are usually too coarse for localised studies. Studies can explore locally available in-situ soil moisture data for mapping at localised areas.

Generally, if an increase in AGGCS values does not result in a comparable increase in the SAR backscatter values, this implies saturation [66]. In a study focusing on forests, Sinha et al. [67] observed that the C-band generally saturates at about 70 t ha^{-1} for tropical forests. In our study, signs of signal saturation were also observed, particularly at 40 g m^{-2}. Nevertheless, actives sensor with high spatial resolution in the L and P spectral bands have been missioned to offer enhanced image contracts due to their multi-looking trait [66]. These sensors can characterize intra-canopy traits and as a result they are able to deal with saturation [29]. However, in many resources constrained savannah regions, such Africa, access to commercial SAR data such as ALOS PALSAR comes at a high cost.

Although the models based solely on SAR parameters were the least effective, when combined with the texture matrices, the prediction capability was enhanced. This may be because of the grey-scale properties of images and the spatial position of image pixels, which makes it possible to combine them with backscatter variables to reduce the underestimation or overestimation. Image textures have been reported to be sensitive to brightness variations arising from canopy structure of vegetation [68]. These findings agree with [29, 43] who reported improved accuracies when incorporating texture variables. Integrating SAR-derived parameters with texture matrices offers several complementary benefits including leverage on both biophysical and spatial attributes. For example, when SAR parameters explain the structural properties of grasses, texture matrices capture the spatial orientation of these properties. Therefore, providing a composite of variables to account most of the variabilities in the landscape, which can improve the accuracy of quantifying AGGCS.

Overall, this study underscores the challenges encountered when utilising single source data to estimate AGGCS, especially in complex savannah ecosystems. Nevertheless, the study also highlights potential benefit of integrating multiple parameter from different sources to quantify AGGCS in complex landscapes such as savannahs.

4.2 Limitations of the Study

From a biological and environmental perspective, several factors including precipitation and temperature, along with other environmental factors such as soil properties and topography, influence the suitability of habitats for different grass species. This, in turn, influences the distribution and abundance of AGGCS across different landscapes. For instance, Liu et al. [69] reported that precipitation and temperature are the most essential climate factors influencing the spatial variability of above-ground biomass. In this study, attempts were made to incorporate soil moisture and rainfall from existing products such as Soil Moisture Active Passive (SMAP) [63] and Rainfall from the Climate Hazards Group Infrared Precipitation with Station data (CHIRPS). However, the resolution (i.e. 9 km and 1 km respectively) of these data was too coarse for the scale of our mapping (i.e. 10 m).

Extrapolating predictions from remote sensing-based AGGCS models is a common and challenging practice in remote sensing, yet essential for mapping at larger scales (regional, landscape, national). This uncertainties between the measured and predicted AGGCS observed in this study can be attributed to the mismatch in the data scale. For example, the observed data was extrapolated from a quadrat scale (0.5 m × 0. 5 m) to the scale of Sentinel-1 (10 m × 10 m), and ultimately to the entire study site. The transition between these two scales introduces some level of error and therefore may have affected the accuracy thereof. Furthermore, the number of plots measured during fieldwork (i.e. two), may have been insufficient to capture the variability of AGGCS in the study site. A solution would have been to make several measures of the same grasses and average such measurement values.

Apart from interferences from the woody layer, the grass layer in the study area was characterized by a mix of tall and short grasses which may have affected the SAR backscatter. The heterogeneity in canopy cover and the presence of low biomass grasses may have resulted in lower backscatter signals. This variability within the savannah landscape introduces uncertainty in SAR-based AGGCS predictions, as SAR sensors may struggle to detect small changes in sparse biomass areas. Another source of uncertainty was sampling far from the lab for wet weight. As shown in Fig. 1, the sampling sites were far apart resulting in increased commuting time from one site to the other and ultimately to the lab. Future studies that may find it impractical to immediately transport wet samples to the lab due to long distances,

should explore portable field-weighing tools, such as hanging scales with zero adjuster for measuring the weight of grasses. However, field drying requires additional equipment, such as ovens or dryers, which may not be feasible in remote or off-grid location. Nevertheless, the combined use of SAR parameters and texture matrices were able to quantify AGGCS with an acceptable accuracy.

