Leveraging Machine Learning for Spatio-temporal Monitoring of Carbon Sequestration in Urban Reforested Landscapes in eThekwini Municipality, South Africa.

by

COLLINS MATIZA (220112661)

Thesis submitted in fulfilment of the academic requirements of

Doctor of Philosophy in Environmental Science

School of Agricultural Earth and Environmental Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

January 2024

1 PREFACE

The research contained in this thesis was completed by Collins Matiza (220112661) while based in the Discipline of Geography, School of Agricultural, Earth and Environmental Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa. The research was financially supported by the South African Research Chair (SARChI) in Land use and the Wood Rights Programme.

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

Signed: Professor Onisimo Mutanga (Supervisor)

Date: 15/02/2024

Signed: Doctor Kabir Peerbhay (Co-Supervisor)

Date: 23/02/2024

Signed: Romano Lottering (Co-Supervisor)

Date: 23/02/2024

2 DECLARATION 1: PLAGIARISM

I, Collins Matiza, declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work.

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university.

(iii) this dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written, but the general information attributed to them has been referenced.

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

(v) where I have used material for which publications followed, I have indicated in detail my role in the work.

(vi) this dissertation is primarily a collection of material, prepared by me, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included.

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: Collins Matiza

Date: 23 February. 2024

DECLARATION 2: PUBLICATIONS

My role in each paper and presentation is indicated. The * indicates corresponding author.

Chapter 2

1. Matiza, C.*, Mutanga, O., Peerbhay, K., Odindi, J., & Lottering, R. 2023 A systematic review of remote sensing and machine learning approaches for accurate carbon storage estimation in natural forests, *Southern Forests: A Journal of Forest Science*, 1-19. https://doi.org/10.2989/20702620.2023.2251946.

Chapter 3

2. Matiza, C.*, Mutanga, O., Peerbhay, K., Odindi, J., & Lottering, R. 2023, Assessing aboveground biomass in reforested urban landscape using machine learning and remotely sensed data, *Journal of Spatial Science, 1–27.*

https://doi.org/10.1080/14498596.2024.2343764

Matiza, C the same paper was presented at the SAEON Indaba Conference in October 2022.

Chapter 4

3. Matiza, C.*, Mutanga, O., Peerbhay, K., Odindi, J., Lottering, R.& Mngadi, M., 2023, Utilizing Planetscope spectral data to quantify above-ground carbon stocks in urban reforested landscapes. *Ecological Informatics, p.102472.* https://doi.org/10.1016/j.ecoinf.2024.102472

Chapter 5

4. Matiza, C.*, Mutanga, O., Peerbhay, K., Odindi, J., Lottering, R.& Mngadi, M., 2023, Quantifying the impact of urban reforestation on carbon accumulation using RapidEye satellite imagery and extreme gradient boosting regression**,** *Ecological Applications*, 183–189. (Under Review).

Chapter 6

5. Matiza, C.*, Mutanga, O., Peerbhay, K., Odindi, J., Lottering, R.& Mngadi, M., 2023, Mapping aboveground carbon stocks under past, present and future climate scenarios in urban reforested Landscapes, *Beyond the Surface: Revealing Ecosystem Services through Remote Sensing: Chapter 9.* (Under Review).

__ [redacted] _____

Signed: Collins Matiza

Date: 23 February 2024

3 ABSTRACT

Urbanization and expanding cities have emerged as major drivers of deforestation, forest degradation, and associated carbon emissions. While urban areas account for a disproportionate share of global greenhouse gas emissions, urban reforestation initiatives provide significant opportunities for climate change mitigation.

This study investigates the utility of remote sensing and machine learning techniques for quantifying, mapping, and tracking carbon accumulation and storage by reforested urban landscapes. Through a systematic literature review, multisource satellite data integration (i.e. Sentinel-1, Sentinel-2, Rapideye and Planetscope spectral data), time-series image analysis, and future climate scenario modelling (i.e., CMIP6 models with 3 representative climate pathways in shared socio-economic pathways RCP-SSP), the research demonstrates the capabilities of geospatial analytics for robust carbon monitoring to strengthen localized climate action planning.

Key findings showcase the ability of remote sensing spectral data, derived from the multisource satellite integration, to map heterogeneous reforested biomass and model spatio-temporal variations in aboveground carbon stocks across an urban landscape in South Africa, over seven years. Comparative assessment of reforestation carbon accumulation based on ecological history informs management prioritization for optimized climate benefits. Furthermore, simulation of carbon stocks using remote sensing data, climate models and machine learning indicate a shift in carbon accumulation under the three shared socio-economic pathways, with the low emissions pathway showing an increase in carbon overtime. The methodological framework in this study delivers an adaptable tool for continuous near real-time quantification of urban forest carbon accumulation to support evidence-based decision-making through targeted monitoring, strategic expansions, and adaptive management. While underscoring the immense potential of urban reforestation for mitigation, the research highlights the need to refine an understanding of realized carbon capture rates, utilize emerging data sources such as Planetscope, Sentinel-1 and 2, and Rapid-Eye, expand spatio-temporal assessments, and integrate economic valuations to maximize climate policy impacts. This contributes to sustainable forest management and climate change mitigation.

4 ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the many organizations and individuals who supported me throughout this research.

First and foremost, I thank the University of KwaZulu-Natal School of Agricultural, Earth and Environmental Sciences for providing me the opportunity to pursue my doctoral studies. I am grateful for the funding generously provided by the WoodRights programme and NRF-SARChI Chair in Landuse Planning and Management that made this work possible. I also thank the eThekwini Municipality for granting access to their study site and sharing invaluable historical data. Moreover, I appreciate Planet Labs Inc. for contributing crucial Planetscope and RapidEye imagery.

Most importantly, I wish to thank my advisor, Prof. Onisimo Mutanga, for his outstanding mentorship. Professor Mutanga taught me not only scientific rigor but also resilience and determination to overcome obstacles. I am honored to have worked under his guidance. I sincerely thank my supervisor, Dr. Kabir Peerbhay, for his constant encouragement and advice. Our discussions always motivated me when I doubted myself. Additionally, I appreciate the support and direction provided by my co-supervisor Dr. Romano Lottering. Knowing I could turn to him with any challenge was reassuring. I also thank Prof. John for pushing me to think deeper scientifically and sharpening my writing skills.

The Department of Geography Pietermaritzburg campus provided unwavering support, for which I am grateful. Special thanks to Shanita Ramanroop, Mr Devon, Brice, Dr Shenelle Lottering, Dr Mthembeni Mngadi, Prof Jemma Finch, Prof Trevor Hill, Dr Lawrence Dube and Sizwe Hlatshwayo.

My dear friends and colleagues, I could not have completed this journey without your support. A special thanks to Adeola, Bongokuhle Sibiya, Dadirai Matarira, Anita Masenyama, Duduzile Diza, and Snethemba Ndlovu for the camaraderie and community you provided during the highs and lows of this journey. Joyce Mlay, your assistance came at a critical time, and I am forever grateful. Dr. Matongera, your encouraging words kept me motivated even when the road seemed darkest. Most of all, this achievement would not have been possible without my family's unconditional love and patience. I wish my parents could have witnessed what I have become.

Finally, I thank my partner for unwavering support, sacrifices, and confidence in me - this belongs to you too. And most importantly, I thank God for the blessings that made this dream attainable.

5 TABLE OF CONTENTS

Page

6 LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

7 ACRONYMS

ANN	Artificial Neural Networks
AGB	Aboveground biomass
ARVI	Atmospheric Resistant Vegetation Index
Bnd	Band
BPNN	Back-propagation neural network
BRT	Boosted regression tree
C	Carbon
CART	Classification and Regression Trees
CB	Cubist
CI_{green}	Green Chlorophyll Index
CIRE	Red-Edge Chlorophyll Index
CNN	Convolutional Neural Networks
Con	Contrast
DBH	Diameter at Breast Height
DNN	Deep Neural Networks
ENS	Ensemble
Ent	Entropy
ENVI	Environment for Visualizing Images
ESA	European Space Agency
EVI	Enhanced Vegetation Index
GAM	General Additive Network
GAM	Generative Additive Models
GEMI	Global Environmental Monitoring Index
GLCM	Gray-level Co-occurrence Matrix
GNDVI	Green normalized vegetation index
GP	Gaussian Processes
GPS	Global Positioning System
GRVI	Green Ratio Vegetation Index
ha	Hectares
IPCC-GPG	Inter-Governmental Panel on Climate Change Good Practice Guidance on Natio Inventories
IRECI	Inverted Red-Edge Chlorophyll Index
K-NN	K-Nearest Neighbor
LCCC	Lin Concordance Correlation
LMSTEP	linear multi-stepwise regression
LR	Linear regression
MAE	Mean Absolute Error
MCARI	Modified Chlorophyll Absorption in Reflectance Index
ML	Machine learning

MLPNN	Multi-layer Perceptron Neural Network
MLR	Multiple linear regression
MNDVI	Modified Normalized Difference Vegetation Index
MSAVI	Modified Soil Adjusted Vegetation Index
NDREII	Normalized Difference Red-Edge-II
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
OLSR	Ordinary least squares regression
OSAVI	Optimized Soil Adjusted Vegetation Index
QGIS	Quantum Geographic Information System
r	Correlation
R^2	Correlation coefficient
REDD+	Reducing Emissions from Deforestation and forest Degradation
RENDVI	Red-Edge Normalized Difference Vegetation Index
RF	Random Forest
RMSE	Root Mean Square Error
RVI	Ratio Vegetation Index
SAR	Synthetic Aperture Radar
SAVI	Soil Adjusted Vegetation Index
SGB	Stichastic Gradient boosting
SOM	Self-organizing map
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
t	Tonnes
TWI	Topographic Wetness Index
UNFCCC	United Nations Framework Convention for Climate Change
VH	Vertical transmit /Horizontal receive
VV	Vertica transmit / Vertical receive
XGBoost	Extreme Gradient Boosting

1 CHAPTER ONE: GENERAL INTRODUCTION

1.1 Introduction

1.1.1 Global forest carbon dynamics

Natural forest ecosystems, which cover 31% of terrestrial land area play a vital role in the global carbon cycle by sequestering and storing significant carbon through trees and woody vegetation. (Apps and Price, 1996, Chen and Xu, 2010, Cockburn et al., 2016). Beyond regulating climate, forests provide ecosystem services that support biodiversity, economic activities, and human livelihoods (Ngo et al., 2021). However, escalating anthropogenic threats from urban expansion, deforestation, and habitat fragmentation degrade these ecosystems. Globally, about 10 million hectares of forests are converted annually (Alig, 2003). In South Africa's KwaZulu-Natal province, over 615,000 hectares of tree cover were lost from 2001-2022 (Ramjeawon et al., 2020). As intact forests decline, their carbon storage, biodiversity, and climate resilience diminish. Expanding cities drive an increase in clearing and fragmentation of adjoining natural forests (Dorning et al., 2015, Kowarik et al., 2019) . Ultimately, urgent integrated management approaches are needed to address these threats and safeguard ecosystem services for climate change mitigation and human well-being (Fink, 2016, Nowak et al., 2013, Sithole et al., 2018). The United Nations REDD+ program provides incentives to conserve and enhance forest carbon stocks while benefiting biodiversity and communities (IPCC, 2006, Shukla et al., 2019).

1.1.2 Urban reforestation as a nature-based solution

Urban reforestation is a promising strategy that enhances carbon sinks. By reintroducing and nurturing tree growth within urban areas, reforestation efforts not only contribute to the sequestration of carbon but also provide a multitude of co-benefits such as improved air quality, reduced urban heat island effects, and enhanced biodiversity (Teo et al., 2021, Dietrich et al., 2014). The strategic development of urban green spaces is increasingly recognized as a key component in sustainable urban planning, offering a nature-based solution to the environmental challenges posed by urban expansion and intensification (Ehdi et al., 2016). Although the importance of urban reforestation is widely recognized, its integration into urban planning, management, and policy has been hindered by a lack of comprehensive information regarding its

impact on carbon accumulation and ecosystem regulation (Shukla et al., 2019). Knowledge gaps exist around quantifying and mapping the effects of urban reforestation on ecological functions like carbon accumulation to manage and enhance mitigation functionality (Zhao and Sander, 2015, Teo et al., 2021). Expanding our understanding of the relationship between urban reforestation and these crucial ecological dynamics is vital for effective decision-making and sustainable urban development. Moreover, improved monitoring and assessment of landscape level ecological processes and functions could ensure balanced interactions between social, economic and environmental functions, while maintaining a sustainable environment and mitigating climate change risk.

1.1.3 Challenges in carbon accounting

A significant body of research has focused on quantifying aboveground carbon accumulation in natural forests (Uniyal et al., 2022, Shen et al., 2021, Zhang et al., 2022b, Dube et al., 2018, Basyuni et al., 2023). However, a gap remains regarding the carbon contributions of urban reforested areas. Bridging this by conducting spatial assessments of biomass and carbon in city reforestation contexts is crucial to comprehend their climate mitigation potential and inform planning (Teo et al., 2021). Incorporating spatio-temporal dynamics allows more effective assessment (Matiza et al., 2024a, Mngadi et al., 2022b). Concise quantification of urban forest carbon has been hindered by limitations like scarce resources, limited standardized methodologies and technical expertise, especially in developing regions (Zhang et al., 2009). Nonetheless, such information could benefit city managers and planners, especially relating to spatial distributions of biomass and carbon in urban reforested landscapes. Establishing cost-effective, robust techniques incorporating available datasets is essential for reliable monitoring and quantification of global forest carbon accumulation and storage.

1.1.4 Role of remote sensing and machine learning in aboveground carbon monitoring

Traditionally, quantifying forest carbon relied on field measurements, either destructive or non-destructive (Gara et al., 2016, Sibanda et al., 2015). While accurate, this approach is unsustainable - laborious, costly, and impractical for large areas (Mngadi et al., 2021a). Regional and global assessments are essentially impossible due to insufficient tree sampling and ethical concerns. Therefore, there is a pressing need to transition towards innovative techniques that leverage remote sensing and freely available data assets. The Global Forest Observations

Initiative (GFOI) (Penman et al., 2016) and the Inter-Governmental Panel on Climate Change Good Practice Guidance (IPCC-GPG) on National Greenhouse Gas Inventories (IPCC-GPG, 2004) have underscored remote sensing as a robust technique in providing, reliable and cost effective data necessary for wall to wall predictions and mapping of forest carbon dynamics. Despite complex landscapes, remote sensing data can provide spectral information at both local and regional scales (Peerbhay et al., 2016). Image analysis with field samples allows scalable carbon estimation without extensive fieldwork. Multispectral data including Sentinel-2, Landsat-8 and WorldView-2 have estimated biomass and carbon in forests (Dube et al., 2018, Gara et al., 2016, Peerbhay et al., 2013). LiDAR provides 3D vegetation data to model biomass (Tian et al., 2021, Zhang and Shao, 2021). Such techniques must balance accuracy with cost-effectiveness for wide adoption.

Remote sensing's efficacy is demonstrated by several key factors including the progressive enhancement of sensor capabilities enabling finer spectral and spatial data acquisition, the comprehensive historical data archive facilitating longitudinal analysis (Odebiri et al., 2023) (IPCC, 2006, Gyamfi-Ampadu et al., 2021), rapid generation of current image datasets ensuring timely insights, and the continuous advancements in hardware and software facilitating robust analysis (Mutanga et al., 2012, Matiza et al., 2023a). Recent multispectral satellite systems like PlanetScope, Sentinel-2, and RapidEye provide 3-60 m resolution across optimized visible and infrared bands achieving success in monitoring forest attributes (Baloloy, 2018, Andreatta et al., 2022, Chinembiri et al., 2023). Additionally, extensive time-series data facilitates long-term stock change analysis. Red-edge bands between red and near-infrared wavelengths demonstrate sensitivity to carbon-linked vegetation characteristics including chlorophyll, biomass and carotenoid (Clevers and Gitelson, 2013, Hernández-Clemente et al., 2012, Mutanga et al., 2012). However, despite capabilities, remote sensing exploitation remains limited for quantifying urban forest carbon to inform planning and climate mitigation. More research is needed to capitalize on modern tools for precise monitoring of urban forest carbon dynamics.

Literature suggests that multispectral sensors face limitations in forest monitoring like shadowing, saturation issues, and impacts of inclement weather, which can degrade output imagery (Mngadi et al., 2021a, Wang et al., 2019b). Synthetic aperture radar (SAR) systems operating at penetrating micro wavelengths can help overcome these constraints. SAR backscatter data demonstrates potential to complement multispectral signals for predicting urban forest carbon patterns (Matiza et al., 2024a, Jiang et al., 2007, Gyamfi-Ampadu and Gebreslasie,

2021). Thus, sensor integration could improve model performance for carbon mapping (Grenier et al., 2020, Han et al., 2017). Additionally, incorporating vegetation indices and strategic spectral bands linked to carbon sequestration permits carbon quantification when embedded into estimation models (Dube et al., 2018, Gara et al., 2016). Further key environmental variables like topography from digital elevation models (DEM) and climate data from sources like WorldClim can be easily integrated (Odebiri et al., 2023, Zhao et al., 2019b, Poggio et al., 2018). However, diverse data inputs pose analytical challenges. Recent machine learning methods capable of handling complex datasets are required for accurate carbon mapping (Campesato, 2020, Tamiminia et al., 2020, Ali et al., 2015). Overall, an integrated modeling approach blending multi-sensor remote sensing data with advanced statistical techniques promises significant enhancements in quantifying and projecting carbon dynamics for sustainable urban landscape management.

Substantial research has established correlations between field-measured forest carbon stocks and environmental variables like climate, vegetation density, and terrain (Lopatin et al., 2019, Duchesne et al., 2016). Diverse geospatial modelling techniques have been formulated to predict carbon accumulation and storage based on these ecological relationships (Lourenço, 2021). Approaches derive statistical connections between observed carbon and remotely sensed surrogates. Traditional regression methods have shown initial promise for carbon mapping (Odebiri et al., 2022, Pham and Yoshino, 2017). However, complex nonlinear interactions between remote sensing data, climate, vegetation indices, terrain metrics, and field carbon pose challenges for conventional statistics (Nandy and Kushwaha, 2021, Mngadi et al., 2022b). Machine learning algorithms like random forest, SVM, XGBoost, KNN and ANN provide powerful modelling frameworks (Salunkhe et al., 2018, Pham et al., 2019). These can capture nonlinear relationships without rigid assumptions. Ensemble approaches integrate decision trees, SVMs maximize feature space separability, KNN enables localized accuracy, and ANNs enable sophisticated pattern recognition ((Shaik and Srinivasan, 2018, Jhamtani et al., 2021, Ghorbanian et al., 2022). Overall, advanced algorithms enable nuanced quantification of links between diverse remote sensing metrics, climate and terrain data to create high-performance predictive carbon models. Harnessing machine learning promises enhancements in leveraging growing remote sensing archives (Das et al., 2020).

While machine learning based remote sensing (RS) techniques for mapping carbon are well-established, research on implementation and transferability to diverse landscapes like urban

landscapes remains limited. Leveraging abundant remote sensing data and ML analytics provides significant potential for reliable, continuous urban reforestation mapping and evaluation. This can inform carbon accounting, support mitigation policies, and meet REDD+ and Paris Agreement reporting (Lovell et al., 2010, Mackey et al., 2022) . However, customized ML-remote sensing frameworks for urban reforested areas are lacking. This study addresses this gap by creating a modelling workflow integrating historical and current high-resolution optical and radar imagery with advanced ML to produce high-accuracy urban forest carbon maps. The framework establishes scalable, operational carbon monitoring to strengthen reforestation planning and sustainability assessments (Wassenaer et al., 2012). Overall, this uniquely highlights prospects at the intersection of proliferating remote sensing streams and ML tools for evidence-based management of urban reforested landscapes tackling climate change.

1.2 Aims

This study aimed to develop and evaluate remote sensing and machine learning based approaches for predicting and mapping aboveground forest biomass and carbon storage in urban reforested areas and explore long-term monitoring capabilities to support carbon balance assessments and climate change mitigation policy.

1.3 Objectives

The objectives of the study were to:

1. To review the adoption of remote sensing and machine learning in forest carbon accumulation and storage in natural forests.
2. To estimate and map aboveground forest biomass in an urban reforestation context using an ensemble of machine learning algorithms and multi-source data.
3. To explore the utility of PlanetScope spectral data to quantify aboveground carbon stocks in urban reforested landscapes.
4. To quantify the spatio-temporal changes in forest carbon stocks in an urban reforested landscape using RapidEye spectral data and extreme gradient boosting regression.
5. To model and map aboveground carbon stocks under past, present and future climate scenarios in urban reforested Landscapes.

1.4 Research questions.

1. How accurately can remote sensing data map and monitor forest aboveground carbon accumulation across an urban reforestation landscape when integrated with machine learning algorithms?
2. Which category of machine learning algorithms (e.g. ensemble methods and/or neural networks) demonstrates optimal performance for spatially explicit quantification of urban reforestation carbon accumulation using remote sensing inputs?

3. What specific combinations of spectral bands and, vegetation indices calculated from multispectral data, offer the highest sensitivity for capturing variability in aboveground carbon stocks across diverse urban reforested landscapes?
5. How do environmental factors including terrain, disturbance history, climate, and land use change trajectories influence the spatial distribution of carbon storage across complex urban reforestation landscapes?
6. How can time-series based neural network models be leveraged to reliably forecast trajectories of aboveground carbon sequestration by urban forests under varying climate and land use change scenarios to inform adaptive management?

1.5 Study area.

The Buffelsdraai reforestation site, situated in the eThekwini Municipality of KwaZulu-Natal Province, South Africa, lies approximately 25 km northwest of the city center of Durban, in proximity to the Inanda suburb and dense residential townships like KwaMashu (Figure 1.1). The geographic coordinates of the reforestation site fall between 30°58′20.08″E and 29°37′55.17″S, encompassing an area of approximately 800-1000 hectares. Established through eThekwini initiatives for the 2010 FIFA World Cup to mitigate greenhouse gas emissions, the reforestation project aimed to create a diverse, functional indigenous forest ecosystem to provide climate regulation services. Specifically, sequestering substantial atmospheric carbon to offset the impacts of the 2010 World Cup and nearby landfill (Mngadi et al., 2022a, Douwes et al., 2015a, Mugwedi et al., 2017).