5 Conclusion

This study examined the prospect of Sentinel-1-derived matrixes coupled with texture matrixes and the XGBoost algorithm for predicting above-ground grass carbon stock in savannah ecosystems. Based on the findings, this study concludes that when utilised separately, SAR parameters and texture matrices tends to underperform ($R^2 = 0.36–0.40$, RMSE = 5.21–6.9). However, the combined use of SAR parameters and texture matrices results with an R^2 of 0.39 and an RMSE of 8.75, suggesting that texture matrices may improve the predictive ability of SAR data for AGGCS estimation. The variables that optimally captured the variability of AGGCS includes RVI, VH_{cor}, and $VV_{cor\ as}$ well as the latitude and longitude. Overall, the integration of Sentinel-1-derived matrixes coupled with texture matrixes and the XGBoost algorithm shows potential in predicting above-ground grass carbon stock in savannah ecosystems.

Nevertheless, as mandated by the Kyoto Protocol for Reducing Emissions from Deforestation and Forest Degradation (REDD +), the findings of this study are essential for providing insight into how conservation efforts contribute to the global carbon balance and the possibility of regulating climate change. Furthermore, adopting this remote-sensing approach for estimating AGGCS will ensure regular and up-to-date statistics on grass carbon stock that are essential for meeting reporting demands and ensuring transparency and accountability in climate actions taken by multiple nations.

Acknowledgements The authors express their gratitude to the Institute for Natural Resource and Engineering at the Agricultural Research Council for supplying the hardware and software utilized in this study. Special thanks are extended to Dr Thomas Fyfield for his editorial assistance, Sabelo Mazibuko, Kgaugelo Mogano, Phathutshedzo Ratshiedana Makgethwa Masemola and Sibongiseni Mgolozeli for assisting with data collection.

Author Contribution "Reneilwe Maake. wrote the main manuscript text. Mahlatse Kganyago. wrote the R scripts and codes used for the analysis. Johannes George Chirima. and Onisimo Mutanga. acquired the funding. All authors reviewed the manuscript."

Funding Open access funding provided by Agricultural Research Council. This research was funded by the Agricultural Research Council-Natural Resources and Engineering project: ISC012403000027 and the National Research Foundation (NRF) Research Chair in Land Use Planning and Management (Grant number: 84157).

Data Availability The remote sensing data supporting the results reported in our paper can be accessed at https://scihub.copernicus.eu/dhus/. The original data generated in our research is available upon reasonable request.

Declarations

Competing Interests The authors declare no competing interests.

References

1. Higgins SI, Bond WJ, February EC, Bronn A, Euston-Brown DI, Enslin B, Govender N, Rademan L, O'Regan S, Potgieter AL (2007) Effects of four decades of fire manipulation on woody vegetation structure in savanna. Ecology 88(5):1119–1125
2. Scholes RJ, Archer SR (1997) Tree-grass interactions in savannas. Annu Rev Ecol Syst 28(1):517–544
3. Moore CE, Beringer J, Evans B, Hutley LB, McHugh I, Tapper NJ (2016) The contribution of trees and grasses to productivity of an Australian tropical savanna. Biogeosciences 13(8):2387–2403
4. Dass P, Houlton BZ, Wang Y, Warlind D (2018) Grasslands may be more reliable carbon sinks than forests in California. Environ Res Lett 13(7):074027
5. Allen CD, Breshears DD, McDowell NG (2015) On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. Ecosphere 6(8):1–55
6. Stevens-Rumann CS, Kemp KB, Higuera PE, Harvey BJ, Rother MT, Donato DC, Morgan P, Veblen TT (2018) Evidence for declining forest resilience to wildfires under climate change. Ecol Lett 21(2):243–252
7. McDowell NG, Sapes G, Pivovaroff A, Adams HD, Allen CD, Anderegg WR, Arend M, Breshears DD, Brodribb T, Choat B (2022) Mechanisms of woody-plant mortality under rising drought, CO2 and vapour pressure deficit. Nature Reviews Earth & Environment 3(5):294–308
8. Das AA, Thaker M, Coetsee C, Slotow R, Vanak AT (2022) "The importance of history in understanding large tree mortality in African savannas." Ecography 2022(1). https://doi.org/10.1111/ecog.06012
9. Shannon G, Thaker M, Vanak AT, Page BR, Grant R, Slotow R (2011) Relative impacts of elephant and fire on large trees in a savanna ecosystem. Ecosystems 14:1372–1381
10. Craine JM, Ocheltree TW, Nippert JB, Towne EG, Skibbe AM, Kembel SW, Fargione JE (2013) Global diversity of drought tolerance and grassland climate-change resilience. Nat Clim Chang 3(1):63–67