The region exhibits a subtropical climate, characterized by a mean annual temperature ranging between 22 and 27 °C, and a mean annual rainfall varying from 600 to 1000 mm. The study area

exhibits an uneven topography supported by a resistant underlying dwyka tillite rock (Mugwedi et al., 2017, Douwes et al., 2015a). Vegetation consists primarily of dense coastal forests dominated by Acacia species (*Caffra and Robusta*), *Syzygium cordatum, Bridelia micrantha* and *Erythrina caffra*. This study analyzes the Buffelsdraai site over a period from reforestation initiation (2012) through 2040 to evaluate historic, current and projected reforestation carbon accumulation potential(Mngadi et al., 2022b).

The Buffelsdraai reforestation area lies within the Indian Ocean Coastal Belt biome, a biodiversity hotspot containing numerous endemic plant and animal species (Mugwedi et al., 2017). The rehabilitated forest aims to recreate the natural subtropical coastal forest vegetation of the region. In addition to the previously mentioned dominant tree species like Acacia, Syzygium cordatum and Erythrina, the site includes other indigenous flora such as the protected milkwood tree (Mimusops caffra), pondo milkwood (Mimusops obovata), red beech (Protorhus longifolia) and forest buddleja (Nuxia floribunda) (Matiza et al., 2024b).

Restoring this diverse forest habitat provides a critical ecological corridor and sanctuary for indigenous fauna species. Small antelope like the endangered Samango monkey and blue duiker thrive in the understorey vegetation(Aryal et al., 2012). The dense canopy shelters endemic bird species of conservation concern such as the spotted ground-thrush and green twinspot. Other avifauna identified include the red-capped robin-chat, green wood-hoopoe and trumpeter hornbill. Ecological challenges include the inherent fragmentation from the surrounding urban matrix, as well as environmental pressures like invasive alien plant species encroachment, soil erosion on steep slopes, uncontrolled pedestrian traffic and illegal biomass harvesting (Mugwedi et al., 2017). Ongoing management efforts focus on clearing invasive species, installing boardwalks to minimize disturbance, and engaging local communities on sustainable utilization (Douwes et al., 2015a).

Overall, the Buffelsdraai reforestation project represents a critical conservation effort to re-establish a regionally-representative, high biodiversity urban forest that supports threatened species and ecosystem services within a rapidly developing coastal metropolitan area. Its success highlights the potential for integrated sustainability initiatives to harmonize urban development with environmental preservation.

The Buffelsdraai reforestation area presents an ideal case study for developing and applying novel machine learning approaches to precisely monitor urban reforestation carbon dynamics (Matiza et al., 2024b). Its location within a rapidly urbanizing coastal metropolitan region grants high relevance for sustainable city planning and climate mitigation policies. The site itself encompasses a heterogeneous matrix of diverse reforestation patches of varying ages, densities, and species compositions set within an urban matrix - characteristics that pose challenges for traditional carbon mapping techniques developed for contiguous natural forests (Sithole et al., 2018). Moreover, the site's distinct topographic variations, documented impacts from climate fluctuations like droughts, and evolving soil conditions induced by reforestation cumulatively influence carbon accumulation rates (Mugwedi et al., 2017). This confluence of complex environmental drivers necessitates advanced machine learning models integrated with multi-source remote sensing data for capturing intricate relationships and generating reliable continuous carbon maps. Successfully quantifying and projecting carbon sequestration across Buffelsdraai's unique urban reforestation landscape would demonstrate machine learning's potential for replicable, effective monitoring to guide sustainable urban forest management globally.

Partners: eThekwini Municipality, uMgungundlovu District Municipality, uMngeni Resilience Project, Sustainable and Healthy Food Systems (SHEFS), Department of Cooperative Governance and Traditional Affairs, Department of Economic Development, Tourism and Environment, and South African National Biodiversity Institute,

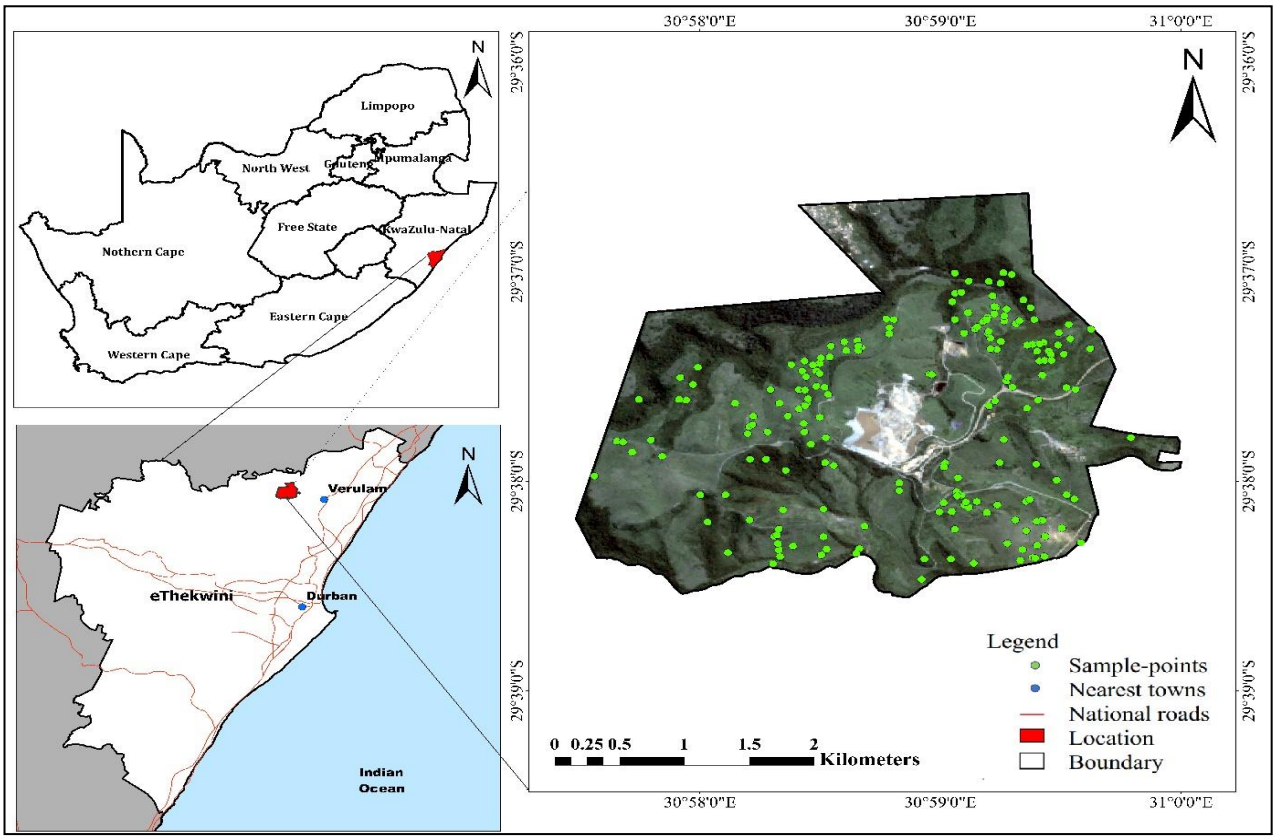


Figure 1.1 The location of Buffelsdraai reforestation site and sample points, within the eThekwini Municipality in KwaZulu-Natal Province, South Africa.

1.6 Outline of Thesis

This thesis comprises five articles excluding the introduction and the synthesis chapters (which respond to the research objectives). The literature review and methodologies are entrenched within the outlined papers. Out of the five research manuscripts two have been published in peer-reviewed journals, one is in-press and two are currently under review. Kindly note that each manuscript is presented as a separate chapter within this thesis. Each chapter is mostly self-contained, containing a literature review, materials and methods, results and discussion, conclusions and a final link to the next chapter in the form of a summary. Hence there is some overlap, commonalities and theoretical repetition between chapters. This was unavoidable due to the parallel transition of doctrines that serve as a foundation of the current scientific knowledge. In this regard each chapter should be viewed as a self-contained, stand-alone piece of writing, without detracting from the overall aim and context of the thesis.

Chapter 1: This chapter presents the general introduction and contextualizes the research study underscoring the importance of reforestation initiatives in improving carbon storage and accumulation over time to understand its contribution in climate change mitigation. This chapter also highlights the methods and datasets that can be adopted to quantify the spatio-temporal changes in aboveground biomass and carbon within urban landscapes. The research aims and objectives were also presented.

Chapter 2: This chapter comprehensively reviews remote sensing and machine learning techniques for quantifying carbon storage and accumulation in forest ecosystems. Given the urgent climate mitigation motivations, this work emphasizes the importance yet challenges of accurate carbon stock assessments. Traditional field measurements can prove limited, especially for vast, remote regions. Thus, we highlight recent remote sensing and computational advancements that offer more efficient, innovative carbon estimation alternatives. Specifically, this review focuses on optical, radar, and LiDAR data coupled with machine learning algorithms for predicting aboveground biomass carbon in natural forests. Prominent techniques include advanced regression algorithms applied to multispectral datasets. However, complex terrain and signal noise can hamper accuracy in specific biomes. Limitations and recommendations for future progress are examined. Uniquely, this review integrates findings across multiple sensors, modeling methods, and forest ecosystems to provide a holistic update on using Earth observation data for forest carbon tracking. Key knowledge gaps are synthesized to guide further developments in operationalizing these technologies for climate change mitigation policy and decisions.

Chapter 3: Focuses on the estimation of aboveground biomass in reforested urban landscape using machine learning and spectral data, vegetation indices and terrain metrics This study used a multi-source remote sensing data (i.e. Planetscope, Sentinel, SRTM) and field measurements to build aboveground biomass regression models using machine learning algorithms such as extreme gradient boosting (XGBoost), random forests (RF), support vector machines (SVM), and k-nearest neighbours (K-NN). Results found that XGBoost achieved the best AGB predictions from 4.1-286.5 tonnes/hectare across the reforestation area, demonstrating the value of the modelling approach for quantifying

urban forest carbon stocks. By leveraging site data and advanced modelling techniques, this research contributes to an innovative methodology for monitoring biomass accumulation and carbon offsets in disturbed landscapes undergoing ecological restoration.

Chapter 4: Although reforestation initiatives considered effective nature-based solution for forest carbon capture and storage, their carbon accumulation has remained largely unknown. Therefore, this chapter examines the prospect of a fine resolution Rapideye (5m) multispectral sensor in quantifying the spatio-temporal changes of aboveground carbon stocks in a reforested urban environment using extreme gradient boosting regression. The chapter also explores the influence of unique vegetation indices and environmental covariates in improving the accuracy of predictions.

Chapter 5: This chapter focuses on quantifying and mapping historical forest carbon accumulation from 2012 to 2019 across a rapidly reforesting urban landscape using a combination of field measurements and multi-temporal RapidEye satellite imagery. Specifically, an extreme gradient boosting (XGBoost) model is developed to estimate carbon density changes over time leveraging the high spatial (5m) and temporal resolution data available through the RapidEye constellation. The changing carbon accumulation and storage across areas experiencing active reforestation under varying land use types, topographic contexts and climatic conditions are evaluated. Comparative assessment of estimated gains based on local biophysical factors provides insights into ecological controls influencing observed carbon trajectories. Additionally, validation against an independent dataset of field observations enables robust evaluation of carbon mapping accuracy. Overall, the chapter delivers an operational methodology for retrospective quantification of reforestation-driven carbon mitigation benefits using extensively available archived optical imagery. Results will equip urban planners and policymakers to better account for, assess and expand high carbon sequestering green infrastructure to support net emissions reduction goals.

Chapter 6: This chapter models and predicts forest aboveground carbon stocks in a South African reforested landscape under changing land use and climate scenarios, using

space-for-time substitution (SFTS) techniques. Specifically, artificial neural networks are developed to map historical (2012) and current (2022) carbon densities. These models provide baseline carbon estimates and allow assessment of recent accumulation rates. Building on this, future trajectories of urban forest carbon sequestration between 2022-2040 are forecasted using CMIP6 climate projections and shared socioeconomic pathway (SSP) land use change narratives for the region. Three scenarios are evaluated – SSP1-26 (sustainability), SSP2-45 (middle-of-the-road) and SSP5-85 (fossil-fuelled development) – representing low, medium and high emissions conditions respectively. Comparative analysis of current and potential future carbon densities provides insights into climate change mitigation potentials of urban reforestation under varying global trajectories. Findings will equip local municipalities to better account for and leverage green infrastructure in achieving emission reduction targets. This research delivers an adaptable, evidence-based framework for near real-time mapping and forward-looking simulations of urban forest carbon to strengthen mitigation policymaking.

Chapter 7: This chapter synthesizes all findings and conclusions from the research objectives. The chapter further provides important recommendations for future research. Future teaching, learning and research possibilities are included. The list of references is provided at the end of this chapter.

2 CHAPTER TWO: A systematic review of remote sensing and machine learning approaches for accurate carbon storage estimation in natural forests

This chapter is based on:

C. Matiza, O.Mutanga , K.Peerbhay, J. Odindi, and R. Lottering (2023). A systematic review of remote sensing and machine learning approaches for accurate carbon storage estimation in natural forests

Abstract

Assessing carbon storage in natural forests is paramount in the ongoing efforts against climate change. While traditional field-based methods for quantifying carbon storage pose challenges, recent advancements in remote sensing and machine learning offer efficient and innovative alternatives. This systematic literature review investigates the latest developments in utilising optical, radar, and light detection and ranging (LiDAR) remote sensing data, coupled with cutting-edge machine learning algorithms, to estimate carbon storage in natural forests. Non-parametric machine-learning algorithms commonly applied to multispectral datasets have emerged as prominent tools for predicting aboveground carbon storage. Nonetheless, accurately assessing forest carbon storage using remote sensing data can be arduous in regions with complex terrain and diverse species where dataset noise may be pronounced. Alternatively, adopting freely available optical sensors with moderate resolution has showcased reliability in estimating forest carbon storage. Hence, leveraging the integration of multi-sensor data with machine learning techniques has yielded substantial improvements in the accuracy of carbon storage estimation. This study identifies the most sensitive remote sensing variables that correlate with measurable biophysical parameters, thus highlighting the pivotal role of geospatial technologies in estimating terrestrial aboveground carbon storage. The study also delineates gaps and limitations inherent in current practices, underscoring the need for further investigations in this rapidly evolving field. By unifying conventional methods with state-of-the-art technologies, this study contributes to the advancement of accurate and efficient carbon storage assessments. This research holds substantial promise in bolstering global climate change mitigation efforts by assuming such a transformative role. Ultimately, this study aimed to demonstrate to researchers, policy makers and practitioners the importance of embracing the combined power of remote

sensing and machine learning as a tool for safeguarding our natural forests and fighting against climate change.

Keywords: carbon sequestration, climate change adaptation, net-zero emissions, precision forestry, remote sensors, sustainable forest management

2.1 Introduction

Natural forests are vital for maintaining global ecosystem services, including carbon sequestration and storage (Rajasugunasekar et al., 2023) . The sequestration of carbon by trees through photosynthesis and the subsequent carbon storage in natural forests is considered one of the most effective and cost-efficient ways to reduce atmospheric carbon dioxide concentrations (Granieri et al., 2014, Austin et al., 2020). To limit global warming to below 2 °C, countries are now required to implement measures that promote large-scale removal of carbon dioxide (Halme et al., 2019b, Cook-Patton et al., 2020). Achieving net-zero emissions and reducing carbon emissions are critical in this context, and natural forests are key to mitigating climate change through carbon assimilation and storage. However, degradation and deforestation in natural forests have resulted in a global increase in atmospheric carbon dioxide, negatively affecting natural forests' carbon sequestration and storage capacity (Cook-Patton et al., 2020, Dube et al., 2014b). Therefore, it is necessary to conduct frequent, detailed assessments of carbon storage in natural forests to maintain their sustainability.

While traditional methods for estimating carbon sequestration in natural forests have proven laborious, time- consuming and expensive, recent developments in remote sensing (RS) have revolutionised the field by offering efficient and accurate ways to estimate carbon storage in natural forests (Gao et al., 2018, Muukkonen, 2006). Remote sensing, which has numerous advantages, such as accessibility to remote locations, availability of historical datasets, repetitive spatial coverage and cost-effectiveness, has become a mainstream method for quantifying carbon storage in natural forests (Mohd Zaki and Abd Latif, 2017, Li et al., 2019b). Specifically, combining remote sensing data with field measurements has proven particularly successful in recent years, enhancing both model calibration and the precision of predictions (Shen et al., 2018b, Zhang et al., 2013). However, it is essential to consider the most appropriate remote sensing dataset to establish good carbon estimation models and characterise ongoing carbon changes in natural forests. Different datasets produce different results. Therefore the choice of remote sensing data significantly affects the accuracy of the prediction (Vicharnakorn et al.,

2014, Li et al., 2020a). While remote sensing data alone is not always comprehensive enough, machine learning (ML) algorithms have emerged as an effective solution for producing robust models for predicting natural forest carbon storage.

Choosing an appropriate machine learning algorithm is crucial for accurately estimating carbon storage in natural forests (Safari et al., 2017, Gao et al., 2018, Jiang et al., 2007, Su et al., 2020). Geostatistical models, which have been widely used, are limited in their ability to establish a non-linear connection between the predictor variables of remote sensing datasets and natural forest carbon storage (Dube et al., 2014a, Su et al., 2020). Conversely, non-parametric machine learning algorithms such as random forests (RF), support vector machines (SVM) and extreme gradient boosting (XGBoost) have emerged as viable alternatives (Pham et al., 2020a, Zhou et al., 2020). These algorithms can handle high-dimensional remote-sensing data and do not require a strictly linear relationship. Non-parametric machine learning approaches also allow for parameter adjustments through a tuning process, improving the accuracy of aboveground forest carbon storage prediction. Evidence suggests that these approaches can replace the traditional methods for estimating natural forest carbon storage at different scales (Li et al., 2019b, Zhang et al., 2020). For example, using machine learning algorithms, we can overcome the limitations of geostatistical regression models and achieve more precise and comprehensive results in estimating natural forest carbon storage (Veronesi and Schillaci, 2019, Zhang et al., 2020).

This systematic literature review investigated recent developments in using optical (multispectral and hyperspectral), radar, and light detection and ranging (LiDAR) remote sensing data, along with machine learning algorithms, for estimating carbon storage in natural forests. Previous reviews have provided an overview of remote sensing applications for carbon storage assessment in different types of forests (Ali et al., 2015, Pham et al., 2019, Salunkhe et al., 2018, Gyamfi-Ampadu and Gebreslasie, 2021), this review concentrates on broader natural forest carbon storage assessment. This review aimed to ascertain the role of machine learning algorithms in predicting carbon storage from the field and remote sensing data in natural forests, determine the most effective sensors for natural forest carbon storage predictions; and evaluate the accuracy achieved by different machine learning algorithms on remotely sensed data, highlighting their strengths and weaknesses. The specific objectives of this study were to identify the most effective and commonly used machine learning algorithms for natural forest carbon storage assessment; synthesize and review published applications of machine learning to remotely sensed data for aboveground carbon storage estimation in natural forests; and analyse trends

shown in the reviewed studies. This study employed descriptive and correlation analysis approaches to identify the relationships between study characteristics and estimation performance, as well as common datasets used in conjunction with machine learning methods and the advantages and disadvantages of such methods. While this study primarily focused on carbon storage estimation, we also acknowledge the contributions of studies related to forest volume prediction using machine learning approaches which are closely related to carbon storage studies, as they offer valuable insights into machine learning algorithm performance (Sun and Liu, 2020), dataset calibration (Fassnacht et al., 2014), model transferibility (Cosenza et al., 2022) and predictor selection (Tompalski et al., 2019).

2.2 Methods

2.2.1 Systematic review literature search

Three main databases (SCOPUS, Web of Science and Elton B Stephens Company host [EBSCOhost]) were used for the literature search. The search strategy was based on the topic (abstract, article title and keywords), and was limited to articles written in English from 2009 to 2022 to ensure consistency, comparability, and focus on high-quality, peer-reviewed literature in the predominant language of scientific publishing in the fields of environmental sciences, remote sensing, and machine learning, while acknowledging the potential language bias and limitations introduced by excluding non-English articles. Studies from 2009 to 2022 were chosen because machine learning, particularly its applications in environmental sciences and remote sensing, experienced a significant surge during this period, driven by several key developments in the field. Prior to 2009, machine learning studies in domains such as carbon sequestration, carbon storage, aboveground biomass estimation, and natural forest ecosystem analysis were relatively limited. This can be attributed to several factors including computational constraints, Scarcity of large datasets, theoretical limitations, and limited interdisciplinary collaborations.

The focus was on peer-reviewed articles. The protocol used for this systematic review was PRISMA-P (Moher et al., 2015). Keywords were selected as search words: ‘natural forests’, ‘carbon sequestration’, ‘carbon storage’, ‘remote sensing’, ‘aboveground biomass’ and ‘machine learning’ The initial search was conducted in September 2021 and was repeated in February 2022.Google Scholar was used for the 2022 search since most databases had not been updated. In total, 1 284 records were retrieved, and exported to EndNote 20 for further screening,

searching and selection. From the 1 284 results obtained from the initial search, studies that were not directly related to carbon storage, did not include at least one machine learning algorithm, did not include remote sensing data and were not conducted in a natural forest ecosystem were eliminated, leaving n = 900 entries. Figure 1 presents the steps taken in the systematic literature review at various stages.

2.2.2 Exclusion Criteria

In EndNote 20, references were screened to exclude off-topic ones unrelated to carbon sequestration (biomass) and storage in natural forests. This involved removing references on remote sensing, machine learning research and carbon sequestration studies in non-natural forest ecosystems. In addition, theoretical articles, such as overviews, reviews, book chapters, and literature that did not have clear methods and study sites, were removed. A total of 535 references were eliminated at this stage.

2.2.3 Inclusion criteria

The full-text screening focused on the following relevant criteria: (1) spatially explicit study site; (2) a well-defined methodology with a mandate for assessing natural forest parameters linked to forest carbon storage and management concerns using remote sensing and machine learning; and (3) records, permitting an explicit assessment of the natural forest parameters related to carbon sequestration using remote sensing. A total of 245 entries were retrieved and retained for further review and analysis, and 120 were eliminated to maintain a manageable and thorough systematic review approach. (Figure 2.1). Furthermore, keyword analysis was performed to identify the links that existed in the 245 retrieved entries.

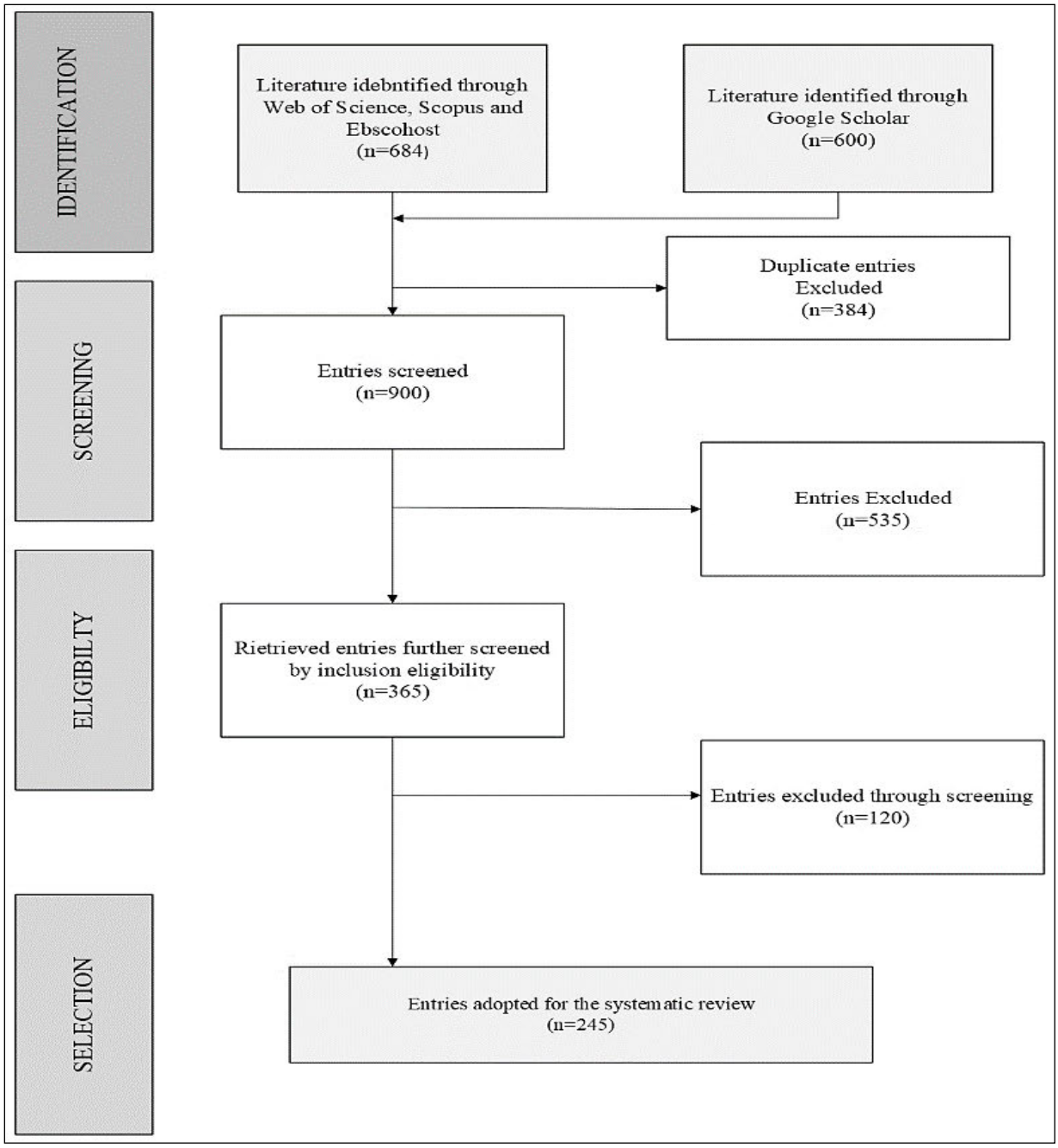


Figure 2.1 The flow process of the systematic review adapted from (Tamiminia et al., 2020) PRISMA (Preferred Reporting Items for Systematic Reviews and Meta Analysis).

2.2.4 Statistical analysis

Quantitative synthesis and analysis of the retrieved articles were performed. Network analysis using Vos Viewer software was conducted to visualize the co-occurrence and occurrence of key

terms (van Eck and Waltman, 2017). This metric allowed for the identification of interconnections among keywords related to published papers in remote sensing, and machine learning technologies for aboveground carbon storage estimation in natural forests.

Basic statistical frequencies and trend analyses were performed using Microsoft Excel to assess the use of remote sensing technologies in aboveground carbon storage estimations for natural forests. This review is divided into two sections to address the research objectives. The first section provides an overview of the existing progress in remote sensing technologies and machine learning frameworks applied to aboveground carbon storage estimations in natural forests. The analysis thoroughly examined the literature search characteristics, distribution of studies, remote sensing data, and methods utilized for aboveground carbon storage estimation in natural forests. Additionally, the study focused on commonly employed machine learning frameworks for evaluating forest carbon storage. Based on the results of the first section, the second section focused on analyzing the results and explaining what they mean regarding machine learning architectures for estimating aboveground carbon storage in natural forests. The analysis involved segmenting the characteristics of each paper, including data sources, study location, nature of the study, explanatory variables, type of natural forest, and accuracies achieved by the machine learning algorithms.

To rigorously evaluate the accuracy and performance of the machine learning algorithms employed for aboveground carbon storage estimation, a comprehensive statistical analysis was conducted. The root mean squared error (RMSE) and coefficient of determination (R^2) were calculated to assess the predictive ability and goodness-of-fit of the algorithms. Additionally, a 10-fold cross-validation technique was implemented to ensure robust and unbiased evaluation of model performance. The statistical significance of the algorithm performance was further analyzed using analysis of variance (ANOVA) and post-hoc Tukey's honest significant difference (HSD) tests at a 95% confidence level ($\alpha = 0.05$). These tests allowed for the identification of statistically significant differences in the accuracy and precision of the various machine learning models. Furthermore, the stability and generalization capability of the algorithms were examined through the calculation of confidence intervals (95% CI) for the RMSE and R^2 metrics. This comprehensive statistical analysis ensured a thorough and statistically sound evaluation of the machine learning algorithms, enabling informed

comparisons and selection of the most appropriate techniques for accurate estimation of aboveground carbon storage in natural forest ecosystems.

2.3 Results

2.3.1 Description of literature search characteristics

From the 245 articles in the study sample, 6 948 keywords were retrieved. The analysis was performed using VOS viewer (van Eck and Waltman, 2017). An analysis of keywords that appeared frequently led to the identification of several research topics that indicated the most interesting lines of research among scholars between 2009 and 2022. It also displayed the general direction of research. The analysis of the co-occurrence of keywords indicated six clusters in which keywords were retrieved and their links. The classification of keywords was based on a single circle with a colour unique to a specific keyword, resulting in different colours for different keywords, as shown in (Figure 2.2). The size of the circle denotes the number of keywords constituting the cluster (i.e., the number of articles that used the keyword). The line linkages display the union between two different groups; an increase in line thickness indicates the number of interactions. The keyword analysis aimed to display the relevance of the literature search and its adoption into the final literature search database after applying the inclusion and exclusion criteria. A total of 176 keywords were the most prominent terms, signified by the six largest clusters, that is, red, green, blue, yellow, purple and light blue. The key terms in the red cluster (40 items) were ‘forest’, ‘carbon stock’, ‘forest structure’ ‘normalised difference vegetation index’ and ‘mapping’, which were closely related. The green cluster (35 items) key terms were ‘remote sensing technology’, ‘predictor variables’, ‘estimation’, ‘bands,’ ‘vegetation indices’, ‘regression model’, and ‘random forest’. This cluster links remote sensing to machine learning algorithms, with biomass estimation as the task. The inclusion of terms such as ‘image’, ‘unmanned aerial systems (UAS)’ and ‘unmanned aerial vehicle (UAV)’ in this cluster presents the linkage between satellite imagery and other sources of data, which implies the utility of various remote sensing data in aboveground biomass (AGB) or aboveground carbon (AGC) stock predictions in natural forests.

The blue cluster had keywords ‘agb’, ‘algorithm’, ‘performance’, ‘artificial neural networks’, ‘agb estimation’, and ‘feature selection’. The yellow cluster had its key terms as ‘LiDAR’,

‘spectral bands’, ‘Sentinel’, ‘red-edge’, and ‘worldview’. Spectral band settings and sensor type influence the performance of the sensor in estimating forest AGB carbon stocks (Wang et al., 2019a). The purple cluster (25 items) had ‘China’ ‘imagery’, ‘mae’, ‘rmse’, and ‘coefficient’ as its key terms which directly implies the standard metrics of evaluation used to validate machine learning models and remote sensing data for aboveground carbon storage estimation in natural forests. Finally, the light blue cluster (17 items) connected terms such as ‘China’, ‘MODIS (Moderate Resolution Imaging Spectroradiometer) normalised difference vegetation index (NDVI)’, and ‘carbon cycle’. This articulates the wide use of MODIS-derived (NDVI) as a proxy for studying forest aboveground carbon storage as a major component of carbon cycling, with most studies carried out in China (Chen and Xiao, 2019, Zhang et al., 2019c).

Six keyword clusters had the highest link strengths. The six keywords corresponded to the search strings used in SCOPUS, Web of Science and EBSCOhost in this systematic review. This indicates the effectiveness of the six keyword clusters in retrieving relevant and valuable literature related to the research topic. The link strength reflects the degree of association between the search terms and the content of the retrieved articles, indicating their significance in addressing the literature search.

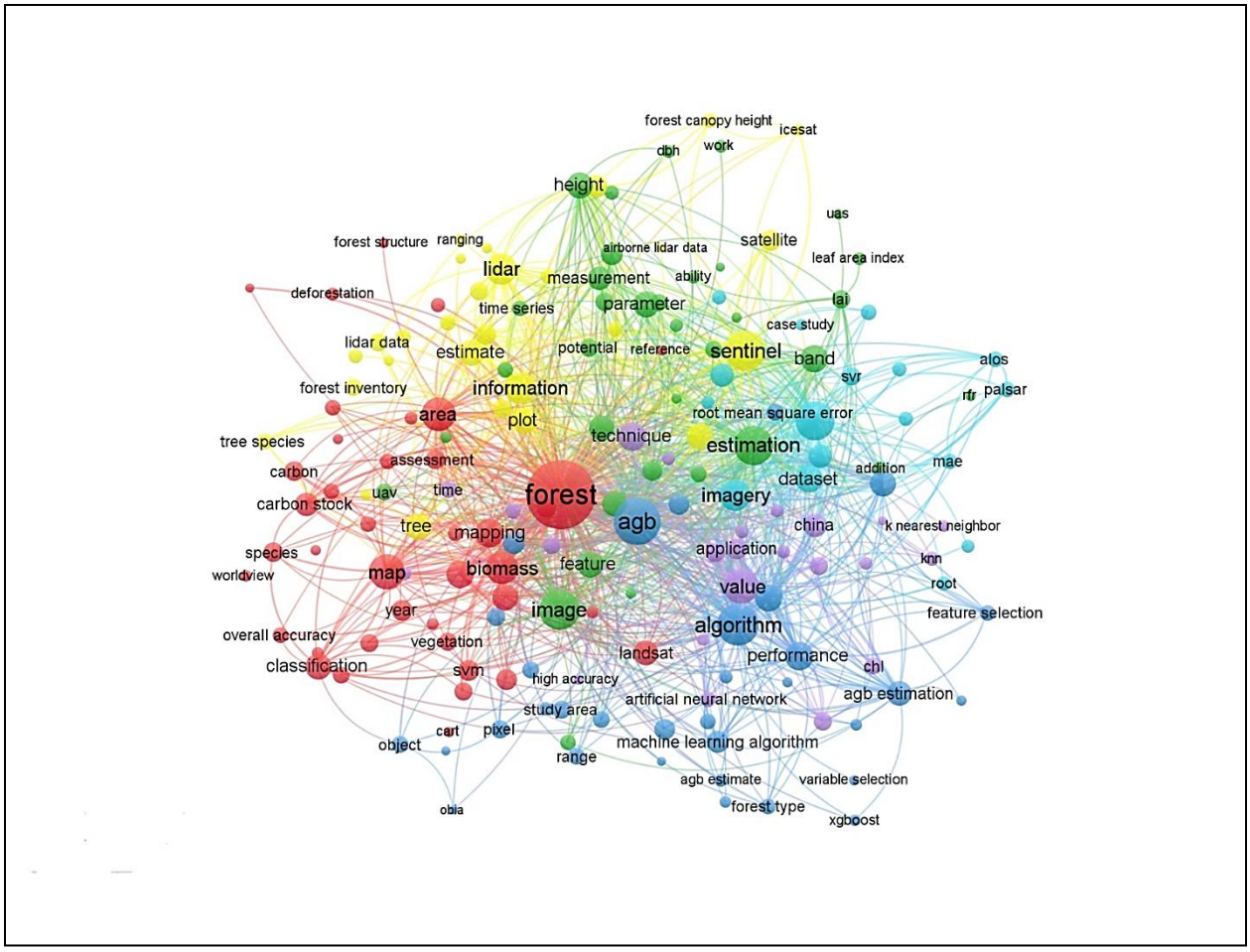


Figure 2.2**:** Network analysis of key terms in the titles and abstracts of the reviewed studies. The term sizes correspond to their frequency and their colors to statistical clusters (determined using the node-repulsion Lin Log method (Noack, 2007)).

2.3.2 Articles selected.

A literature search in Scopus, Web of Science, Ebscohost and Google Scholar retrieved 1 284 papers. The retrieved papers underwent further screening and exclusion that resulted in the removal of duplicate papers and papers that did not meet the inclusion criteria (n = 1 039), reducing the number of papers (n = 245). The study used a final list of 245 papers that used ML and RS approaches to estimate aboveground forest carbon storage. Figure 3 presents the total number of papers published per year for our sample.

Initially, in 2009, less than five publications focused on estimating aboveground carbon storage in natural forests using ML approaches (Goetz et al., 2009, Tamm and Remm, 2009, Wijaya and

Gloaguen, 2009). Publications slightly increased in 2011, and one of the papers investigated the relationship between urbanization and vegetation carbon storage in Taiwan using ML and RS (Ren et al., 2011) . After 2011, the frequency of publications on ML and RS applications for aboveground carbon storage estimation in forests was inconsistent and increased in 2017. Since 2017, the number of publications has increased annually. The annual numbers of publications between 2017 and 2022 (i.e., until September 2022) were 10, 18, 25, 30, 68 and 56, respectively. It can be noted that 2021 represents the peak of publications in our sample (Figure 2.3). Additionally, the period between 2017 and 2022 had the highest number of publications. Studies are expected to continue to steadily increase as technologies and sensors improve.

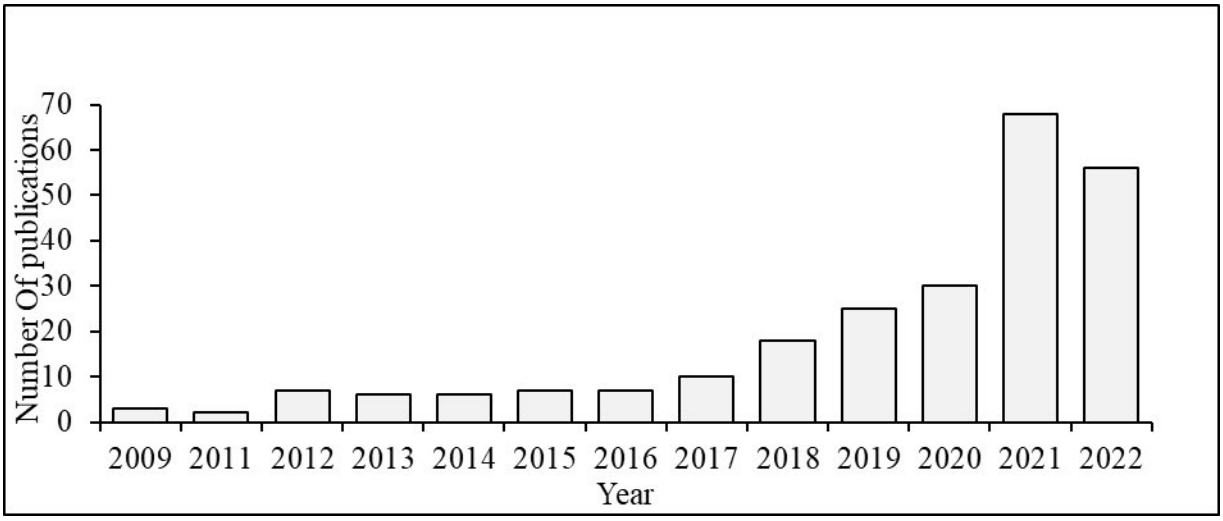


Figure 2.3 Total number of papers published by year (n = 245).

2.3.3 Geographical extent and coverage

Figure 2.4 presents the spatial distribution of the retrieved entries related to the aboveground carbon storage of natural forests and the geographical location of the publications and their study sites (country and continent). In some cases, a study can appear in numerous countries and on many continents, such as Avitabile et al. (2012) and , which were conducted at continental and global levels, respectively. Figure 2.4 shows that 33% of these studies were conducted in Asia, with China having the most studies. Europe has 22% of the studies, 18% in both North and South America and 10% of the studies were in Australia. Africa had 10% of the studies, and half of the

total number came from South Africa. Figure 2.4 presents the significant gaps in the geographic distribution of the published articles.

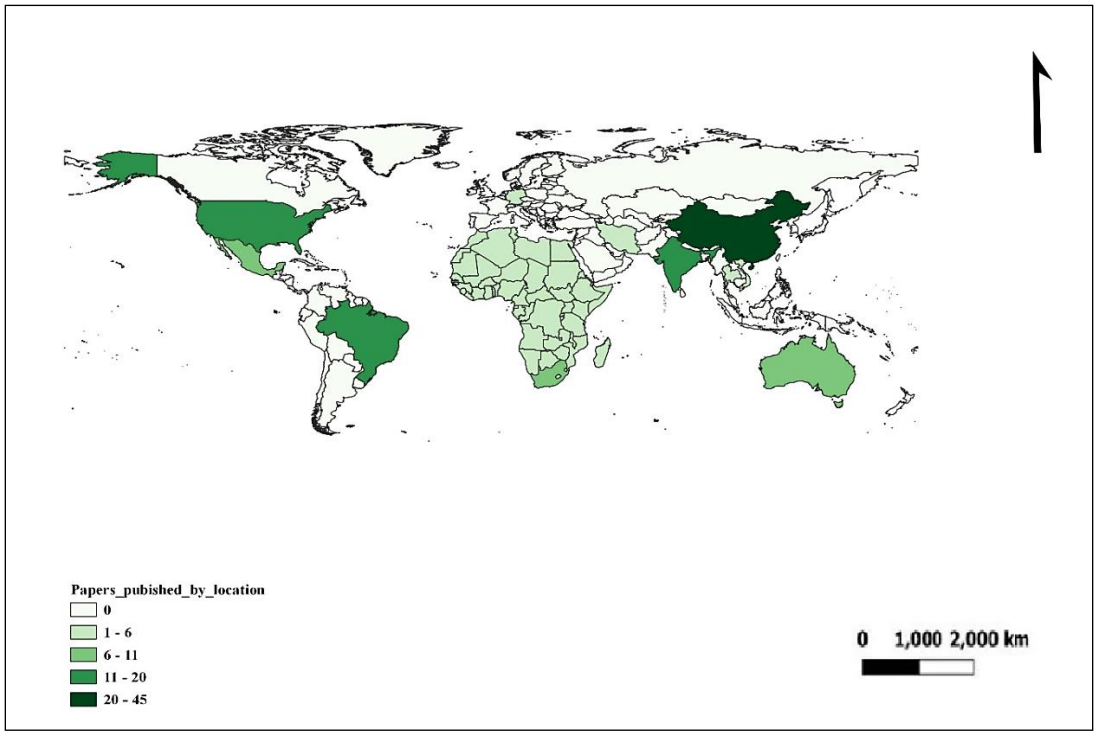


Figure 2.4 Global distribution of studies on estimation of forest carbon storage in natural forests.

2.3.4 Integration of machine learning frameworks and remote sensing data.

Remote sensing and machine learning (ML) are highly complementary technologies that can be seamlessly integrated to enhance precision carbon storage estimation in natural forest ecosystems(Cheng et al., 2024, Zhang et al., 2022a). Remote sensing provides a wealth of data on tree traits and environmental conditions through various platforms, such as satellites, aerial vehicles, and ground-based sensors. These platforms can capture multispectral, hyperspectral, and 3D structural information about trees over large spatial and temporal scales. However, extracting meaningful insights from this raw data can be a complex task (Matiza et al., 2024b). This is where machine learning comes into play. By leveraging advanced algorithms, ML models can learn to interpret the remote sensing data and extract relevant tree traits (Upreti, 2022, Mosin et al., 2019). For example, artificial neural networks (ANNs) can be trained to detect and measure forest features, such as leaf area, plant height, and biomass, from high-resolution imagery (Matiza et al., 2024a, Shen et al., 2021). Similarly, regression models can be used to

estimate biophysical parameters, like chlorophyll content or water stress, from spectral signatures. The integration of remote sensing and ML can be illustrated through a case study on estimating aboveground carbon (AGC) storage in natural forest ecosystems. Satellite-derived vegetation indices, combined with LiDAR-derived structural information and environmental data (e.g., precipitation, temperature), can be used as inputs to an ML model. The model can then be trained to predict AGC at the landscape or regional scale, allowing for more accurate monitoring and management of forest carbon stocks.

Remote sensing and machine learning (ML) are highly complementary technologies that can be seamlessly integrated to enhance precision carbon storage estimation in natural forest ecosystems. Remote sensing provides a wealth of data on tree traits and environmental conditions through various platforms, such as satellites, aerial vehicles, and ground-based sensors. These platforms can capture multispectral, hyperspectral, and 3D structural information about trees over large spatial and temporal scales. However, extracting meaningful insights from this raw data can be a complex task. This is where machine learning comes into play. By leveraging advanced algorithms, ML models can learn to interpret the remote sensing data and extract relevant tree traits. For example, artificial neural networks (ANNs) can be trained to detect and measure forest features, such as leaf area, plant height, and biomass, from high-resolution imagery. Similarly, regression models can be used to estimate biophysical parameters, like chlorophyll content or water stress, from spectral signatures. The integration of remote sensing and ML can be illustrated through a case study on estimating aboveground carbon (AGC) storage in natural forest ecosystems. Satellite-derived vegetation indices, combined with LiDAR-derived structural information and environmental data (e.g., precipitation, temperature), can be used as inputs to an ML model. The model can then be trained to predict AGC at the landscape or regional scale, allowing for more accurate monitoring and management of forest carbon stocks.