11. Walsh B, Ciais P, Janssens IA, Penuelas J, Riahi K, Rydzak F, Van Vuuren DP, Obersteiner M (2017) Pathways for balancing CO2 emissions and sinks. Nat Commun 8(1):14856
12. Abbas S, Wong MS, Wu J, Shahzad N, Muhammad Irteza S (2020) Approaches of satellite remote sensing for the assessment of above-ground biomass across tropical forests: Pan-tropical to national scales. Remote Sensing 12(20):3351
13. Bindu G, Rajan P, Jishnu E, Joseph KA (2020) Carbon stock assessment of mangroves using remote sensing and geographic information system. The Egyptian Journal of Remote Sensing and Space Science 23(1):1–9
14. Chapungu L, Nhamo L, Gatti RC (2020) Estimating biomass of savanna grasslands as a proxy of carbon stock using multispectral remote sensing. Remote Sensing Applications: Society and Environment 17:100275
15. Clementini C, Pomente A, Latini D, Kanamaru H, Vuolo MR, Heureux A, Fujisawa M, Schiavon G, Del Frate F (2020) Long-term grass biomass estimation of pastures from satellite data. Remote Sensing 12(13):2160
16. Ding L, Li Z, Wang X, Yan R, Shen B, Chen B, Xin X (2019) Estimating grassland carbon stocks in Hulunber China, using Landsat8 oli imagery and regression kriging. Sensors 19(24):5374
17. Guo L, Sun X, Fu P, Shi T, Dang L, Chen Y, Linderman M, Zhang G, Zhang Y, Jiang Q (2021) Mapping soil organic carbon stock by hyperspectral and time-series multispectral remote sensing images in low-relief agricultural areas. Geoderma 398:115118
18. Abdi AM, Brandt M, Abel C, Fensholt R (2022) Satellite remote sensing of savannas: Current status and emerging opportunities. Journal of Remote Sensing 2022:1–20
19. Potin P, Rosich B, Miranda N, Grimont P, Shurmer I, O'Connell A, Krassenburg M, Gratadour J-B (2018) Sentinel-1 constellation mission operations status. IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, IEEE. https://doi.org/10.1109/IGARSS.2018.8517743
20. Ghasemi N, Sahebi MR, Mohammadzadeh A (2011) A review on biomass estimation methods using synthetic aperture radar data. International Journal of Geomatics and Geosciences 1(4):776–788
21. Ghosh SM, Behera MD (2021) Aboveground biomass estimates of tropical mangrove forest using Sentinel-1 SAR coherence data-The superiority of deep learning over a semi-empirical model. Comput Geosci 150:104737
22. Vatandaşlar C, Abdikan S (2022) Carbon stock estimation by dual-polarized synthetic aperture radar (SAR) and forest inventory data in a Mediterranean forest landscape. Journal of Forestry Research 33(3):827–838
23. Zeng P, Zhang W, Li Y, Shi J, Wang Z (2022) Forest total and component above-ground biomass (AGB) estimation through C-and L-band polarimetric SAR data. Forests 13(3):442
24. Hajj ME, Baghdadi N, Belaud G, Zribi M, Cheviron B, Courault D, Hagolle O, Charron F (2014) Irrigated grassland monitoring using a time series of terraSAR-X and COSMO-skyMed X-Band SAR Data. Remote Sensing 6(10):10002–10032
25. Mandal D, Kumar V, Ratha D, Dey S, Bhattacharya A, Lopez-Sanchez JM, McNairn H, Rao YS (2020) Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. Remote Sens Environ 247:111954
26. Nguyen TT, Pham TD, Nguyen CT, Delfos J, Archibald R, Dang KB, Hoang NB, Guo W, Ngo HH (2022) A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. Sci Total Environ 804:150187
27. Wang X, Ge L, Li X, Gherardi S (2014) The feasibility of using ENVISAT ASAR and ALOS PALSAR to monitor pastures in Western Australia. Photogramm Eng Remote Sens 80(1):43–57