The progression and improvement of sensors in remote sensing have significantly improved aboveground carbon storage estimates in natural forest ecosystems (Ni-Meister et al., 2010, Gao et al., 2018, Halme et al., 2019a). It is essential to note the types of remote sensing data involved in forest carbon storage retrieval to understand the advancements of ML-frameworks about AGB and AGC storage prediction. Accordingly, statistical analysis of the 245 entries indicated that thirty-two different types of sensors were noted in the reviewed literature (Figure 2.5). Landsat

imagery has the highest frequency of use (16 %) (Figure 2.5), which corresponds to the multispectral image type (Figure 2.6). Based on the sample data, Sentinel-2 multispectral was identified as a key term in the network analysis (yellow cluster) (Figure 2.2) and was ranked second with 12% application. In contrast, a similar 8% application of MODIS and LiDAR data was found (Figure 2.5), listed among the key terms in the network analysis, suggesting its importance in forest AGB and carbon stock retrieval. The findings of this study further illustrate the growing popularity of newer sensor technologies such as unmanned aerial vehicle light detection ranging (UAV-LiDAR), UAV-hyperspectral, and high-resolution imagery such as World-View and Quickbird, which had similar applications of 4% each (Figure 2.5) Commercial sensor systems with a cost attached to their access are less popular than those with open-access sensors. IKONOS, Rapid-Eye, Planetscope, and Gaofen-hyperspectral had 2% use in forest aboveground carbon stock estimation-related studies. With an increase in the cost of access, the application of the sensor in forest biomass and carbon stock retrieval decreased, which can be seen in the low percentages of sensors such as Enmap (1%), Hyperion (1%), AVHRRS (0.4%), NOAA-AVHRRS (0.4%), and AIR-hyperspectral (0.4%). Ariel images were used in one study.

This growing interest in UAV-based and very high-resolution commercial sensors can be attributed to several key advantages they offer: 1) Improved spatial detail and precision - UAVs can capture extremely high spatial resolution data (<10cm) compared to traditional sources, enabling precise 3D mapping of individual tree and canopy metrics. 2) Targeted, flexible deployment - UAV sensors can be deployed over specific areas of interest on-demand, rather than relying on static satellite orbits/schedules. 3) Understory mapping capability - Low-altitude UAV sensors can "see" beneath canopy openings to better characterize the full vertical forest structure. However, the limited coverage area and high computational demands of processing massive UAV-derived point clouds remain notable challenges for operational upscaling. Integrating these datasets with broader coverage airborne LiDAR or optical imagery is likely required.

The cost considerations are also evident, with the highest adoption of newer commercial sensors concentrated in the mid-range price tiers like WorldView and RapidEye. Very expensive sensors like Enmap and Hyperion were seldom used (<1%), likely due to restrictive data costs for carbon mapping applications over large areas.

Overall, 36% of the study sample used more than one type of dataset (Figure 2.6) The authors combined passive and active sensors or used two or more sensors that did not belong to the same constellation, such as ALOS PALSAR and Landsat 8 OLI data or Landsat and Sentinel data, unmanned aerial vehicle light detection ranging UAV-LiDAR, and Sentinel multispectral data, which made multi-source data the most popular dataset (Morin et al., 2019, Peng et al., 2020). Multispectral data was second, with a total of 25% of the studies choosing to use sensors such as Sentinel-2, Landsat 8 OLI, MODIS, UAV-hyperspectral, SPOT, and IKONOS (Chen et al., 2020, Zhao et al., 2011). This shows that ML architectures generally perform well with multispectral data. Twenty-one percent of the studies used radar data, and LiDAR had a 14% use in the sample, indicating a potential rise in the popularity of their usage in geospatial applications related to biomass estimation in forests (Figure 2.6) (Ahmed et al., 2015, Navarro et al., 2020, Tian et al., 2021). A small percentage of studies applied high-resolution data (2%), such as EO-01 Hyperion and Gaofen-hyperspectral data, to aboveground carbon stock estimation (Marshall and Thenkabail, 2015, Jacon et al., 2021). Three studies used aerial (active) imagery for aboveground carbon stock estimation within forest ecosystems (Figure 2.6) (Apostol et al., 2020, Navarro-Cerrillo et al., 2018, Reiersen et al., 2022).

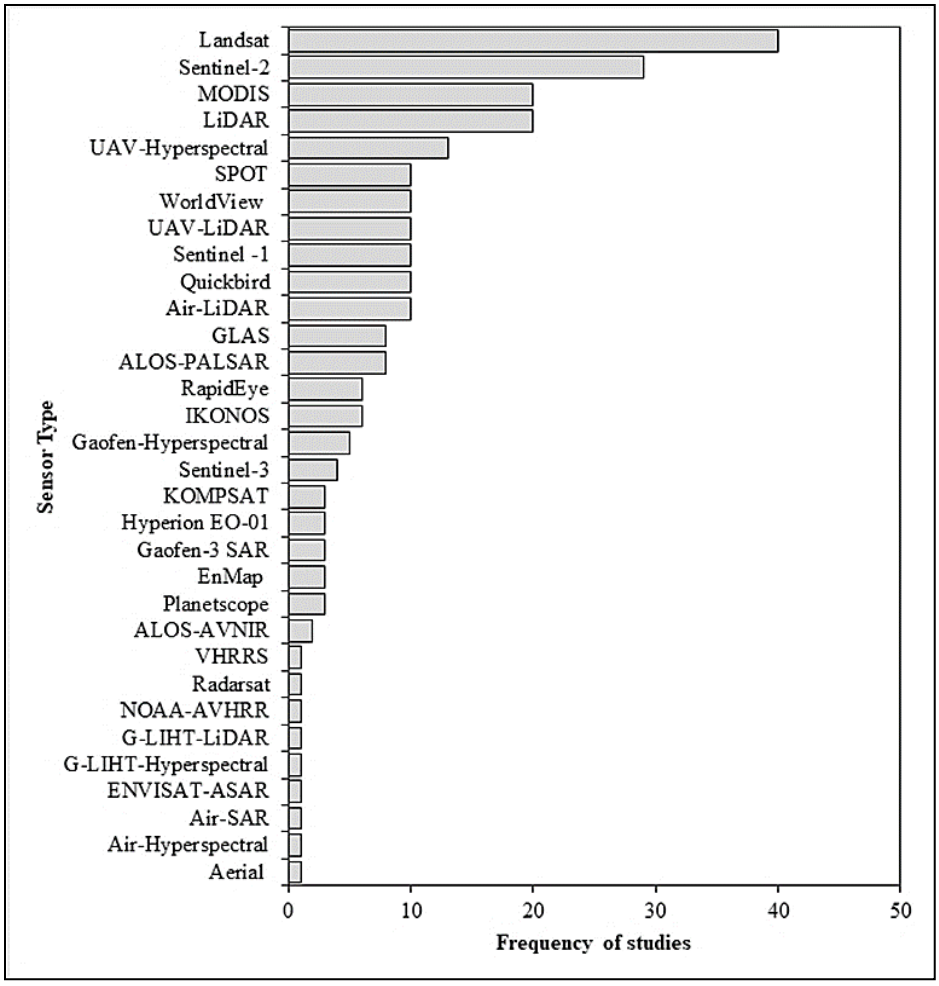


Figure 2.5 Frequency of studies utilizing a specific sensor system from 2009 to October 2022. Multi-sensor studies were treated as being unrelated (i.e., different studies, each sensor representing a study).

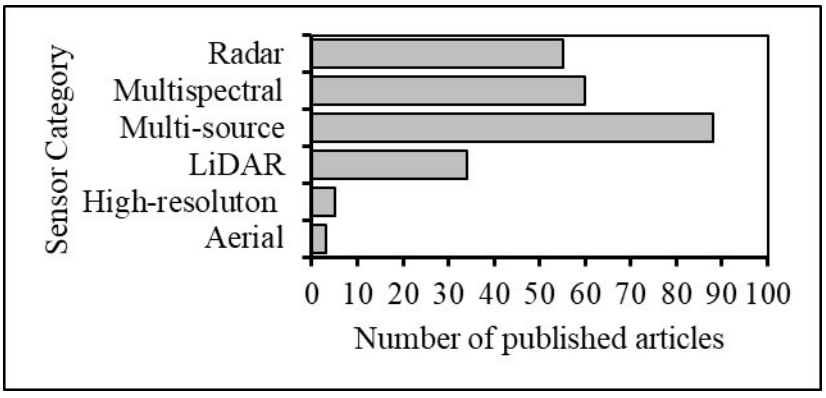


Figure 2.6 Common sensors adopted from 2009 to October 2022, aerial, high-resolution, LiDAR, multi-source, multispectral, radar. Studies using multiple sensors were treated as unrelated (i.e., different studies, each sensor representing a study).

On the other hand, sensors have been used to transform and derive information in the form of principal components (PCA), vegetation indices (VI) and texture metrics for the estimation and mapping of AGB carbon stocks in natural forests (Odebiri et al., 2022, Pham et al., 2020b). Despite numerous sensor derivatives in literature (Halme et al., 2019b, Gara et al., 2016, Peerbhay et al., 2016), the study reported derivatives that had the highest usage, and less common derivatives were not included. As a result, this study reports general radiometric image transformations widely applied to aboveground carbon storage estimates in natural forests. The Normalized Vegetation Difference Index (NDVI), a VI that contrasts the reflectance of the red and near infrared (NIR) wavelengths of light to reveal the volume of chlorophyll pigments, was the most frequently used index with 18% usage in the sample of studies (Gara et al., 2016, Ali et al., 2018, Dube et al., 2018). The enhanced vegetation index (EVI) was followed by 10%, and the green normalized vegetation index (GNDVI), and soil-adjusted vegetation index (SAVI) were 9%. The simple ratio (SR), red edge indices such as the chlorophyll index red edge (CIRE), and normalized difference red edge (NDRE) had a similar 6% use in the sample of studies. This study used, entropy and gray-level co-occurrence matrix (GLCM) the most in 7% and 11% of the studies, respectively. Finally, principal components (PCA) were only used in 3% of the studies to retrieve biomass and carbon stocks. Despite most studies using VIs, texture features, and image transformations, this review noted the use of topographic and climatic factors derived from digital elevation model images and climate data. However, few studies have used these sources (Zhou et al., 2013, Zhang et al., 2019c, Jiang et al., 2021) . Table 2.1 lists commonly used vegetation indices, image transformations, and texture features. The table also shows how often each feature is used in the literature.

Table 2.1 Commonly used Vegetation indices, image transformations and texture features.

Feature	Type of feature	Equation	Usage in Publications	Source
Normalized Difference Vegetation Index (NDVI)	Vegetation Index	$\frac{(NIR - Red)}{(NIR + Red)}$	43	(De Ocampo, 2023)
Enhanced Vegetation Index (EVI)	Vegetation Index	$2.5 * \frac{(NIR - Red)}{(NIR + 6 * Red + 7.5 * Blue + 1)}$	25	(Lin, 2012)
Simple Ratio (SR)	Vegetation Index	$\frac{NIR}{Red}$	15	(Melillos and Hadjimitsis, 2020)
Green Normalized Vegetation Index (GNDVI)	Vegetation Index	$\frac{(NIR - Green)}{(NIR + Green)}$	23	(García et al., 2018)
Soil Adjusted Vegetation Index (SAVI)	Vegetation Index	$\frac{(NIR - Red)}{(NIR + Red + L)}$	23	(Huete, 1988)
Chlorophyll Index Red-Edge (CIRE)	Vegetation Index	$\frac{NIR}{Red} - 1$	15	(Li et al., 2024a)
Normalized Difference Red Edge (NDRE)	Vegetation Index	$\frac{(NIR - RedEdge)}{(NIR + RedEge)}$	15	(Li et al., 2024b)
Gray Level Co-occurrence Matrix (GLCM)	Image texture	*Frequency of cooccurrence of pixel pairs*	29	(Kurniati et al., 2024)
Entropy	Image texture	$\sum(P * log2(P))$	16	(Kurniati et al., 2024)
Principal Component Analysis (PCA)	Image transformation	*Linear dimensionality reduction to find most important pixel patterns.*	10	(Estornell et al., 2013)

[1]Vegetation indices are calculated using image bands represented by *NIR near-infrared band, Red*-red band, *Green*-green band, *Blue*-blue band, *RedEdge*-rededge band, L-coefficient.

[1]

Different types of sensors have unique limitations and benefits when predicting carbon storage in natural forests using machine-learning algorithms (Li et al., 2020c, Dube and Mutanga, 2015). For example, high-resolution optical sensors provide data on forest structure, composition and vegetation density, which can help improve the accuracy of carbon storage predictions (Pasher et al., 2014, Gao et al., 2018, Halme et al., 2019a). However, they are limited by cloud cover, and cannot penetrate dense canopies. On the other hand, radar sensors can penetrate dense canopies and operate regardless of cloud cover, but they have a lower spatial resolution than optical sensors (Ghosh and Behera, 2018, Lu et al., 2017). LiDAR sensors offer valuable high-resolution data for modeling forest structure and aboveground carbon storage (Coops et al., 2021, Singh et al., 2024). Still, they have limitations, such as high costs, cover a small area, require considerable time to acquire data for terrestrial laser scanning (TLS), difficulty in detecting small trees and restricted vertical penetration for airborne laser scanning (ALS). The unmanned aerial vehicle laser scanning UAV-LS has challenges with measurement accuracy in complex canopy structures. While in contrast bathymetric laser scanning (BLS) has limited range and accuracy (Tian et al., 2021, Almeida et al., 2019, Urbazaev et al., 2018). Therefore, selecting the most suitable sensor type for a specific study will depend on the research question, the desired accuracy of carbon storage predictions, and the cost and availability of different sensor types**.** Table 2.2 provides information on different types of sensors used to map forest aboveground biomass (AGB) and aboveground carbon (AGC). The table shows the accuracy, cost, mapping extent, predictive error, and saturation problem of each type of sensor.

The type of remote sensing data has important implications for the choice and performance of machine learning algorithms applied to estimate forest biomass and carbon stocks. Different sensor modalities capture complementary information about forest structural properties relevant for biomass modeling (Kumar et al., 2021). Optical data like high-resolution imagery and vegetation indices provide detailed information about species composition, canopy cover and health, suiting techniques like Random Forests/Decision Trees for delineating land cover using spectral signatures, and Deep Learning (CNNs) for automatically extracting complex vegetation patterns (Li et al., 2019a). LiDAR point clouds offer 3D structural information highly valuable for biomass estimation, lending itself to Random Forests relating LiDAR metrics to biomass using reference plots, and Point Cloud Segmentation (CNN/Deep Learning) methods to directly predict biomass (Oehmcke et al., 2024, Lu and Jiang, 2024). Radar backscatter data is sensitive

to vegetation structure and moisture, enabling Random Forests to combine optical, LiDAR and radar sources for enhanced biomass modeling, and Physics-Based Retrieval inverting radar models with reference data. The increasing availability of multi-sensor data is driving ensemble/fusion approaches like Model Stacking/Ensembles combining outputs from different sensor-specific models, and Multi-Task/Transfer Learning with shared feature learning across data modalities in deep neural networks(Lu and Jiang, 2024). Ultimately, machine learning and deep learning algorithms are well-suited to leverage the rich information content across the rapidly evolving suite of remote sensing data sources for improved forest carbon mapping and modeling performance(Cheng et al., 2024, Matiza et al., 2023a).

Table 2.2 limitations and benefits of RADAR, LIDAR and optical sensors used to estimate aboveground carbon storage in natural forests.

Sensor types	Resolution	Platform	Forest AGB/AGC accuracy	Cost	Mapping extent	Predictive error	Saturation problem
Optical (passive)	Coarse	Satellite	poor	Free/low	global	≤40%	Yes
	Medium	Satellite	medium	Free/low	global	≤30%	Yes
	Fine	Satellite	High	high	Local-regional	≤20%	Yes
		Airborne					
		UAV					
	hyperspectral	Satellite	Very high	high	Local-regional	≤20%	Yes
		Airborne					
		UAV					
RADAR (active)		Satellite	high	Free/low	global	≤40%	Yes
LiDAR (active optical)		Airborne, Satellite	Very High	Very high	Local-global	<20%	No
		UAV					

* High (>40%), Moderate (≤40% and >30%), Low (≤30% and >20%), and V. Low (<20%)

2.3.5 Machine learning algorithms

Remote sensing data can best be analyzed using machine learning (ML) algorithms, and the rich information contained in these images enables ML to extract more representative features that could benefit AGC storage predictions in natural forests (Luo et al., 2021, Safari et al., 2017, Dube et al., 2018). (Figure 2.7). Twenty-one algorithms were found in the literature, and their use dates back to 2009. ML algorithms such as random forests and support vector machines are the most popular, and their use steadily increased (Figure 2.7). Because algorithms have advanced and improved, the use of machine learning algorithms has been increasing in natural forests for the estimation, and mapping of AGC storage. Since 2015, algorithms like the k-nearest neighbor method (k-NN), Gaussian process algorithm (GP), and artificial neural networks (ANN) have been continuously applied to studies relating to forest aboveground carbon estimation in natural forests (Chen et al., 2022b, Gao et al., 2018, Zhou et al., 2020). A major shift in the use of ML algorithms occurred between 2015 and 2022, with peak use in 2022. Extreme gradient boosting (XGBoost), categorical boosting (CatBOOST), stochastic gradient boosting (SGB), and gradient boosting (GB) were increasingly adopted in 2016, going forward after their introduction, and proved to be more robust than existing algorithms. Based on the chart (Figure 2.7), nonparametric algorithms are generally more prevalent than parametric algorithms such as linear regression (LR), multiple linear regression (MLR), and ordinary least squares regression (OLSR). Deep learning is also gaining popularity with algorithms such as multiple layer perceptron neural networks (MLPNN), backpropagation neural network (BPNN), and deep neural networks DNN, which have been used to estimate AGC stocks in natural forests (Cao et al., 2016, Campesato, 2020, Shao et al., 2017). Ensemble stacking algorithms (ENS) and generative additive networks/models (GAN/GAM), although limited in use, have reported positive results.

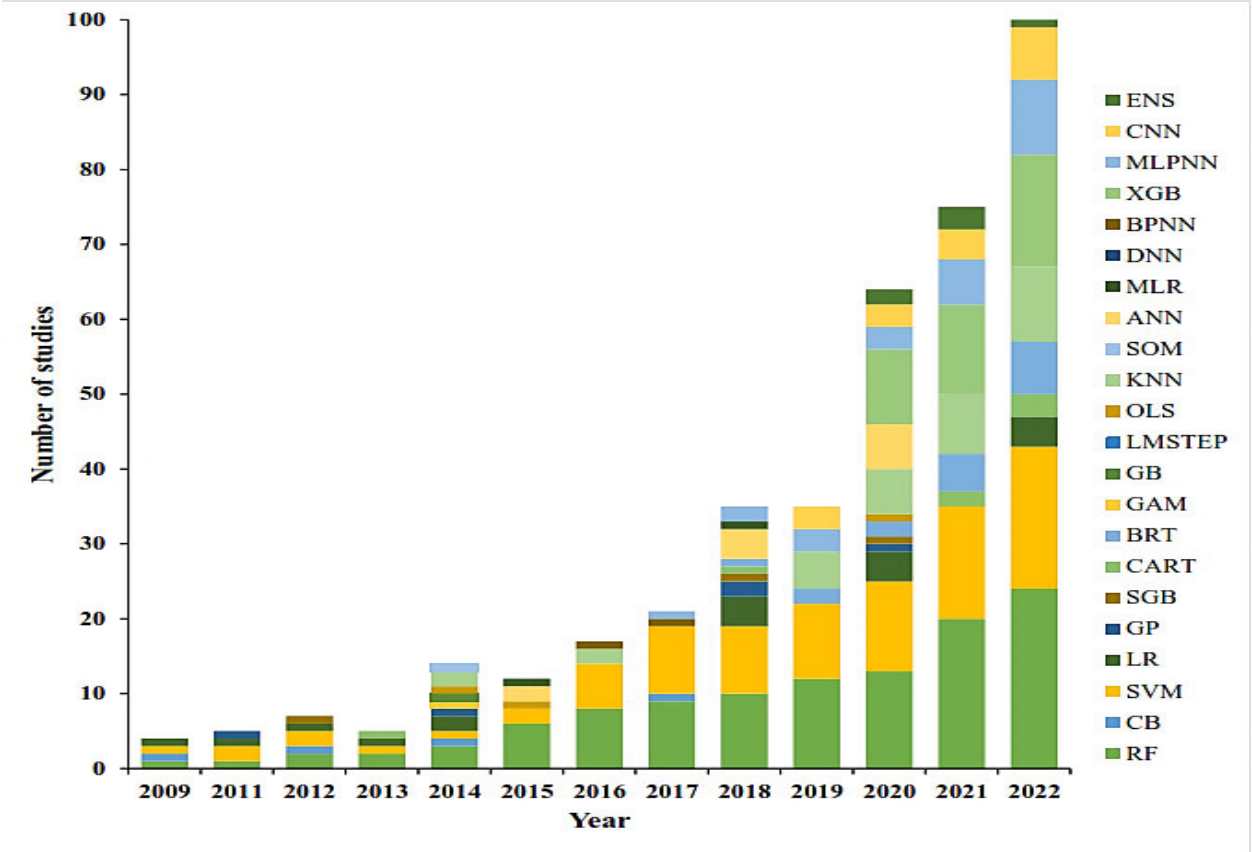


Figure 2.7 Progression of ML algorithms within the reviewed studies between 2009 and October 2022. Abbreviations in the chart represent the following algorithms ensemble (ENS), convolutional neural networks (CNN), multi-layer perceptron neural network (MLPNN), extreme gradient boosting (XGB), Back-propagation neural network (BPNN), deep neural network (DNN), multiple linear regression (MLR), artificial neural networks, self-organizing map (SOM), k-nearest-neighbor (KNN), ordinary least squares regression (OLSR), linear multi-stepwise regression (LMSTEP), gradient boosting (GB), generative additive models (GAM), boosted regression trees (BRT), classification and regression trees (CART), stochastic gradient boosting (SGB), gaussian processes (GP), linear regression (LR), support vector machine (SVM), cubist (CB) and random forest (RF).

2.3.6 Machine learning algorithm performance and accuracy

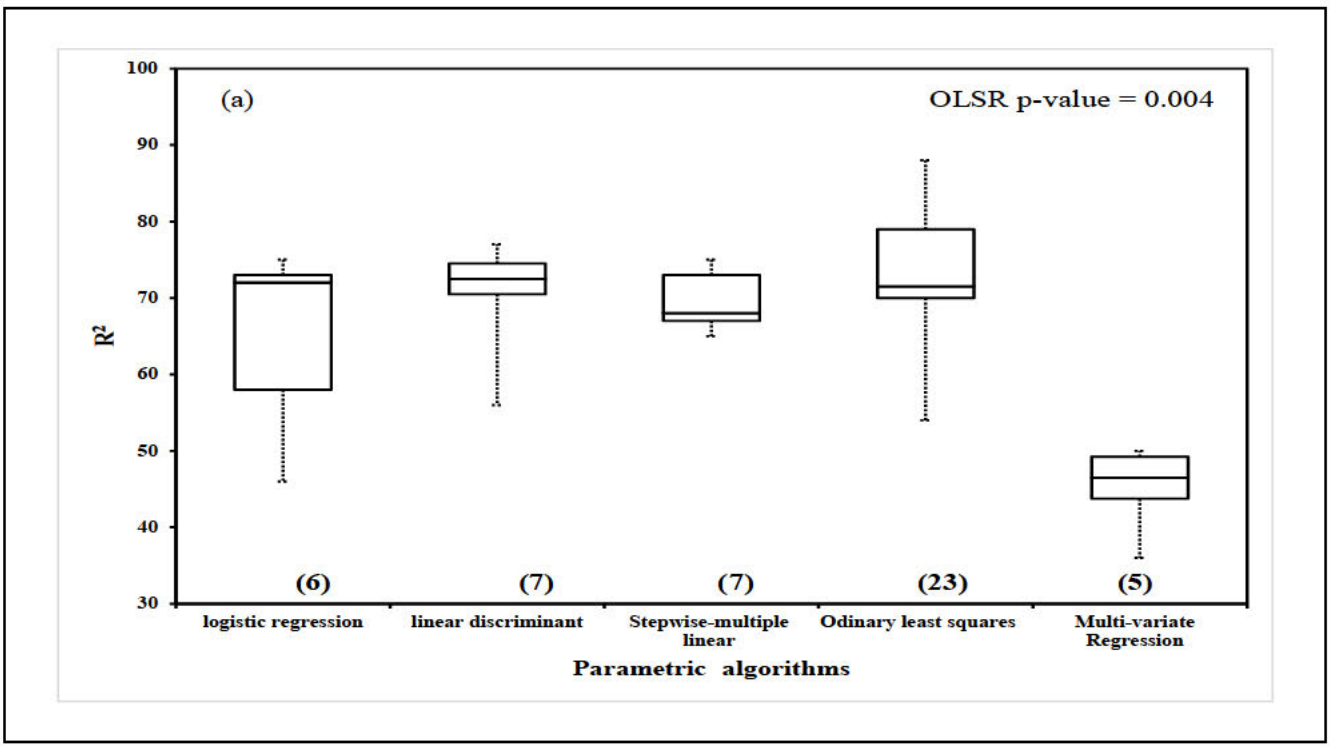
Machine learning algorithms are beneficial for estimating and mapping of AGC stocks in various natural forest types (Pham et al., 2020b, Zhou et al., 2020, Pacheco-Pascagaza et al., 2022). Accuracy assessment is an integral part of any modeling technique to determine a model's reliability in classification or regression (Li et al., 2019b, Zhang et al., 2020). Figure 2.8a shows the box and whisker plots of the average coefficient of determination accuracies of

studies that applied parametric ML algorithms for AGC storage predictions in natural forests. The box and whisker plots show the variance across the different forest types and RS data types used in the selected study sample. Among the five most used ML parametric algorithms, there were statistical differences in estimation accuracy (Figure 8a). For instance, the 95% confidence interval for the logistic linear regression accuracies was 15.49 to 64.8%. In other words, there is a 95% probability that the true mean accuracy of the linear regression models lies within this range. The mean accuracy of the linear regression models was 65.6%. The highest median R^2 was obtained by linear discriminant analysis (72.5%), followed by logistic linear regression boxplot (72%) which represented three algorithms including linear regression and multiple linear regression, stepwise multiple linear (68%), and ordinary least squares regression (71.5%). Multi-variate linear regression had the lowest median R^2 value (46.5%). A t-test of independent samples indicates that ordinary least squares regression has a statistically significantly higher R^2 than all other parametric algorithms p-value = 0.004 (Figure 2.8a).

Statistically significant differences were observed among the 14 most used non-parametric ML algorithms in the study sample (Figure 2.8b). For example, the RF and CB plots were comparatively taller than all other plots, suggesting that scholars have obtained differing accuracies in estimating of AGC/AGB stocks in natural forests. The 95% confidence levels of the two plots RF and CB (9.08 and 7.76) are relatively higher than other algorithms, SVM (4.26), XGB (4.07), CatBOOST (2.99), K-NN (5.05), SGB (4.93), GAN (2.71), Stacking-ENS (1.37), and CNN (4.73). RF and CB have the widest interquartile ranges, showing dispersion in the accuracies obtained by various scholars in AGC/AGB stock estimates. There was a smaller interquartile range for CatBOOST, GAN, stacked-ENS, ANN, and BRT, which indicates that most scholars obtained similar accuracies, and the values were generally high. The median R^2 values of SVM (76%) and XGB (81.5%) were within the RF interquartile range, signifying no significant statistical difference.

Despite the small differences in the means and median accuracies of the ML frameworks, there were numerous asymmetries in terms of variability. For instance, although RF has the highest number of applications (n=68), it also has the widest range in performance, that is, the difference between the first or third quartile and the median. Multi-variate linear regression has a much lower number of applications (n=6) and a much smaller range of estimation

performance. Moreover, the study sample demonstrated the strength of non-parametric algorithms over parametric ones.



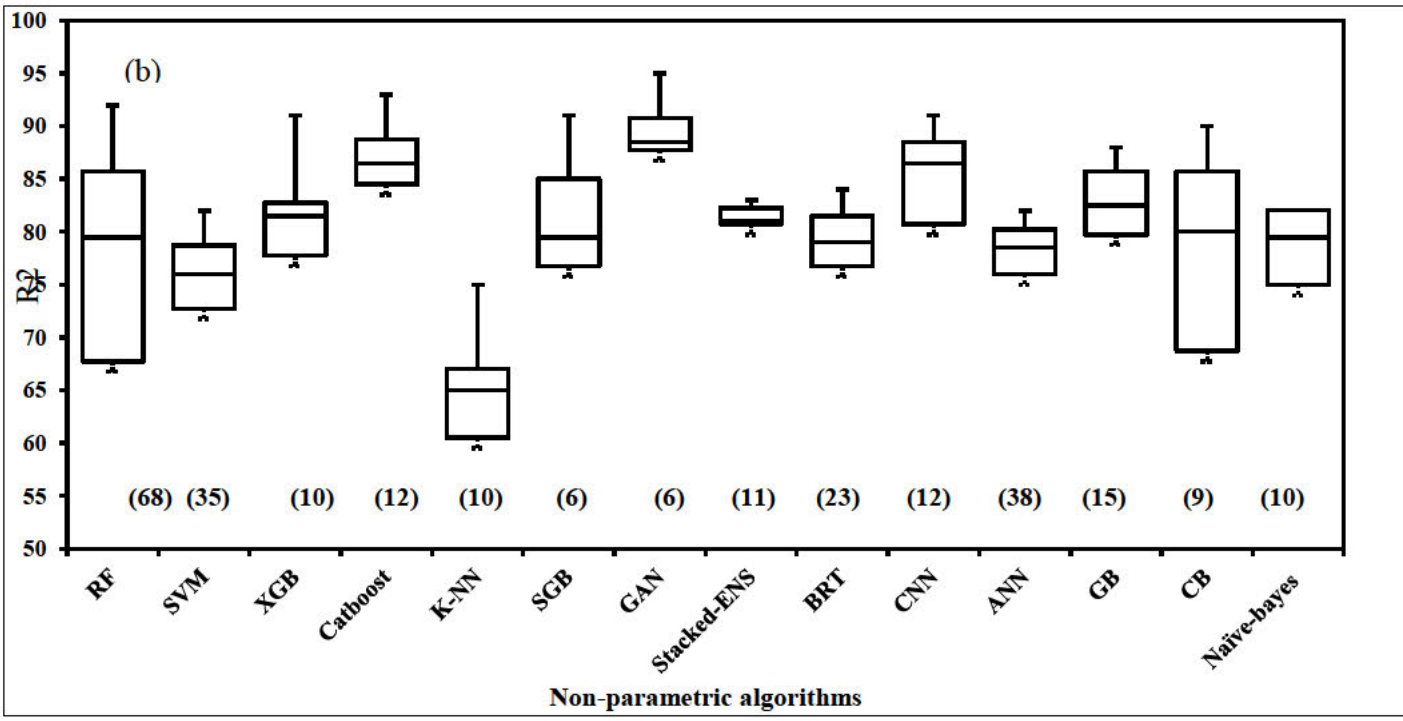


Figure 2.8 Box and whisker plots showing the average coefficient of determination (^{R2}) values produced by machine learning algorithms applied to remote sensing studies for the estimation

of AGC stocks in natural forests. Numbers in bold represent the number of studies that used the algorithm (a) logistic regression, linear discriminant, stepwise multiple linear, Ordinary least squares regression (OLSR) which is combined with linear and multiple linear regression studies , multivariate linear regression (b) random forest (RF), support vector machine (SVM), extreme gradient boosting (XGB), categorical boosting (CatBOOST), k-nearest neighbor (KNN), stochastic gradient boosting (SGB), generative additive networks (GAN), stacked ensemble (Stacked-ENS), boosted regression tree (BRT), convolutional neural network (CNN), artificial neural network (ANN) which is a combination of neural networks(NN),multi-layer perceptron neural networks (MLPNN) and perceptron, gradient boosting (GB), and cubist (CB).

Gaps /Limitations

Gap/Limitation	**Potential Studies/Methodologies to Address**
Lack of multi-sensor data fusion	Integrate multiple remote sensing datasets (e.g. optical, radar, LiDAR) using ensemble/stacking machine learning models
Limited transferability of models across sites	Develop domain adaptation techniques to transfer knowledge from data-rich to data-poor regions
Uncertainty in field measurement data	Employ data cleaning, exploratory analysis and robust loss functions to handle noisy training data
Lack of temporal analysis	Utilize time-series remote sensing and recurrent neural networks to model carbon dynamics
Black-box nature of some ML models	Explore interpretable machine learning methods (e.g. rule ensembles, feature importance) for insights
Scalability to large areas	Distribute model training across high-performance computing clusters or leverage cloud platforms
Lack of operational systems	Develop end-to-end pipelines integrated with field data collection and monitoring programs

2.4 Discussion

Natural forests are an important resource in managing and mitigating carbon accumulation; hence, methods to keep track of any changes in carbon storage in natural forests are intrinsic. Traditional methods of forest assessment alone are not capable of keeping track of carbon changes in natural forests, as such, remote sensing and machine learning approaches can keep a timeous record of natural forests by providing information on forest parameters (Halme et al., 2019a, Zhang et al., 2019b). This review investigates common machine learning methods applied to remote sensing datasets for aboveground carbon storage estimation in natural forests between 2009 and 2022. Furthermore, we reviewed the remote sensing sensors, applied algorithms, and predictors used in 245 articles. Notably, the popularity of remote sensing combined with machine learning has surged by an impressive 533% since 2009, as indicated by the trend in e publications. A considerable number of articles have been published between 2020 and 2022. Advances in remote sensing imagery across all platforms have enabled the steady rise of studies and the adoption of methodologies. The rising environmental problems, such as climate change, has contributed to a greater need for solutions and policy directions.

2.4.1 Sensor data

Remote sensing methods, including optical (e.g., Landsat, Sentinel-2, Hyperion), radar (e.g., ALOS PALSAR, Sentinel-1), and LiDAR (e.g., GEDI) sensors, have proven effective for estimating aboveground carbon (AGC) storage over large areas (Cao et al., 2016, Zhang et al., 2020). These techniques allow for the capture of spatial variations in biomass and carbon stocks, making it possible to monitor remote areas reliably (Goetz and Dubayah, 2011, Safari et al., 2017). Therefore, they are increasingly used to estimate forest AGC. Compared with traditional field-survey measurement approaches, remote sensing techniques have several advantages for estimating AGC, including the ability to estimate AGC at different scales (Zhang et al., 2019c, Shen et al., 2021). Various types of remote sensing data, such as radar, LiDAR, and optical data, are available for predicting and mapping AGC stocks in forests. Satellite-based approaches have the potential to provide “wall-to-wall" observations of carbon stock proxies and estimations over large areas using a uniform method in a short time (Pham

et al., 2019, Lu et al., 2016). However, errors can occur due to mismatches between field plots and satellite pixels (Réjou-Méchain et al., 2019, Qi et al., 2019).

This review highlights the predominance of optical sensors in estimating forest biophysical parameters owing to the availability of freely accessible remote-sensing datasets. Low-, medium-, and high-resolution imagery, such as MODIS, Sentinel, Landsat, and Hyperion, have been widely used for natural forest carbon storage estimation (Mngadi et al., 2022b, Dube et al., 2018, Gyamfi-Ampadu and Gebreslasie, 2021). Landsat and Sentinel are the most popular sensors in this field, possibly because of their consistent global coverage and longer spectral ranges that can better detect vegetation characteristics (Lu et al., 2016, Dube et al., 2014a, Ahmad et al., 2021). The increasing sensitivity of Landsat sensor bands to vegetation has contributed to their popularity among researchers (Dube and Mutanga, 2015, Labrecque et al., 2006). Multispectral data from these sensors combined with ML algorithms have been maximally utilized to extract information for image analysis tasks. Compared to multispectral data, hyperspectral data can accurately identify objects of interest, enabling the accurate estimation of forest biophysical parameters, such as AGB/AGC stocks (Popescu and Hauglin, 2014, Barbosa et al., 2014, Lu et al., 2016). Studies have used field-based methods and multispectral or hyperspectral data and ML algorithms, to estimate forest AGB and AGC stocks. Passive multispectral sensors have limitations in capturing forest vertical structural attributes, but their successful combination with active sensors has contributed significantly to improving natural forest carbon storage estimation studies (Abdel-Hamid et al., 2018, Pontius et al., 2020, Dainelli et al., 2021).

Remote sensing systems that rely on optical data have limitations caused by cloud cover and daylight requirements, which decrease the number of potential revisits (Mitchell et al., 2017, Lu et al., 2017, Jiang et al., 2021). However, newer technologies, such as radar systems and LiDAR, can provide data day and night, making it possible to penetrate clouds and reduce the chances of cloud cover (Wulder et al., 2012, Zhang and Ni-meister, 2014, Gyamfi-Ampadu and Gebreslasie, 2021). Traditional image processing and analysis techniques may not be accurate when dealing with newer remote sensing datasets since technology has evolved. Hence, machine learning methods are important ongoing research topics for providing adequate methods, in line with recent remote-sensing data (Sudha and Aji, 2019, Safari et al., 2017, Dube et al., 2014a).

Some high spatial resolution data often lack the shortwave infrared band, which is crucial for estimating the forest carbon storage (Kempeneers et al., 2011, Halme et al., 2019a). However, research acknowledges that fusion with other sensors such as SAR and LiDAR can reduce and effectively address this limitation (Mitchell et al., 2017, Montesano et al., 2013). Optical imagery is suitable for forest area mensuration, vegetation health monitoring, and forest classification but may have limited correlation with forest AGB/AGC after canopy closure. In contrast, synthetic aperture radar and LiDAR sensors can penetrate forest crowns, estimate trunk biomass under foliage cover, and are not influenced by atmospheric or temporal conditions. SAR and LiDAR sensors capture abundant information about forests that is typically laborious to gather, enabling machine learning to handle their characteristic nonlinear relationships. As a result, these sensors are widely used for estimating forest carbon storage. For instance, machine learning algorithms, such as RF and SVM, have been applied to estimate biomass in the natural forests of China using ALOS PALSAR data (Zhang et al., 2020, Pham et al., 2020b). By leveraging machine learning algorithms such as artificial neural networks (ANN) and support vector regressions (SVRs) researchers have utilized polarimetric synthetic aperture radar (SAR) data to estimate forest biomass and carbon (Almeida et al., 2019, Zhang et al., 2022d, Ghosh and Behera, 2018). This method was successfully applied to assess aboveground biomass (AGB) in natural mangrove forests across West Africa, using Sentinel 1 SAR data (Navarro et al., 2019, Forkuor et al., 2017).

Although there has been a significant amount of research on using machine learning and remote sensing for predicting and monitoring forest carbon storage, several aspects of this approach have been underexplored in the literature (Goetz and Dubayah, 2011, Liu et al., 2011). One aspect is the integratng of multisource remote sensing data for forest carbon storage prediction and monitoring. Although some studies have used multiple data sources, such as combining LiDAR and radar data (Tian et al., 2021), there is still room for the integration of different types of remote sensing data for a more comprehensive analysis (Baccini et al., 2004, Zhang et al., 2019c). For instance, integrating satellite-based remote sensing data with aerial or ground-based data can enhance the accuracy and precision of the predictions (Heenkenda et al., 2014, Reiersen et al., 2022) Integrating multiple remote sensing data sources can enhance forest carbon monitoring by providing a more comprehensive understanding of the spatial and temporal variations in forest carbon stock (Lu et al., 2017, Lourenço, 2021).

2.4.2 Spectral features (texture metrics, vegetation indices and image transformations)

This review highlights that the widely used vegetation indices, image texture features, and image transformations are all important covariates for forest carbon storage estimation. Vegetation indices are calculated from spectral bands, image texture measures spatial variability, and image transformations improve image quality or extract features (Eckert, 2012). Each method has its advantages and disadvantages, so the best choice depends on the application. However, combining methods can improve accuracy and understanding of factors affecting forest biomass or carbon storage (Tian et al., 2021, Uniyal et al., 2022). Studies have used texture measurements with optical and synthetic aperture radar (SAR) data for biomass estimation. Sarker and Nichol (2011) explored the potential of optical imagery using ALOS AVNIR-2 texture indices for biomass estimation. They obtained a significant improvement when using the ratio of texture parameters. also obtained similar results in the estimation of forest carbon and biomass while using the texture measurement from WorldView-2 satellite data. In another study by Cutler et al. (2012), a combination of SAR image texture and LANDSAT TM data and vegetation indices was used to estimate tropical forest biomass. This study suggests that SAR texture and vegetation indices can be used to predict biomass at different times and locations.

2.5 Machine learning for carbon storage estimation for natural forests

Machine learning is responsible for creating algorithms used to train models to make predictions and decisions based on data input. Recent advances in pattern recognition and computer vision have led to the emergence of new techniques (Campesato, 2020, Gyamfi-Ampadu and Gebreslasie, 2021). This has improved the accuracy, visualization, and estimation of carbon storage in natural forests compared to generic remote sensing approaches (Li et al., 2019b, Zhang et al., 2023). The recent application of machine learning algorithms and data integration to aboveground carbon storage prediction has provided insight into various aspects of natural forest ecosystems, especially those characterized by complex data dimensionality (Li et al., 2019b). A major advantage of ML methods is their ability to estimate complex nonlinear relationships from large amounts of data in extremely fast (Campesato, 2020, Pham et al., 2019). Furthermore, most ML frameworks do not assume data distribution. Several ML methods have been applied to studies related to forest biomass or carbon storage prediction, from simple logistic regression to cutting-edge ML methods, such

as generative additive models GAM, CNN, or CatBOOST (Safari et al., 2017, Uniyal et al., 2022, Pham et al., 2020a).

Choosing the appropriate ML algorithm for AGB/AGC stock estimates depends on several factors, including data size, application purpose, diversity, and quality (Halme et al., 2019b, Li et al., 2019b). Additional considerations include the training time, output interpretability, and accuracy (Campesato, 2020). Hence, choosing the best ML algorithm for the task (i.e., prediction and mapping of AGC) requires experimentation, time, and specifications (Pham and Yoshino, 2017, Kempeneers et al., 2011). It is therefore important to note that each ML algorithm has its strengths and weaknesses in any application, and using more than one technique can provide greater insight than isolated use.

This study reviewed the maturity, readiness, and performance of ML algorithms in estimating natural forest biomass and carbon storage. The results show that non-parametric algorithms tend to have higher estimation accuracy than parametric algorithms. For instance, the highest R^2 value was obtained using CatBOOST (0.94), whereas the highest median R^2 value (0.90) was obtained for CNN. The lowest R^2 value was obtained using multi-variate linear regression (0.46), but the average R^2 of the RF applications was one of the highest overall values (0.80). These results are consistent with the existing literature (Gyamfi-Ampadu and Gebreslasie, 2021), which noted that non-parametric algorithms were superior in performance compared to parametric algorithms (López-Serrano et al., 2016a, López-Serrano et al., 2016b, Dube et al., 2014b). Models based on data have certain limitations. An accurate result depends heavily on the quantity and quality of data and the relationship between the training dataset and the outputs for the study region (Fassnacht et al., 2014, Hosseini et al., 2021); outliers and erroneous values in the training data may adversely affect the performance of the model (Pham et al., 2018). There can be computationally demanding and suboptimal solutions when defining ANNs and SVRs, as well as choosing the kernel function for classifying tree species using multispectral (Shang and Chisholm, 2013, Chaves et al., 2020, Zhang et al., 2018) and hyperspectral imagery (Koedsin and Vaiphasa, 2013, Heenkenda et al., 2014, Wong and Fung, 2014, Wang et al., 2020), as well as SAR data (Pham et al., 2020a, Santi et al., 2020, Li et al., 2020c). The different machine learning algorithms commonly applied in natural forest carbon and biomass estimation include random forests (RF), artificial neural networks (ANN), convolutional neural networks (CNN), support vector machines (SVM), decision trees (DT), extreme gradient boosting (XGBoost), and k-nearest neighbors (k-NN),

which are continuously updated with technological advancement(Li et al., 2020c, Pham et al., 2018, Uniyal et al., 2022). Machine learning techniques have proven more accurate when combined with other metrics such as spectral and textural metrics (Sarker and Nichol, 2011, Mngadi et al., 2022a). Machine learning techniques that can be applied to estimate natural forest carbon storage can be grouped into non-parametric and parametric algorithms. In this study, we examined the effectiveness of machine learning algorithms in predicting aboveground forest carbon storage. Our review identifies interesting facts and critical aspects related to these algorithms. We observed that these methods have become increasingly popular among practitioners, but often do not consider their interpretability in diverse ecosystems. In other words, many researchers have applied these algorithms without fully understanding how they operate in different environments. Furthermore, we found a lack of clear evidence to help understand the accuracy of each algorithm for different ecosystems. In our review, we noticed that different studies used different methods to interpret the models they have developed. For example, some studies only report the coefficient of determination. In contrast, others include additional measures such as the root mean square error (RMSE), mean absolute error (MAE), mean squared error (MSE), and mean absolute percentage error (MAPE). Consequently, it is difficult to compare the accuracy of different algorithms or draw clear conclusions about their effectiveness. Therefore, more research is needed to establish the best practices for interpreting machine learning models to predict aboveground forest carbon storage.