28. Liao Z, He B, Quan X (2020) Potential of texture from SAR tomographic images for forest aboveground biomass estimation. Int J Appl Earth Obs Geoinf 88:102049
29. Naidoo L, Mathieu R, Main R, Wessels K, Asner GP (2016) L-band Synthetic Aperture Radar imagery performs better than optical datasets at retrieving woody fractional cover in deciduous, dry savannahs. Int J Appl Earth Obs Geoinf 52:54–64
30. Nichol JE, Sarker MLR (2010) Improved biomass estimation using the texture parameters of two high-resolution optical sensors. IEEE Trans Geosci Remote Sens 49(3):930–948
31. Kruger A, Makamo L, Shongwe S (2002) An analysis of Skukuza climate data. Koedoe 45(1):87–92
32. Wessels K, Li X, Bouvet A, Mathieu R, Main R, Naidoo L, Erasmus B, Asner GP (2023) Quantifying the sensitivity of L-Band SAR to a decade of vegetation structure changes in savannas. Remote Sens Environ 284:113369
33. Low AB, Rebelo AG (1998). Vegetation of South Africa, Lesotho and Swaziland: a companion to the vegetation map of South Africa, Lesotho and Swaziland. Department of Environmental Affairs and Tourism, Pretoria, South Africa
34. Mucina L, Rutherford MC, Palmer AR, Milton SJ, Scott L, Lloyd J, Van der Merwe B, Hoare D, Bezuidenhout H, Vlok J (2006) "Nama-karoo biome." The vegetation of South Africa. Lesotho and Swaziland Strelitzia 19:324–347
35. Trollope W, Potgieter A, Zambatis N (1989) Assessing veld condition in the Kruger National Park using key grass species. Koedoe 32(1):67–93
36. Adjorlolo C, Mutanga O (2013) Integrating remote sensing and geostatistics to estimate woody vegetation in an African savanna. J Spat Sci 58(2):305–322
37. Peichl M, Arain MA (2006) Above-and belowground ecosystem biomass and carbon pools in an age-sequence of temperate pine plantation forests. Agric For Meteorol 140(1–4):51–63
38. Torres R, Snoeij P, Davidson M, Bibby D, Lokas S (2012) The Sentinel-1 mission and its application capabilities. 2012 IEEE International Geoscience and Remote Sensing Symposium, IEEE. https://doi.org/10.1109/IGARSS.2012.6351196
39. Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. https://doi.org/10.1145/2939672.2939785
40. Beltran JC, Valdez P, Naval P (2019) Predicting protein-protein interactions based on biological information using extreme gradient boosting. 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), IEEE. https://doi.org/10.1109/CIBCB.2019.8791241
41. Kganyago M, Adjorlolo C, Sibanda M, Mhangara P, Laneve G, Alexandridis T (2022) "Testing Sentinel-2 spectral configurations for estimating relevant crop biophysical and biochemical parameters for precision agriculture using tree-based and kernel-based algorithms." Geocarto International: 1–25. https://doi.org/10.1080/10106049.2022.2146764
42. Shrestha N (2020) Detecting multicollinearity in regression analysis. Am J Appl Math Stat 8(2):39–42
43. Chen L, Wang Y, Ren C, Zhang B, Wang Z (2019) Optimal combination of predictors and algorithms for forest above-ground biomass mapping from Sentinel and SRTM data. Remote Sensing 11(4):414
44. Luo S, Jiang X, He Y, Li J, Jiao W, Zhang S, Xu F, Han Z, Sun J, Yang J (2022) Multi-dimensional variables and feature parameter selection for aboveground biomass estimation of potato based on UAV multispectral imagery. Front Plant Sci 13:948249
45. Zhang X, Liu C-A (2023) Model averaging prediction by K-fold cross-validation. Journal of Econometrics 235(1):280–301