2.5.1 Parametric

Several studies have applied parametric regression algorithms to remote-sensing datasets to assess carbon storage in natural forests. This category of machine learning refers to common statistical regression, where algorithms within the parametric category apply a limited number of assumptions or parameters that are not dependent on the number of available training samples (Tamiminia et al., 2020, Salunkhe et al., 2018).

Linear or multiple linear regression algorithms are commonly used in the assessment of natural forest carbon storage, assuming a linear relationship between the dependent variable (aboveground biomass/ carbon) and independent variables (textural image features and spectral image bands) (Liu et al., 2019). However, choosing appropriate variables that can represent carbon / biomass is crucial as natural forest AGB/AGC is typically affected by

many factors such as species diversity and tree height (Pham et al., 2019). Remote sensing data are often difficult to express using a simple linear regression algorithm (Lourenço, 2021, Lu et al., 2016). Therefore, various derivatives of simple linear regression, such as multiple linear regression (MLR), stepwise linear regression, ordinary least squares regression (OLSR), and partial least squares regression (PLS), have been developed.

In the review, 48 studies used parametric algorithms to estimate forest carbon storage. For instance, linear models have been applied to assess natural forest biomass in Cameroon's savannah forests using ALOS-PALSAR SAR data (Mermoz et al., 2014). In China, Landsat and ALOS-PALSAR data have been integrated to estimate biomass in mixed forests using linear models. However, parametric algorithms have certain shortcomings, they are constrained to a specified form, which can limit their flexibility and make them less suitable for complex problems (Laurin et al., 2018b). Moreover, parametric algorithms are based on assumptions about the underlying data distribution, which may not be true in practice. This can lead to poor fitting of the model to the data and inaccurate predictions (Anaya et al., 2009, Eckert, 2012, García-Gutiérrez et al., 2016, Hame et al., 2013, Laurin et al., 2018b). The review noted that they remain extremely useful and crucial in estimating natural forest biomass for carbon storage quantification due to their simplicity and general accuracy. However, they are less popular among researchers because of their linear nature, which affects their accuracy.

Comparisons have also been made between parametric and non-parametric algorithms for the assessment of forest carbon or biomass in natural forests. Many scholars have found it useful to compare the two types of algorithms, and in some instances, parametric algorithms have presented better results than non-parametric methods (Castillo et al., 2017, Luo et al., 2020, Zhang et al., 2020). However, most studies indicate that non-parametric algorithms present better and more accurate results in the estimation of natural forest carbon storage because of their non-linear nature (Almeida et al., 2019, López-Serrano et al., 2016a). Parametric algorithm models proved to be faster than non-parametric algorithm models when processing forest carbon estimation (Forkuor et al., 2017).

2.5.2 Nonparametric

This systematic review analyzed 245 publications that applied machine learning algorithms for carbon storage assessment in natural forests. The review noted that most studies preferred

non-parametric algorithms because of their flexibility and ability to handle large datasets without prior assumptions. Random Forest was the most commonly used algorithm, followed by SVM/SVR, XGBoost, CatBoost, ANN, and CNN. Recent research has focused on analyzing the factors that affect the accuracy of ML algorithms, such as the spatial and temporal resolution of remote sensing data, and the trade-offs between accuracy and computational cost. Non-parametric tend to perform better than parametric algorithms, with random forest and boosting algorithms such as CatBoost, XGBoost and BRT showing promising results. However, the accuracy can vary depending on the forest and RS data used. SVM/SVR had a median overall accuracy of 0.85, whereas ANN (0.82) was precise, but computationally expensive. CNN has a high accuracy of 90% but requires a large number of covariates. CatBoost outperformed RF and improved the AGB estimation using Landsat 8 OLI imagery in China (Li et al., 2020b). Sentinel-1, Sentinel-2, and ALOS-PALSAR are the most commonly used remote sensing datasets noted in this review. To improve accuracy and reliability, researchers should consider the specific context of each study when selecting an ML algorithm and conduct in-depth analyses of the factors affecting accuracy.

The study examined various non-parametric machine learning algorithms and found that specific algorithms require more training than others. For example, the XGBoost and SVM algorithms require less data for accurate predictions (Pham et al., 2020b). The study also found that advanced techniques such as stacked ensemble learning significantly improved forest carbon estimation by combining several base learners to contribute to the final meta-learner (Uniyal et al., 2022, Chan et al., 2022). The study noted that 11 other studies have reported success and high prediction performance from ensemble stacking (Jafarzadeh et al., 2021). The precision and variety of base learners affect the performance of a stacking algorithm in addition to their diversity and accuracy (Singh et al., 2022, Dang et al., 2019). While adding more base learners does not always result in more accurate predictions, it requires additional memory usage and computing time. This study suggests that stacking three to four base learners together is suitable for achieving accurate predictions.

Generalized additive models (GAMs) are useful for estimating forest carbon stocks because of their ability to model nonlinear relationships between variables (Chen et al., 2022a, Björk et al., 2020). Research has shown that GAMs can accurately estimate carbon stocks in different forest types and account for the effects of spatial autocorrelation (Björk et al., 2022, Chen et al., 2022a). The advantage of GAMs is that they can incorporate complex

relationships between predictors, without needing prior knowledge of the functional form. However, their effectiveness depends on the quality and quantity of the input data, as well as the appropriate selection of covariates (Wood, 2017). Additionally, GAMs can be computationally intensive and may require more time and resources to train than other simpler models (Campesato, 2020). Despite these limitations, GAMs remain valuable tools for forest carbon estimation and management.

The review noted that the use of machine learning models to predict aboveground biomass and belowground biomass and soil organic carbon was under explored (Wang et al., 2022b). Most studies have focused on aboveground biomass, neglecting the significant contributions of belowground biomass and soil organic carbon to forest carbon storage (Salunkhe et al., 2018, Zhang et al., 2019c). Furthermore, there is need for more research on the transferability of machine-learning models across different regions and forest types for forest carbon storage estimation. Most studies have been conducted in specific regions and forest types, making it difficult to generalize these findings to other regions or forest types. However, valuable insights onto the transferability of machine-learning models have been provided by studies that have applied machine learning on forest growing stock volume (GSV) estimation (Tanase et al., 2020). Most studies on machine learning for GSV estimation have been conducted in specific regions and forest types, making it difficult to generalize these findings to other regions or forest types. These studies have shown that the similarity of the training and test data is the most important factor affecting the transferability of machine-learning models (Cosenza et al., 2022, Tompalski et al., 2019). They have also shown that support vector regression models are more transferable than other machine learning algorithms (Awad et al., 2015).

The statistical differences in aboveground biomass estimation between algorithms stem from several factors. These factors include the type of remote sensing data used, where single-sensor data may not capture the full variation, while multi-sensor data provides more comprehensive information (Vincent and Saatchi, 1999, Xi et al., 2023). Additionally, the sensor's spatial resolution captures more forest stand detail, and its temporal resolution tracks changes in aboveground biomass over time (Zhang et al., 2019b, Myeong et al., 2006). The model's accuracy depends on the quality and representativeness of the field training data collected (Fassnacht et al., 2014). Environmental conditions at the study site, such as climate, soil, and topography, influence aboveground carbon storage (Odebiri et al., 2022). Moreover,

the size, age, and type of forest stand being estimated impact the results, necessitating a model designed specifically for the forest stand type. Furthermore, the history of forest management and the forest's response to climate variability, climate change, and human disturbance also play a role in aboveground biomass variation.

This comprehensive review highlights the powerful capabilities of machine learning coupled with remote sensing data for predicting forest biomass and carbon stocks. Non-parametric algorithms like Random Forest and gradient boosting models have proven particularly effective, with incorporating vegetation indices, texture features, and terrain factors further boosting predictive performance. The fusion of multi-sensor data streams (optical, LiDAR, radar) capitalizes on complementary information sources, yielding superior biomass mapping accuracies compared to single-sensor approaches. However, challenges remain in further improving model precision, efficiency, interpretability, and distilling predictions into actionable intelligence for policymakers and forest managers. Emerging research avenues include leveraging deep learning's potential for even higher accuracy given sufficient training data, developing hybrid techniques that synergistically combine physics-based and machine learning models, and evaluating new satellite sensor capabilities (e.g. Sentinel, Landsat 9, WorldView) and UAV platforms for fine-scale biomass mapping across diverse forest ecosystems. Prioritizing understudied regions like Africa is critical for comprehensive global carbon monitoring. Enhancing model interpretability and quantifying uncertainty will build trust for practical implementation. Ultimately, translating research prototypes into robust, efficiently updated operational systems at regional/national scales is a key objective for harnessing these technologies to support sustainable forest management and climate change mitigation policies through improved biomass/carbon accounting capabilities.

2.5.3 Limitations

Machine learning algorithms can offer promising solutions for accurately estimating the aboveground biomass and carbon in natural forests. However, challenges related to remote-sensing sensors, such as the lack of spectral variation between different tree species and the shadow effect from canopies, can introduce errors in the estimation process. Additionally, the absence of well-distributed field data further complicates the production of accurate and standardized carbon storage estimates. Systematic reviews have identified the limitations of both parametric and non-parametric algorithms for biomass estimation, with parametric

algorithms being less suitable for complex environments with diverse features, and non-parametric algorithms requiring more training data and computational power. Moreover, the limited accessibility of certain datasets such as LiDAR, Hyperion, and RapidEye can pose a challenge. Despite these challenges, the application of machine learning algorithms for estimating aboveground biomass in natural forests in the South African context remains limited.

2.5.4 Critical Analysis and Research Gaps

While the literature review presents a thorough examination of the existing research, several notable gaps and limitations can be identified. A common trend observed is the predominant focus on estimating aboveground biomass (AGB) and aboveground carbon (AGC), with limited exploration of belowground biomass and soil organic carbon, which are significant contributors to overall forest carbon storage. Additionally, most studies have been conducted within specific regions and forest types, limiting the transferability and generalization of the developed models to other geographical areas or forest ecosystems. Furthermore, there is a lack of clear consensus on the interpretability and accuracy assessment of different machine learning algorithms across diverse ecosystems, hindering the selection of the most appropriate algorithm for a given study area. Moreover, the accessibility and availability of certain remote sensing datasets, such as LiDAR, Hyperion, and RapidEye, remain a challenge, potentially limiting the accuracy and comprehensiveness of the estimations. Finally, while the literature review provides a global perspective, there is a notable gap in the application of these techniques within the South African context, particularly for estimating aboveground carbon storage in natural forests. These identified gaps and limitations highlight the need for further research to address the existing shortcomings and contribute to a more comprehensive understanding of forest carbon storage estimation using machine learning and remote sensing techniques.

2.6 Conclusions

This paper provides a comprehensive review of machine learning and remote sensing techniques for forest carbon storage prediction. The review evaluates different algorithms, features, and factors in predicting forest carbon storage, and highlighting the impact of vegetation indices and texture features on prediction accuracy. The results indicate that non-parametric machine learning approaches outperform parametric methods, with RF and

gradient boosting-based (XGBoost, CatBOOST) models being the most effective for forest carbon storage prediction. Vegetation indices, texture and terrain-factors are the most commonly used features for prediction. Factors influencing forest carbon storage prediction are selected based on their correlation with other factors. The review also identifies challenges in using machine learning approaches and remote sensing for forest carbon storage prediction, such as improving model accuracy, the time taken, interpretability, providing accurate information to policymakers and practitioners, and addressing the black box problem. The paper highlights the importance of natural forests and the use of remote sensing techniques to estimate their carbon storage services. The review recommends using machine learning algorithms for estimating natural forest biomass or carbon. It highlights the benefits of emerging sensors like Sentinel, Landsat 8 OLI, Landsat 9, IKONOS, Planetscope, and Worldview for natural forest biomass assessment studies. The paper calls for more research on natural forest biomass and carbon in Africa, which is critical for carbon sequestration. The review suggests that deep learning is a promising approach that may require more data but provides higher accuracies than conventional machine learning approaches. Hybrid approaches that combine physically based models with remote sensing data can complement the estimation process.

Acknowledgement

This study was supported by the UKZN funded by the Wood RIGHTS Programme fund and the DST-NRF SARChI Chair in Land Use Planning and Management at UKZN **(**Grant **No.** 84157 and 114898**)**.

2.7 Summary

The literature review reveals significant research on utilizing remote sensing and machine learning to monitor and map forest carbon accumulation and storage across predominantly natural ecosystems. However, key knowledge gaps persist regarding the carbon contributions and accumulation rates of reforestation projects, especially concerning rapidly developing urban areas. While studies have quantified biomass and carbon densities for commercial plantations and intact forests, assessments of more complex reforestation initiatives are lacking. Additionally, optimized integration of multi-source satellite data and modern algorithms has yet to be demonstrated to accurately map urban forest carbon sequestration

and storage. Therefore, Chapter 3 assesses aboveground biomass in a reforested urban landscape using machine learning and remotely sensed data.

3 CHAPTER THREE: Assessing Above-ground biomass in Reforested Urban Landscapes Using Machine Learning and Remotely Sensed Data.

This chapter is based on:

Matiza, C., Mutanga,O., Peerbhay, K.,Lottering,R. Odindi,J., 2023. Assessing Above-ground biomass in Reforested Urban Landscapes Using Machine Learning and Remotely Sensed Data. *Journal of Spatial Science. (Accepted)*

Abstract

Forest loss and degradation are major contributors to climate change. Hence, urban reforestation is a promising nature-based solution for mitigating these effects. However, it is essential to accurately estimate aboveground biomass to quantify carbon gains and offsets in reforested urban environments. In this regard, this study sought to estimate carbon sequestration within a reforested urban landscape. The study used four remote sensing image datasets (i.e., Planetscope, Sentinel-1A, Sentinel-2A, and Shuttle Radar Topography Mission) and in-situ field measurements to build aboveground biomass estimation regression models based on non-parametric machine learning algorithms, such as k-nearest neighbour, support vector machines, extreme gradient boosting, and random forests. Thirty-nine predictor features were used to generate a spatial distribution of aboveground biomass and carbon density maps. The study results show that the extreme gradient boosting model performed the best, achieving a predicted aboveground biomass of 4.1-286.5t ha-1 for minimum and maximum values of AGB. This suggests that extreme gradient boosting models can effectively model reforested aboveground biomass using multi-source data. The findings of this study have several implications for urban forestry and carbon sequestration research. Firstly, they suggest that extreme gradient boosting models are promising for estimating aboveground biomass in reforested urban landscapes. Secondly, they highlight the importance of using multi-source data for this task, as this can improve the accuracy of the models. Thirdly, they demonstrate the potential of machine learning in addressing important environmental challenges, such as climate change. Overall, this study provides valuable insights into accurately estimating aboveground biomass and carbon sequestration in reforested urban landscapes. The findings of this study have the potential to improve the management of urban forests and mitigate climate change. The use of non-parametric machine learning algorithms and multiple remote sensing image datasets makes this study a

significant contribution to urban forestry and carbon sequestration.

Keywords: *Urbanization, climate change, urban forests, nature-based solutions, reforestation, machine learning*

3.1 Introduction

Urban areas are a significant contributor to global warming and climate change, as evidenced by various studies (Kumar and Shiva Nagendra, 2016, McCarthy et al., 2010). Anthropogenic activities like burning of fossil fuels, deforestation, and rapid urbanization contribute to the progressive accumulation of greenhouse gases, notably carbon dioxide, in the atmosphere (Ramachandra et al., 2015). In fact, urban areas account for over 75% of CO^2 emissions from fossil fuel combustion (Seto and Christensen, 2013). Urban expansion, especially in tropical zones, also plays a significant role in annual emissions from land-use change (Seto and Christensen, 2013, Churkina, 2016). Additionally, urban areas exacerbate other environmental problems such as air, water, and noise pollution, as well as urban heat islands and high ozone levels (Zhao et al., 2019a, Odindi et al., 2012, Xu et al., 2016). As the urban population is forecasted to increase significantly in the next 50 years, strategies to curb carbon emissions in urban areas have become increasingly important (Sithole et al., 2018, FAO, 2015). This has led to a greater emphasis on tracking the causes of global climate change and the role of urbanization in upsetting the carbon balance at both regional and global scales (Cleugh and Grimmond, 2011, Zhao et al., 2019a). It is therefore imperative to focus on mitigating carbon emissions in urban settings to minimize the ecological impact of urbanization.

Urban reforestation (UR) has emerged as an important approach to mitigate rising carbon emissions in urban areas (Den Besten et al., 2014, IPCC, 2019, Mngadi et al., 2022a). In addition to capturing 0.4% of the carbon stored in forests worldwide, urban forests also have significant impacts on local biophysical systems (Raciti et al., 2014, Hutyra et al., 2014). In his regard, Reducing Emissions from Deforestation and Forest Degradation (REDD+) has identified reforestation as the most efficient way to lower carbon footprint and ameliorating environmental transformations, particularly in urbanized contexts (Mulatu et al., 2019). Urban reforestation also provides a range of ecosystem services, including the regulation of key environmental parameters such as wind patterns, humidity levels, and solar radiation. In addition, UR delivers diverse health benefits and recreational prospects to urban inhabitants (Endsley, 2018, Buyantuyev and Wu, 2010, Nowak et al., 2006, Crompton, 2005). Consequently, reforestation initiatives within urban cities are a major infrastructural

component of any urban ecosystem combined with sustainable living strategies such as wind, solar, and other renewable energy innovations (Mngadi et al., 2021a, Lucke and Beecham, 2019, Dobbs et al., 2014, Sousa-Silva et al., 2023). These approaches can be useful in mitigating the urban carbon footprint. Various studies support the efficacy of urban reforestation as an effective tool for mitigating carbon emissions in urban settings (Mngadi et al., 2021a, Escobedo et al., 2011, Sousa-Silva et al., 2023, Nowak et al., 2014). Therefore, it is imperative that such initiatives be given priority in the quest to reduce carbon emissions in urban areas.

However, while studies on other forest landscapes including natural, commercial and conserved forests exist (Li et al., 2020c, Dube and Mutanga, 2015, Navarro et al., 2019), there is a dearth of literature on the quantity of carbon sequestered within a peri-urban area that has been restored with forest cover (Uniyal et al., 2022, Mngadi et al., 2021a), particularly in terms of its impact on above-ground biomass. In addition, although many city governments and authorities have begun to embrace approaches that promote tree-planting and UR, there is still a significant knowledge gap regarding urban forests' ability to reduce atmospheric CO^2 concentrations and mitigate global warming, particularly in resource constrained zones like Africa (Velasco et al., 2016, Demuzere et al., 2014). This is primarily due to: (1) the limited number of reforestation sites, (2) the type of data and auxiliary information used in biomass/carbon simulation, and (3) the effectiveness and accuracy of interpolation algorithms. Reforestation initiatives in peri-urban areas may not receive the same level of funding, incentives or oversight compared to high-profile urban greenspace projects or large-scale rural reforestation efforts (Hedblom et al., 2017, Nevzati et al., 2024). The interface between urban and rural policy domains can create governance gaps. Technically, the heterogeneity of peri-urban landscapes with mixed land use patterns like informal settlements alongside remnant forests can introduce modeling complexities(Nevzati et al., 2024). Nonetheless, as the call for UR within urban settings persists globally, a cost-effective and robust modeling framework to accurately monitor the aboveground biomass AGB of these reforested zones is also becoming essential for evaluating the gains and progress of reforestation initiatives.

In recent years, remote sensing (RS) has become essential valuable tool in earth observation technology due to advancements in sensor technology (Lu et al., 2016, Weng, 2012, Odebiri et al., 2020a, Li et al., 2020c). Remote sensing data has found diverse applications in various fields, including long-term climate studies, forest structural analysis, and disaster mitigation (Hamida et al., 2018). Field-based estimates of above-ground biomass (AGB) carbon in urban reforestation (UR) are accurate but limited in scope and costly (Xu et al., 2016). On the other

hand, when combined with geostatistical or machine learning models, analysis of RS data provides a cost-effective way to assess reforested AGB carbon and its dynamics in large areas with limited accessibility (Ojoyi et al., 2016). Multispectral medium to high resolution images such as Sentinel-2A, Planetscope, and RapidEye provide better predictions of AGB due to their ability to capture complex urban surface objects (Xiao et al., 2022, Li et al., 2019a). To compensate for the limitations posed by different types of sensors, scholars have integrated sensors to retrieve AGB in different forest types (Li et al., 2020c, Zhang et al., 2022d, Zolkos et al., 2013, Xiao et al., 2022, Nandy et al., 2017). Integration of different sensors, such as Sentinel-1, Sentinel-2, and topographic data received less attention in reforested urban sites, however, many scholars have used Sentinel-1 and Sentinel-2 and topographic data to enhance AGB prediction in other forest types (Laurin et al., 2018a, Zhang et al., 2023). Additionally, radar remote sensing can obtain structural information about trees without interference from weather or lighting conditions, making it popular in biomass-related studies in sub-tropical areas with a high incidence of cloud cover and haze (Akhtar et al., 2020). Analysis of RS data by experts provides a way to monitor and measure reforestation progress and its impact on carbon emissions, making it an essential tool in mitigating carbon emissions in urban environments.

Several statistical models have been developed to relate field-measured aboveground biomass (AGB) carbon to remote sensing (RS) variables (Zheng et al., 2004, Cunha-Lignon et al., 2009, Nuthammachot et al., 2018). AGB carbon is strongly correlated with vegetation density, terrain, and climatic variables, all of which can be obtained using multisource remotely sensed datasets. Spectral data together with relevant vegetation indices (VIs) defined by mathematical combinations of reflectance values in the visible, near-infrared (NIR), and shortwave infrared (SWIR) regions of the electromagnetic spectrum as well as topographic variables , are the most widely used for retrieving AGB carbon (Odebiri et al., 2022, Tamiminia et al., 2022, Li et al., 2015). However, some statistical models are parametric and cannot account for the non-linear relationships between RS data and field data (Halme et al., 2019b, Main-Knorn et al., 2011). While optical spectral and radar data may saturate at high biomass densities, LIDAR data does not saturate. This makes LIDAR data a valuable tool for measuring biomass densities that are much higher than those that can be measured with other types of RS data (Tanase et al., 2014). Non-parametric machine learning (ML) models, on the other hand, have grown in popularity due to their ability to reveal complex non-linear patterns as well as address data dimensionality issues when fitting models with a substantial number of features (Pham et al., 2020b, Uniyal et al., 2022, Zhang et al., 2019c). The study employs

several powerful machine learning techniques including Support Vector Machines (SVM), Random Forests (RF), Artificial Neural Networks (ANN), k-Nearest Neighbors (kNN), and Extreme Gradient Boosting (XGBoost) for biomass prediction. These models are chosen for their ability to effectively capture non-linear relationships between input remote sensing features and the target biomass values. SVMs can handle high-dimensional data and are insensitive to outliers (López-Serrano et al., 2016b, Ramdani and Furqon, 2022). RFs are ensemble learning methods robust to overfitting. ANNs excel at discovering intricate patterns in data. kNN is a simple yet effective non-parametric algorithm. XGBoost is a highly efficient and parallelizable implementation of gradient boosting that can optimize arbitrary differentiable loss functions (Ramdani and Furqon, 2022). The diversity of these models allows capturing complementary aspects of the data, potentially improving overall prediction accuracy for this challenging regression task in the context of peri-urban reforestation monitoring.

In this regard, leveraging ML methods on the abundance of RS data including Planetscope, Sentinel-2 and Sentinel-1 could provide significant insight for continuous and reliable reforested peri-urban biomass/carbon predictions, subsequently informing policymakers and reforestation managers on the status of initiatives. To this end, the study sought to quantify the aboveground biomass distribution in a peri-urban reforested site using multisource image datasets and different ML frameworks.

3.2 Materials and methods

3.2.1 Field data collection

Field data was conducted during the summer of 2019 (between February and March) to collect data on forest biomass. Prior to the data collection, 194 sample points were randomly created in ArcGIS and uploaded to a GPS for easy navigation to the sampling locations. A sample size of 194 plots was chosen to balance accuracy and practicality in estimating aboveground biomass. This selection aimed to achieve the desired level of accuracy of ninety-five percent. We employed statistical power analysis considering expected variability and limitations like spatial heterogeneity, and limited finances. The samples were strategically distributed to represent the diversity of the study area using random sampling. Furthermore, the study utilized these points as the basis for data collection during the survey. At each randomly selected location, a 10 x 10 m plot-size window was established, and the structural

dimensions of reforested trees, including tree height and diameter at breast height (DBH), were measured and recorded. The study employed a clinometer to measure tree height and a caliper to measure diameter at breast height (DBH) of the trees. The location of each sampled tree was recorded using a Trimble Global Positioning System with an accuracy of 0.5 m. These data collection activities provided valuable information necessary for forest biomass calculation. Subsequently, allometric computations were performed to estimate aboveground biomass based on the measured tree characteristics. The resulting biomass estimates were later used to calibrate the ML models used in the estimation of forest biomass over the study area.

The Intergovernmental Panel on Climate Change (IPCC) has recently validated allometric models as a non-destructive technique that can be adopted to compute biomass (Shukla et al., 2019). Literature shows that the allometric equation, which relates tree diameter to height, is a critical factor in the provision of biomass estimates which are a prerequisite in the training of models for estimation of tree biomass, making it possible to calculate in-situ biomass based on these measurements (Latifah et al., 2021, Chave et al., 2014). Therefore, this study integrated tree height and DBH into an allometric model (Eq. 3.1) (Latifah et al., 2021, Chave et al., 2014) to compute aboveground biomass.

$$AGB = 0.0673 \times \left(p^{D^2H}\right)^{0.976} \quad \text{Eq. 3.1}$$

Where p represents wood density constant (1.2 g/cm^3), D for diameter (cm) and H indicates height (m).

While the study followed a robust methodology for field data collection and biomass estimation, it is essential to acknowledge and address potential sources of error or bias. These may include measurement errors due to instrument calibration or observer bias, sampling biases from inadequate representation of spatial or species diversity, uncertainties in allometric models, errors in GPS positioning, and temporal variability in tree growth or biomass dynamics. Mitigating these issues could involve implementing stringent quality control measures, optimizing sampling designs, developing site-specific or species-specific allometric models, utilizing high-precision positioning techniques, and incorporating temporal adjustments. By addressing these potential sources of error and bias, the study can enhance the accuracy and reliability of the biomass estimates, ultimately improving the robustness of the modelling efforts.

3.2.2 Image acquisition and preprocessing

This study utilized various Earth observation datasets including Sentinel-1A (S1A), Sentinel-2A (S2A) Multispectral Instrument, Planetscope, and Shuttle Radar Topography Mission Digital Elevation Model (SRTM-DEM) for reforested above-ground biomass estimation. Sentinel-1A operates at a C-band frequency with VV and VH polarizations for Level-1 Ground Resolution Distance. On the other hand, Sentinel-2A has 13 bands ranging from ultra-blue to SWIR with spatial resolutions from 10-60m. Planetscope is a high-resolution 4-band image with a spectral range of blue, green, red, and near-infrared, with a 3.7m resolution and 1-day revisit time. Lastly, the Shuttle Radar Topography Mission Digital Elevation Model is a 30m resolution image obtained from single pass SAR interferograms provided by the National Geospatial Agency and Jet Propulsion Laboratory.

S1A, S2A MSI and Planetscope were acquired on the 6th of February 2019. A Planetscope MSI level-3B image with 5% cloud cover was downloaded on November 15, 2022, from the Planetscope Explorer portal (https://www.planet.com/explorer/). The S1A and S2B MSI were downloaded from the Google earth engine (https://code.earthengine.google.com/) on 16 September 2022, and SRTM-DEM accessed on 17 September 2022. It is worth noting that Sentinel-2, Sentinel-1, and SRTM are freely available, while Planetscope is a commercial satellite.

The Sentinel 1A image in Google Earth Engine underwent several preprocessing steps to ensure its accuracy and usability. These processes included applying the satellite orbit correction for precise positioning, radiometric calibration of intensity values (VH and VV) to obtain actual backscatter coefficients, noise reduction via speckle filtering, and sensor geometry registration into projection coordinates through terrain correction. Additionally, the backscatter values were converted into decibel (dB) units for consistent representation. A level-2A product from Sentinel-2 was acquired and the product is accessible after orthophoto radiometric, geometric, and atmospheric correction have been applied. Nevertheless, the level-2A product does not apply cloud filtering and this was accomplished by using the sen2cor function in Google Earth Engine (GEE). GEE also homogenized the spatial resolution of the Sentinel-2 MSI to 10 m using the nearest neighbor technique. Planetscope image products are already radiometrically and atmospherically corrected, hence no image pre-processing was carried out in this study. The SRTM GL1 version is supplied by google earth engine corrected for all model errors and is considered the definitive version of the SRTM-DEMs (Mouratidis et al., 2010).

Image transformations

Vegetation indices were statistically derived from Sentinel- 2 MSI and Planetscope MSI using the raster calculator embedded in Quantum GIS 3.28.8 software. The vegetation indices used in this study, representing wet green biomass were calculated using pre-existing formulas and respective image wavebands. Eight vegetation indices were selected for each sensor (Planetscope and Sentinel-2A). Additionally, the vegetation indices used in the study were a combination of traditional and red edge indices. For a comprehensive list of the vegetation and red edge indices used, please refer to **(**Table 3.1**),** In addition to spectral information, topographic variables such as elevation, slope and topographic wetness index were extracted from a 30m SRTM-DEM in QGIS 3.28.8 software since topography affects the distribution of green biomass. To ensure consistency, all datasets were resampled to a 10x10 meters resolution using nearest neighbor interpolation technique in QGIS 3.28.8 software. Furthermore, they were clipped to the same extent, retaining only the data within a predefined region of interest. Finally, the datasets were projected to UTM Zone 36S with the WGS84 datum to achieve spatial alignment. These preprocessing steps allowed for seamless comparison and analysis across all datasets, ensuring a standardized and accurate representation of the data.

Table 3.1.Overview of predictors used.

Predictor variable	**Formulae**	**Data source**	**Sources**
NDVI	$\frac{(NIR - Red)}{(NIR + Red)}$	Sentinel-2A, Planetscope	(Pettorelli, 2013)
EVI	$2{,}5 \times \frac{(NIR - Red)}{(NIR + 6 \times Red - 7.5 \times Blue + 1)}$	Sentinel-2A	(Jiang et al., 2008)
ARVI	$\frac{NIR - Red - y(Red - Blue)}{NIR + RED - y(Red - Blue)}$	Sentinel-2A	(Kaufman and Tanre 1992
MSAVI	$\frac{2NIR + 1\sqrt{(2NIR + 1) - 8(NIR - Red)}}{2}$	Planetscope	(Qi et al., 1994)
EVI-2	$2{,}5 \times \frac{(NIR - Red)}{(NIR + Red + 1)}$	Sentinel-2A,	(Jiang et al., 2007)
GRVI	$\frac{NİR}{GREEN}$	Sentinel-2A	(Silleos et al., 2006)
MCARI	*[(RedEdge – Red) – 0.2 × (RedEdge − Green)] × (RedEdge ÷ Red)*	Sentinel-2A	(Wu et al., 2008)
IRECI	$\frac{(RedEdgeIII - Red)}{(RedEdgeI/RedEdgeII)}$	Sentinel-2A	(Bhattarai *et al.* 2020)

reNDVI	$\frac{(RedEdgeIV - RedEdgeII)}{RedEdgeIV + RedEdgeII}$	Sentinel-2A	(Imran et al., 2020)
CIG	$\left(\frac{NIR}{GREEN}\right) - 1$	Sentinel-2A	(Wu et al., 2012)
CIRE	$\left(\frac{RedEdgeIII}{RedEdgeI}\right)$-1	Sentinel-2A	(Clevers and Gitelson, 2013)
NDREII	$\frac{(NIR - RedEdgeII)}{(NIR + RedEdgeII)}$	Sentinel-2A	(Imran et al., 2020)
GEMI	$\frac{(2 \times (NIR^2 - RED^2) + 1.5 \times NIR + 0.5 \times RED)}{(NIR + RED + 0.5)}$	Planetscope	(Pinty and Verstraete, 1992)
RVI	$\frac{NIR}{RED}$	Planetscope	
GNDVI	$\frac{(NIR - GREEN)}{(NIR + GREEN)}$	Sentinel-2A	(Imran et al., 2020)
SAVI	$\frac{(NIR - RED)}{(NIR + RED + 0.5)}(1 + 0.5)$	Planetscope	(Clevers and Gitelson, 2013)
MNDVI	$\frac{NIR - Red}{NIR + RED - SWIR}$	Sentinel-2A	(Gitelson et al., 1996)
OSAVI	$(1 + 0.16)\frac{(NIR - Red)}{(NIR + Red + 0.16)}$	Planetscope	(Qi et al., 1994)
Band 2-8, 11,12, 8A,	Raw bands without any calculation	Sentinel-2A MSI imagery	(Jutz and Milagro-Pérez, 2020)
Band 1-4	Raw bands without any calculation	Planetscope Dove	(Frazier and Hemingway, 2021)
VV, VH	Raw bands without any calculation	Sentinel-1A MSI imagery	(Jutz and Milagro-Pérez, 2020)
TWI	Topographic wetness index	SRTM-DEM	(Sørensen et al., 2006)
SLOPE	Slope	SRTM-DEM	(Sørensen et al., 2006)

3.2.3 Prediction techniques

In this study, four ML algorithms (XGBoost, RF, K-NN, SVM) were used to estimate and map aboveground biomass (AGB) in reforested urban landscapes. The algorithms have different capabilities in analyzing remotely sensed data, and the study employed a combination of remote sensing sources, including Sentinel-2A, Sentinel-1A, Planetscope, and

SRTM-DEM. The research consisted of several steps, including pre-processing, creating training and testing datasets, evaluating the models, selecting optimal variables, and re-evaluating the model for AGB estimation in the reforested urban area.

The selection of the XGBoost, Random Forest (RF), K-Nearest Neighbors (K-NN), and Support Vector Machine (SVM) algorithms for estimating and mapping aboveground biomass (AGB) in reforested urban landscapes was based on their proven performance in handling complex, non-linear relationships between predictor variables and the target variable (Ahmad et al., 2018). These algorithms have demonstrated robustness in dealing with remote sensing data and have been widely applied in biomass estimation studies across various forest types (Ramdani and Furqon, 2022). Specifically, ensemble techniques like XGBoost and RF have shown superior predictive accuracy compared to traditional regression models, while SVM and K-NN offer flexibility in handling high-dimensional data and capturing intricate patterns(Lu et al., 2017). A comprehensive review of the literature on biomass modeling in urban environments guided the choice of these algorithms, considering their strengths and suitability for the unique characteristics of the study area. The study's steps are detailed in a flowchart (Figure 3.1**).**

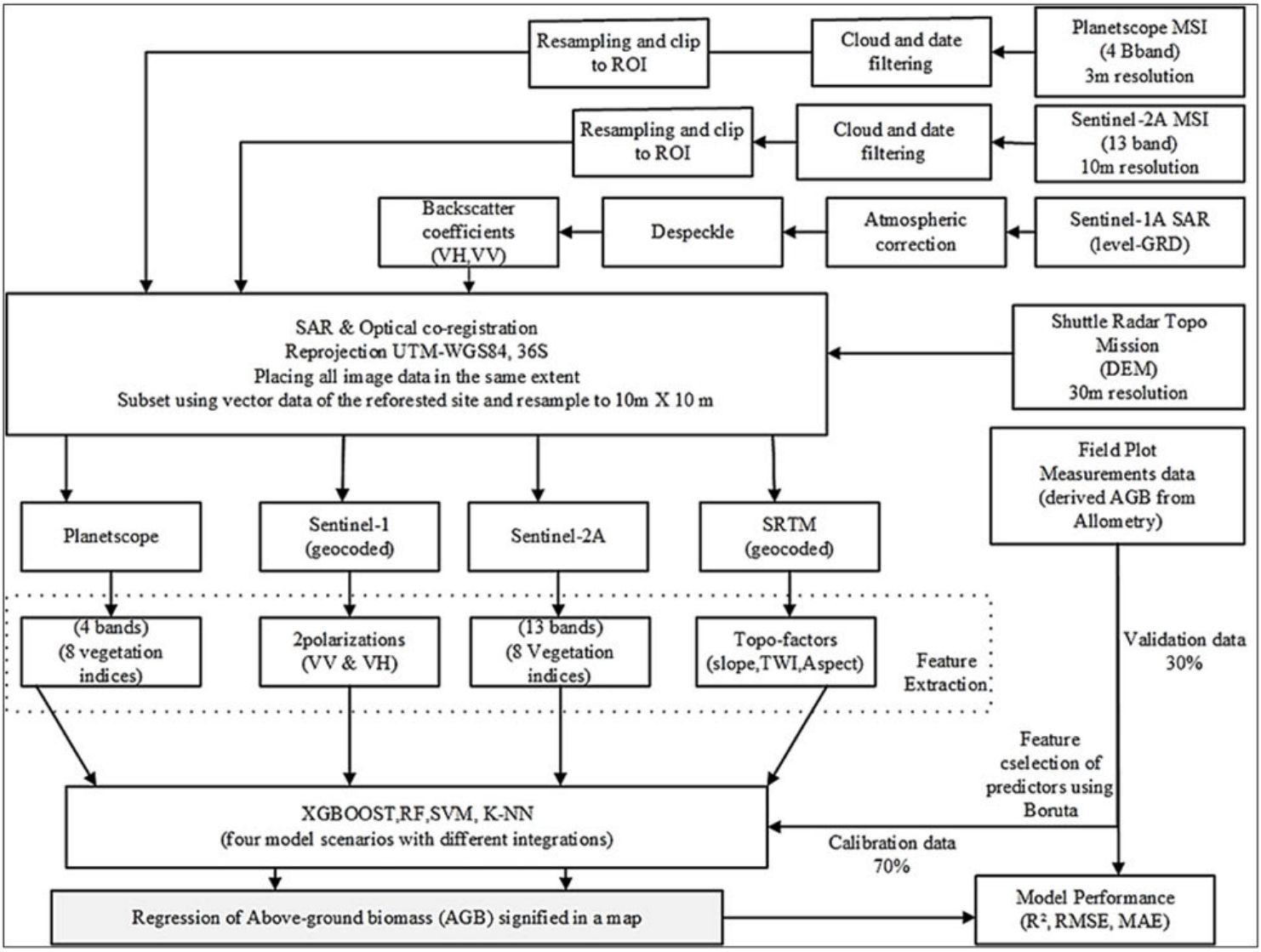


Figure 3.1 Flowchart illustrating the steps involved in processing the satellite images and developing modified AGB predictive models based on ML techniques.

3.2.3.1 Extreme gradient boosting (XGBoost)

XGBoost is an efficient and scalable boosting algorithm that has achieved widespread application in a variety of fields due to its ability to solve complex scientific problems (Chen and Guestrin, 2016a). The XGBoost algorithm utilizes ensemble decision tree learning to build trees in a sequential manner and improve their performance iteratively through gradient boosting(Chen and Guestrin, 2016a) . This makes it a popular machine learning (ML) approach for a variety of tasks, including estimating aboveground biomass and carbon stock(Tamiminia et al., 2022, Zhang et al., 2022d, Huang et al., 2022b).

XGBoost stands out for its efficiency, adaptability, and ability to mitigate bias errors. These qualities are particularly relevant for challenges in estimating aboveground biomass and carbon stock, where data is often sparse and noisy. To address these challenges, XGBoost uses early stopping and an additional regularization term in its training process to prevent overfitting(Chen and Guestrin, 2016a, Nguyen et al., 2022) . Hyperparameter tuning further enhances its performance, tailoring the solution to the unique aspects of the geospatial analysis.

In addition to its technical merits, the significance of the XGBoost algorithm extends to its versatility. XGBoost has found applications across a wide range of fields, including finance, healthcare, and natural resources management. In the field of carbon and aboveground biomass estimation, XGBoost's capacity to reveal underlying patterns solidifies its status as a promising tool for comprehending and quantifying complex ecological processes in reforested urban landscapes.

3.2.3.2 Random Forest (RF)

The Random Forest algorithm is a powerful ensemble machine learning approach that is well-suited for estimating aboveground biomass in reforested urban landscapes. The algorithm utilizes multiple binary decision trees to address the challenges posed by correlated and noisy predictor variables (Breiman, 2001). This is particularly relevant to our context as it enables us to navigate the complexities inherent in estimating aboveground biomass in urban landscapes.

The construction of decision trees within the Random Forest algorithm is underpinned by randomized training schemes and bootstrapping techniques, which collectively mitigate the risk of overfitting to our data. This is important as our focus is on robust predictions that account for potential noise and variability in the input data. An innovative aspect of the Random Forest algorithm lies in its utilization of a modified bagging approach, wherein a subset of features is randomly selected at each tree node split (Breiman, 2001, Gleason and Im, 2012). This introduces an element of controlled randomness, preventing the model from becoming overly reliant on any single feature and enhancing its capacity to generalize effectively.

The two primary optimization parameters, ntree and mtry, play a pivotal role in refining the regression model's predictive accuracy. To harness the full potential of the Random Forest approach, we engaged in a comprehensive search for optimal ntree and mtry values through a

grid search. The decision to select ntree and mtry based on the smallest root mean square error rate of the calibration dataset reflects our commitment to precision and reliability in our estimations (Belgiu and Drăguț, 2016). The ability to handle complex predictor variables, mitigate overfitting, and optimize parameter values through data-driven validation strategies positions it as a compelling technique for analyzing the intricate dynamics of AGB in reforested urban landscapes.

3.2.3.3 K-nearest neighbor (KNN)

KNN is a proximity-based algorithm that assumes that similar data points are located in close proximity to each other. It uses the Euclidean distance technique to categorize or predict each data point based on the collective consensus of its neighboring samples (Xiao et al., 2022, Sun and Huang, 2010, Taunk et al., 2019).. A pivotal hyperparameter, weight, gauges the influence of individual points on their neighboring counterparts(Jhamtani et al., 2021, Taunk et al., 2019) .

KNN's distinct attributes lend themselves inherently to the complexities of forest AGB estimation within urban landscapes. Due to its exceptional ability to manage non-linear relationships, it aligns with the complex nature of urban environments. Furthermore, KNN eliminates the need for pre-model training, which accommodates situations in which timely and comprehensive training data may be difficult to collect. An additional layer of refinement was introduced by the identification of the optimal k-value parameter crucial to KNN's performance (Xiao et al., 2022). The methodical determination of this k value through experimentation reinforces the meticulous approach adopted in this study, ensuring the balance between model complexity and generalizability.

3.2.3.4 Support Vector Machine (SVM)

The Support Vector Machine (SVM) algorithm is a powerful machine learning approach that is effective for nonlinear regression problems, such as AGB estimation, crop yield estimation and soil organic carbon prediction (Zhao et al., 2019b, Mountrakis et al., 2011). SVMs can handle complex relationships between predictor variables by mapping input vectors into a higher-dimensional feature space (Mountrakis et al., 2011, Chen et al., 2022c, Han et al., 2012). This allows SVMs to establish more accurate decision boundaries between vectors than would be possible in the original feature space. SVMs can also be optimized with support vectors to mitigate overfitting. This is important for AGB estimation, as the relationships between predictor variables may vary in different landscapes.

The versatility of SVMs is further extended by the incorporation of various kernel optimization parameters such as radial basis and sigma width (Tanveer et al., 2022, Mountrakis et al., 2011). These parameters can be tuned to customize the SVM to the specific variations of the dataset under analysis. Ideally, SVMs are not only technically capable, but they are also adaptable to diverse datasets and can yield accurate predictions even when confronted with limited training data. This adaptability positions SVMs as a potent and flexible regression tool. It is an appropriate approach for analyzing the complexities of the estimation of AGB within dynamic urban landscape.

3.2.4 Hyperparameter tuning.

During the construction of an ML model, optimizing hyperparameters is crucial for achieving the best regression performance. Hyperparameters control the learning process, and in this study, a grid search with validation was employed in R to tune and maintain these parameters during the calibration and validation phases for the XGBoost, RF, K-NN, and SVM algorithms. Grid search involves the selection of the best hyperparameter combinations from a grid of all possible values to maximize the model's predictive power. The 70-30 hold out validation technique was used to estimate the model's performance by partitioning the data into subsets, training the model on 70% of the dataset, and using the remaining 30% for validation. This process is repeated ten times, with each iteration serving as an opportunity to refine and improve the results.