46. Wong T-T (2015) Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recogn 48(9):2839–2846
47. James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning with Applications in R (2nd ed), Springer
48. Dube T, Mutanga O (2015) Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. ISPRS J Photogramm Remote Sens 101:36–46
49. Qadeer A, Shakir M, Wang L, Talha SM (2024) Evaluating machine learning approaches for aboveground biomass prediction in fragmented high-elevated forests using multi-sensor satellite data. Remote Sensing Applications: Society and Environment 36:101291
50. Hamdan O, Hasmadi IM, Aziz HK, Norizah K, Zulhaidi MH (2015) L-band saturation level for aboveground biomass of dipterocarp forests in peninsular Malaysia. J Trop For Sci 27(3):388–399
51. Crabbe RA, Lamb DW, Edwards C, Andersson K, Schneider D (2019) A preliminary investigation of the potential of sentinel-1 radar to estimate pasture biomass in a grazed pasture landscape. Remote Sensing 11(7):872
52. Eisfelder C, Kuenzer C, Dech S (2012) Derivation of biomass information for semi-arid areas using remote-sensing data. Int J Remote Sens 33(9):2937–2984
53. Voormansik K, Zalite K, Sünter I, Tamm T, Koppel K, Verro T, Brauns A, Jakovels D, Praks J (2020) Separability of mowing and ploughing events on short temporal baseline Sentinel-1 coherence time series. Remote Sensing 12(22):3784
54. Wang X, Ge L, Li X (2013) Pasture monitoring using sar with cosmo-skymed, envisat asar, and alos palsar in otway, australia. Remote Sensing 5(7):3611–3636
55. Sinha S, Jeganathan C, Sharma LK, Nathawat MS (2015) A review of radar remote sensing for biomass estimation. Int J Environ Sci Technol 12:1779–1792
56. Ali I, Schuster C, Zebisch M, Förster M, Kleinschmit B, Notarnicola C (2013) First results of monitoring nature conservation sites in alpine region by using very high resolution (VHR) X-band SAR data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6(5):2265–2274
57. Nasirzadehdizaji R, Balik Sanli F, Abdikan S, Cakir Z, Sekertekin A, Ustuner M (2019) Sensitivity analysis of multi-temporal Sentinel-1 SAR parameters to crop height and canopy coverage. Appl Sci 9(4):655
58. Rapiya M, Ramoelo A, Truter W (2023) Seasonal evaluation and mapping of aboveground biomass in natural rangelands using Sentinel-1 and Sentinel-2 data. Environ Monit Assess 195(12):1544
59. Shen Z, Yong B, Gourley JJ, Qi W, Lu D, Liu J, Ren L, Hong Y, Zhang J (2020) Recent global performance of the Climate Hazards group Infrared Precipitation (CHIRP) with Stations (CHIRPS). J Hydrol 591:125284
60. Beriaux E, Jago A, Lucau-Danila C, Planchon V, Defourny P (2021) Sentinel-1 time series for crop identification in the framework of the future CAP monitoring. Remote Sensing 13(14):2785
61. Wang J, Xiao X, Bajgain R, Starks P, Steiner J, Doughty RB, Chang Q (2019) Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. ISPRS J Photogramm Remote Sens 154:189–201
62. Ali I, Cawkwell F, Dwyer E, Green S (2016) Modeling managed grassland biomass estimation by using multitemporal remote sensing data—A machine learning approach. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10(7):3254–3264
63. Chan SK, Bindlish R, O'Neill P, Jackson T, Njoku E, Dunbar S, Chaubell J, Piepmeier J, Yueh S, Entekhabi D (2018) Development and assessment of the SMAP enhanced passive soil moisture product. Remote Sens Environ 204:931–941
64. Cawse-Nicholson K, Anderson MC (2018) Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) Mission. Jet Propulsion Laboratory, California Institute of Technology, California, pp 91109–8099
65. Dorigo W, Hahn S, Hohensinn R, Paulik C, Wagner W, Drusch M, van Oevelen P (2010) The international soil moisture network-a data hosting facility for in situ soil moisture measurements in support of SMOS cal/val. EGU General Assembly Conference Abstracts. https://doi.org/10.5194/hess-15-1675-2011
66. Mutanga O, Masenyama A, Sibanda M (2023) Spectral saturation in the remote sensing of high-density vegetation traits: A systematic review of progress, challenges, and prospects. ISPRS J Photogramm Remote Sens 198:297–309
67. Sinha S, Sharma L, Jeganathan C, Nathawat M, Das A, Mohan S (2015) Efficacy of InSAR coherence in tropical forest remote sensing in context of REDD. International Journal of Advancement in Remote Sensing, GIS and Geography 3(1a):38–46
68. Haralick RM, Shanmugam K, Dinstein IH (1973) "Textural features for image classification." IEEE Transactions on systems, man, and cybernetics (6): 610–621. https://doi.org/10.1109/TSMC.1973.4309314
69. Liu W, Xu C, Zhang Z, De Boeck H, Wang Y, Zhang L, Xu X, Zhang C, Chen G, Xu C (2023) Machine learning-based grassland aboveground biomass estimation and its response to climate variation in Southwest China. Front Ecol Evol 11:1146850