3.2.5 Selection of optimal predictor variables.

Before constructing a ML model, it is essential to remove spurious data features and select the most relevant variables to reduce multicollinearity, complexity, redundancy, and improve accuracy (Odebiri et al., 2022, Mngadi et al., 2021a). Some techniques such as SVM and RF can handle this selection, but others such as K-NN may suffer from the Hughes effect due to an excessively large combination of variables (Souza et al., 2019). This issue frequently occurs in remote sensing data analysis due to the broad range of predictor variables available. In this study, we used the Boruta package in R to identify the optimal variables for aboveground biomass retrieval in reforested urban landscapes. Boruta is an all-encompassing embedded feature curation method that employs the RF algorithm wrapper technique to recognize both robustly and weakly pertinent spectra, thereby diminishing data complexity and ameliorating regression precision. (Agjee et al., 2016). Boruta's uniqueness lies in its use of statistical measures (Z-scores) to determine the importance of variables (Kursa and

Rudnicki, 2010, Agjee et al., 2016). Duplicate copies of independent variables, known as shadow features or permuted copies, are created and shuffled to eliminate correlations and redundancies with the target variable. Boruta's selection of variables may not always be characterized by a high RF mean decreases in accuracy, distinguishing it from other feature reduction algorithms for instance the recursive feature elimination (RFE) algorithm that selects subsets of ranked bands based on the mean decrease in accuracy specified by RF (Agjee et al., 2016, Rudnicki et al., 2015).

3.2.6 Model performance evaluation and validation.

The effectiveness of ML models to estimate AGB was evaluated using 10 repetitions of train-test split validation with 70% of the data in the training set and 30% in the validation set. The data (n = 194) was randomly split into a training set (n = 136) and a validation set (n = 58). The training set was used to optimize and train the models, while the validation set was used to assess the goodness-of-fit, accuracy, and reliability of the final prediction model. To assess the ability of ML algorithms to estimate AGB in a reforested urban environment, this study employed three performance metrics: the mean absolute error (MAE) Eq. 3.2, the coefficient of determination (R^2),Eq. 3.3and the root mean square error (RMSE) Eq. 3.4. The formulas for these statistical parameters are as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad \text{Eq. 3.2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \qquad \text{Eq. 3.3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \qquad \text{Eq. 3.4}$$

To further validate the results, model outputs were compared against existing biomass datasets and studies conducted in similar reforested urban environments. Additionally, ground truth validation was performed by collecting field measurements of tree diameter, height, and species information across multiple sites within the study area. The model predictions were compared against biomass estimates derived from locally calibrated allometric equations

using the field data. This validation step aimed to demonstrate the agreement between the model outputs and ground truth observations, with RMSE values within acceptable ranges for practical applications. The rigorous evaluation, comparisons, and ground truth validation collectively strengthen the reliability and applicability of the findings for accurate biomass estimation in reforested urban landscapes.

3.3 Results

3.3.1 Reforested aboveground biomass exploratory analysis.

The allometric equations and forest plot measurements used to estimate aboveground biomass within a reforested site reveal a range of 2.1 t ha-1 to 327.68 t ha-1. According to the descriptive statistics, the mean value of aboveground biomass is 96.84 t ha-1, with a standard deviation of 77.46 t ha-1.

3.3.2 Machine learning model optimization

Based on grid search, **Table 2** displays optimal hyperparameters. The XGBoost regression model was found to be relatively neutral to the gamma value as a result of the hyperparameter tuning procedure. Interestingly, when the minimum_child_weight and maximum depth is the same, the RMSE and R^2 values remain unchanged. The RMSE was calculated based on the validation set, and the optimal combination of hyperparameters was selected by minimizing the average RMSE across ten iterations. Once the optimal hyperparameters were identified for the AGB modeling algorithm, they were set for the model.

Table 3.2**.** Results of Grid search of the best hyperparameters for machine learning algorithms

Algorithm	**Parameter**	**Value**
XGBoost	*Min_child_weight*	1
	eta	0.3
	nrounds	200
	Learning_rate	0.8
	gamma	0
	subsample	1
	Max_depth	1
RF	*ntree*	500
	mtry	10

K-NN	*K_max*	9
SVM	*C_control*	500
	gamma	0.05

3.3.3 Specifying the optimal features of importance.

Results in Figure 3.2 show the selected top relevant variables according to the averaged importance score estimated by the statistical Z-score in the Boruta algorithm. Confirmed variables were labeled green while blue represented the shadow min and max values and the red color represented unimportant or rejected variables. Using a statistical z-score criterion, a total of 18 variables from the n=36 were selected including Planetscope-RVI, Planetscope-band 4, Planetscope-GEMI, Planetscope-SAVI, Sentinel2-NDVI, Sentinel2-EVI2, Sentinel2-ARVI, Sentinel2-MNDVI, Sentinel2-IRECI, Sentinel2-OSAVI, Sentinel2-GNDVI, Sentinel2-GRVI Sentinel2-CLRE, Sentinel2-band 7, Sentinel2-band 6, Sentinel2-band 4, Sentinel2-band 3, and Sentinel2-band 1 for optimum AGB quantification after 100 iterations. The integration of these variables into the four ML models produced the lowest RMSE of 49.63 t ha-1 for the fourth scenario of the XGBoost and a 55.40 t ha-1 RMSE validation. When all 39 variables were utilized the RMSE for the XGBoost fourth scenario increased to 65.63 t ha-1 (training) and RMSE validation was 52.30 t ha-1. The final scenario adopted the Boruta feature selected variables for the final comparison of the four regression models for quantifying aboveground biomass in the study site. By reducing the number of variables, Boruta feature selection generally results in a decrease in RMSE values.

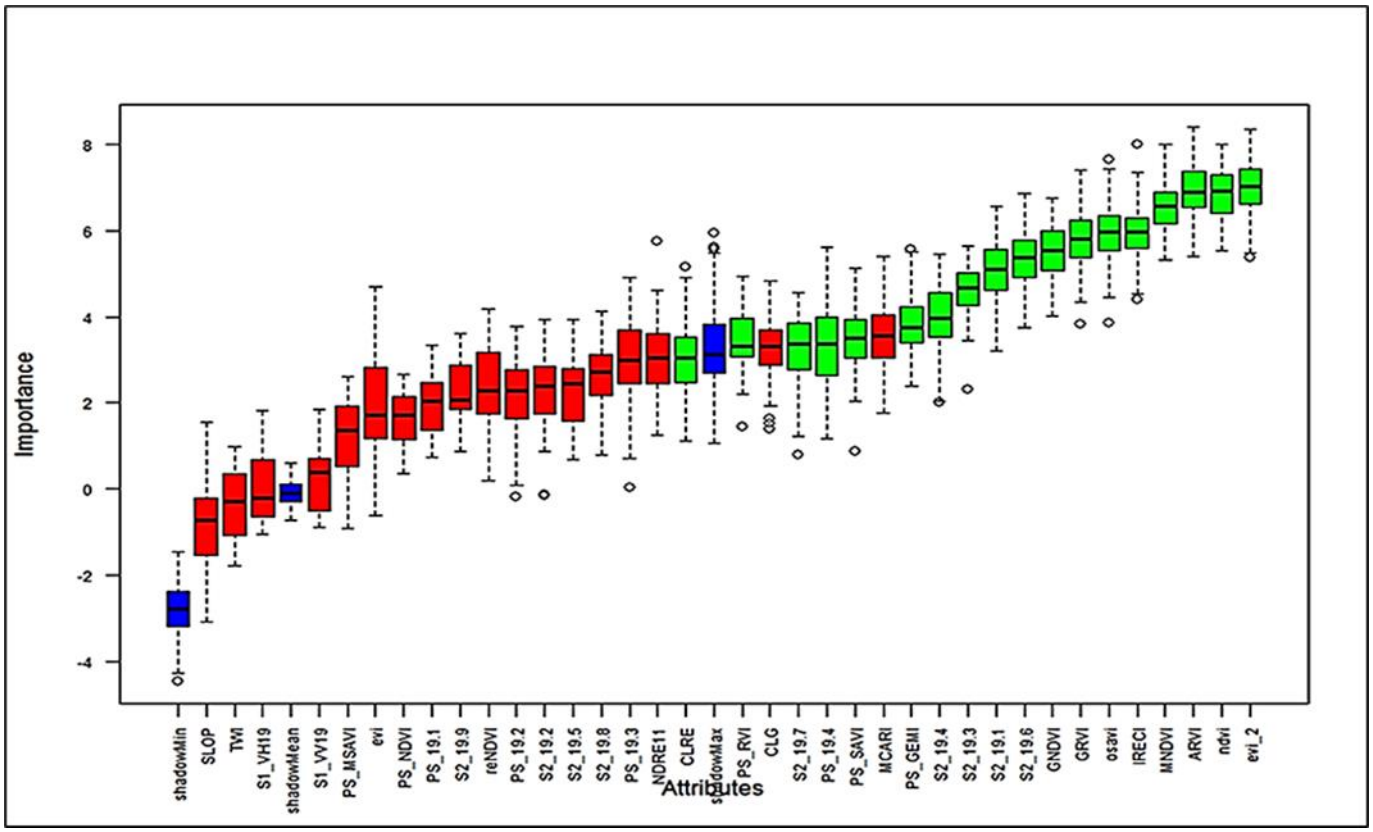


Figure 3.2. The optimal variables in estimating aboveground biomass using the Boruta algorithm. Important variables are labelled as Confirmed in green, unimportant as rejected in red and shadow min and max in navy blue.

For the fourth scenario that encompasses all variables,

Figure 3.3 illustrates the top eighteen selected features for XGBoost, RF, K-NN, and SVM regression based AGB quantification techniques, ranked by feature significance. XGBoost models rely heavily on one or a few predictors, whereas RF, SVM and K-NN models allow importance to be spread across a wider range of variables. Since AGB is predominantly found in woodlands, predictions from the AGB model are heavily influenced by in-situ data. Sentinel2-NDVI and Sentinel2-MNDVI were found in the top five of all four models, indicating the significance of these variables in biomass modeling.

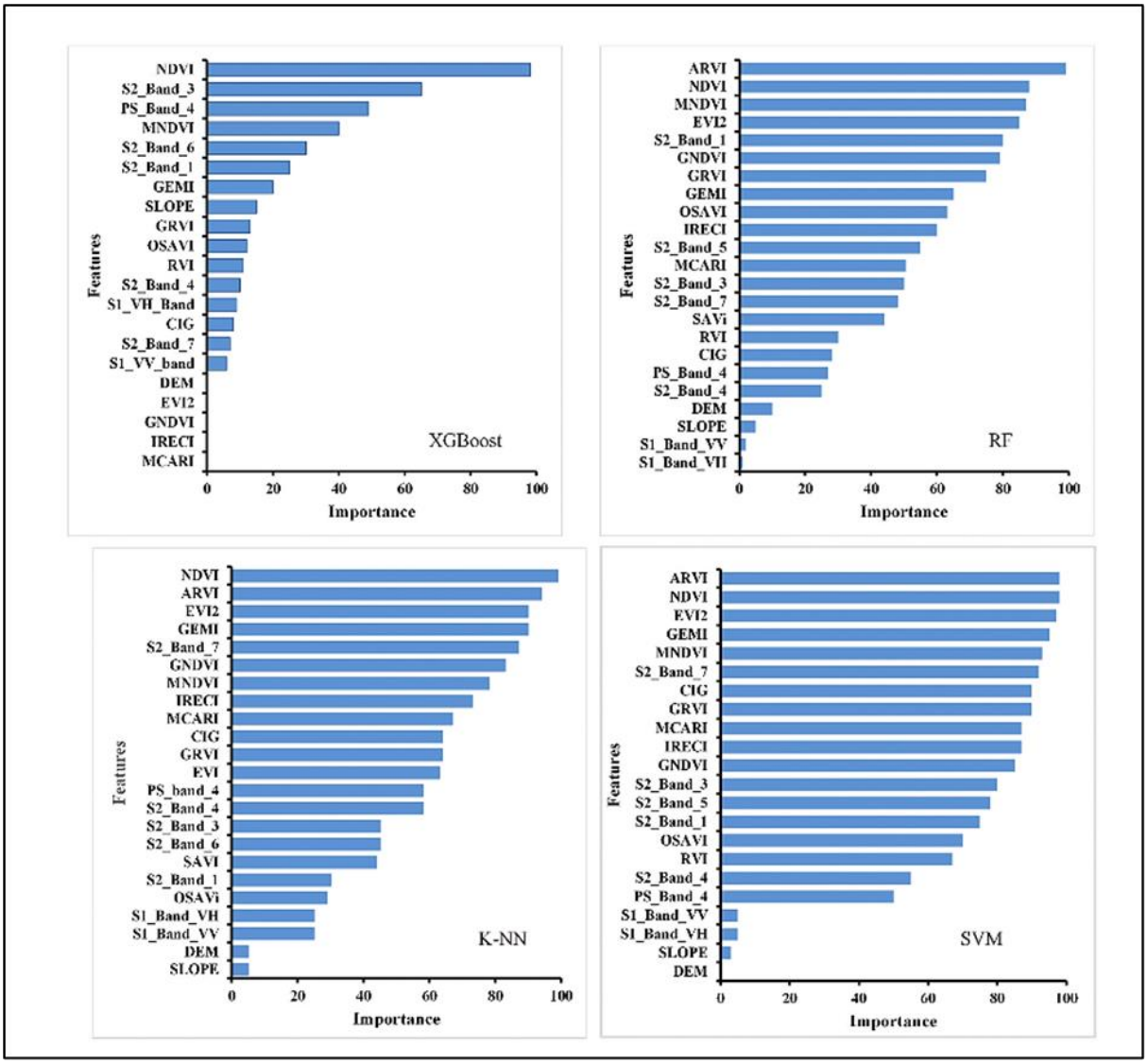


Figure 3.3 Variable of importance of the four models for the fourth scenario (that encompasses all variables.

3.3.4 Model performance, assessment and comparison.

In **Table 3**, we demonstrate the prediction performance of Sentinel-1, Sentinel-2, Planetscope, SRTM-DEM remote sensing spectral information, as well as four ML algorithms including XGBoost, RF, K-NN and SVM. The fourth model scenario, which used the optimal predictor variables selected by the Boruta feature selection algorithm, achieved the best overall performance. The RF algorithm achieved an average aboveground biomass (AGB) value of 52.43 t ha-1 in the calibration set and 53.43 t. ha-1 in the validation set. The XGBoost regression exhibited the lowest root mean square error (RMSE) of 49.63 t. ha-1 and 55.40 t. ha-1, the highest coefficient of determination (R^2) of 0.86 and 0.82, and the mean absolute

error (MAE) of 43.42 t. ha-1 and 45.40 t. ha-1 when using the 18 optimal variables rather than the entire set of 39 variables. RF followed with an RMSE value of 52.43 t. ha-1 and 53.43 t. ha-1, R^2 0.80 and 0.77, as well as MAE 42.44 t. ha-1 and 44.51 t. ha-1 using the calibration and validation set, respectively.

The integration of sensors and sensor derivatives including vegetation indices and topographic variables significantly affected the RMSE and R^2 of the models, for instance, the performance of raw bands of Sentinel-1, Sentinel-2A and Planetscope (scenario 1) for RF RMSE 72.26 t. ha-1 and 74.32 t. ha-1, R^2 0.79 and 0.76 as well as MAE 63.02 t. ha-1 and 67.22 t. ha-1 using calibration and validation was not as high as the performance of scenario 2,3 and 4 for the same algorithm (Table 3.3**)**. The subsequent changes in combinations and addition of more valid predictors coupled with Boruta feature selection resulted in a significant reduction in RMSE and an increase in the R^2 as noted by the performance of the fourth scenario with all input features. As a whole, all models were capable of predicting AGB with a reasonable degree of accuracy (Table 3.3**)**. Despite this, XGBoost proved to be significantly more effective in all four scenarios evaluated.

Table 3.3 Performance of the XGBoost, RF, K-NN and SVM model using different numbers of variables (bold values highlight the best-performing model variables (bold values highlight the best-performing model).

Modelling Algorithm	Model Scenario	Number of variables	RMSE (t. ha-1)		MAE (t. ha-1)		R^2	
			Calibration	Validation	Calibration	Validation	Calibration	Validation
XGBoost	Scenario I	Raw bands (S1+S2+PS)	70.82	73.63	60.58	63.69*	0.83	0.79
	Scenario II	Raw bands + topographic variables	70.34	73.29	60.41	63.43	0.84	0.79
	Scenario III	Vegetation indices + topographic variables	70.20	73.09	62.33	65.33	0.80	0.78
	Scenario IV	Raw bands + vegetation indices + topographic variables	**49.63**	**55.40**	**43.42**	**45.40**	**0.86**	**0.82**
RF	Scenario I	Raw bands (S1+S2+PS)	72.26	74.32	63.02	67.22	0.79	0.76
	Scenario II	Raw bands + topographic variables	72.18	74.87	62.94	65.04	0.81	0.75
	Scenario III	Vegetation indices + topographic variables	72.53	58.83	61.77	60.89	0.79	0.75
	Scenario IV	Raw bands + vegetation indices + topographic variables	52.43	53.43	42.44	44.51	0.80	0.77
SVM	Scenario I	Raw bands (S1+S2+PS)	73.63	76.45	62.68	65.43	0.77	0.70
	Scenario II	Raw bands + topographic variables	71.99	73.53	62.56	63.64	0.77	0.72
	Scenario III	Vegetation indices + topographic variables	63.14	66.31	49.62	58.03	0.73	0.67
	Scenario IV	Raw bands + vegetation indices + topographic variables	54.61	57.78	49.45	52.81	0.77	0.74
K-NN	Scenario I	Raw bands (S1+S2+PS)	74.85	76.82	64.06	65.31	0.75	0.66
	scenario II	Raw bands + topographic variables	74.52	75.91	63.51	64.44	0.80	0.77
	Scenario III	Vegetation indices + topographic variables	71.43	73.23	58.37	61.53	0.71	0.67
	Scenario IV	Raw bands + vegetation indices + topographic variables	55.63	58.31	48.91	52.42	0.75	0.66

In the fourth and final scenario of this study, a multisource dataset was used to illustrate the relationship between observed and predicted aboveground biomass (AGB) in a reforested urban environment (Figure 3.4**)**. The results showed a strong positive coefficient of determination ranging from 0.86 to 0.82 for XGBoost and 0.80 to 0.77 for RF between observed and predicted AGB. In contrast, the coefficient of determination for SVM and K-NN was relatively lower at 0.75 to 0.66 and 0.77 to 0.74, respectively, indicating an average positive relationship (Figure 3.5). The findings from all four models confirmed the efficacy of integrating remote sensing data, including Sentinel-1, Sentinel-2, Planetscope, and SRTM-DEM, for predicting AGB in a reforested urban environment. Notably, the XGBoost model outperformed the other three ML algorithms, as demonstrated by its highest R^2 score of 0.82 and lowest RMSE value of 55.40 t ha-1 among all the models evaluated in this study. Overall, this study highlighted the superior performance of the XGBoost and RF models in predicting AGB in a reforested urban environment, with the XGBoost model showing the highest accuracy. In comparison, the SVM and K-NN models had lower R^2 scores of 0.74 and 0.66, and higher RMSE values of 57.78 t ha-1 and 58.31 t ha-1, respectively.

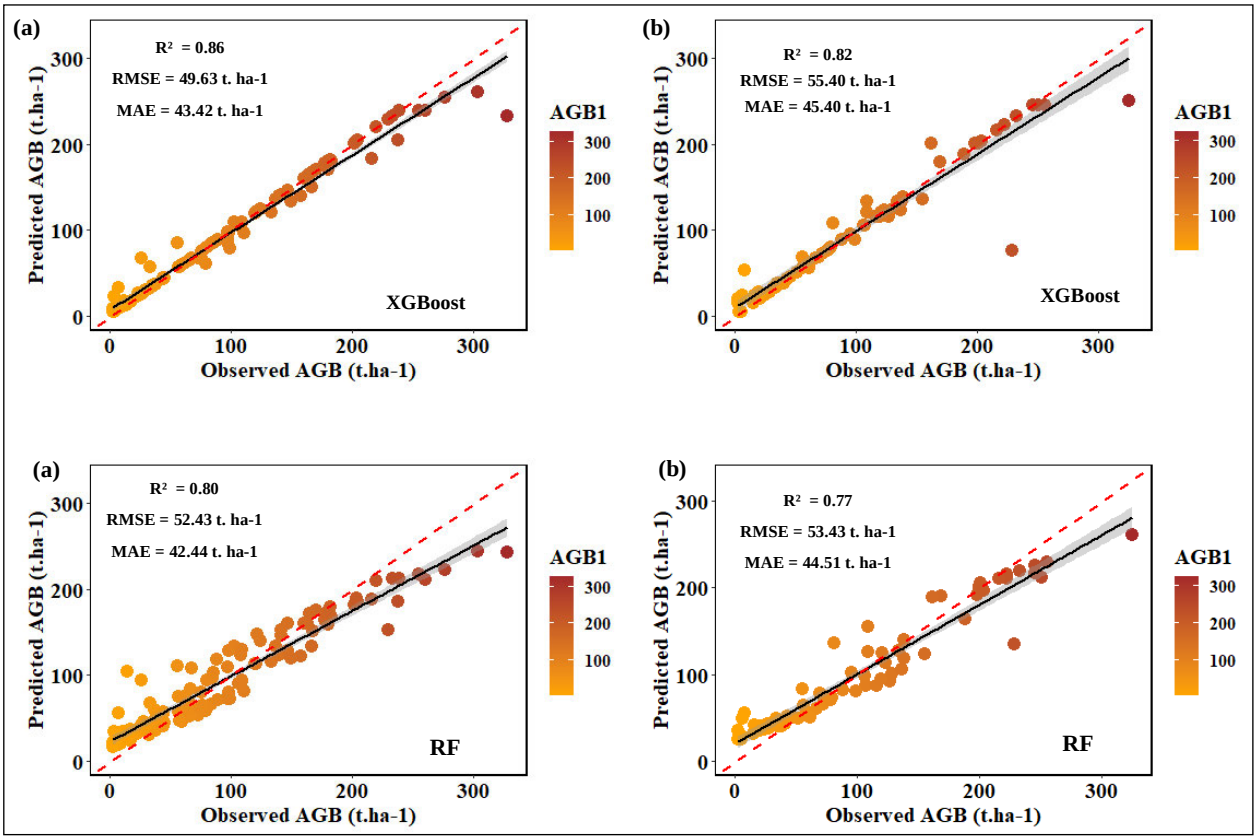


Figure 3.4. Relationship between predicted and observed AGB in a reforested urban environment for XGBoost and RF calibration (a) and validation (b).

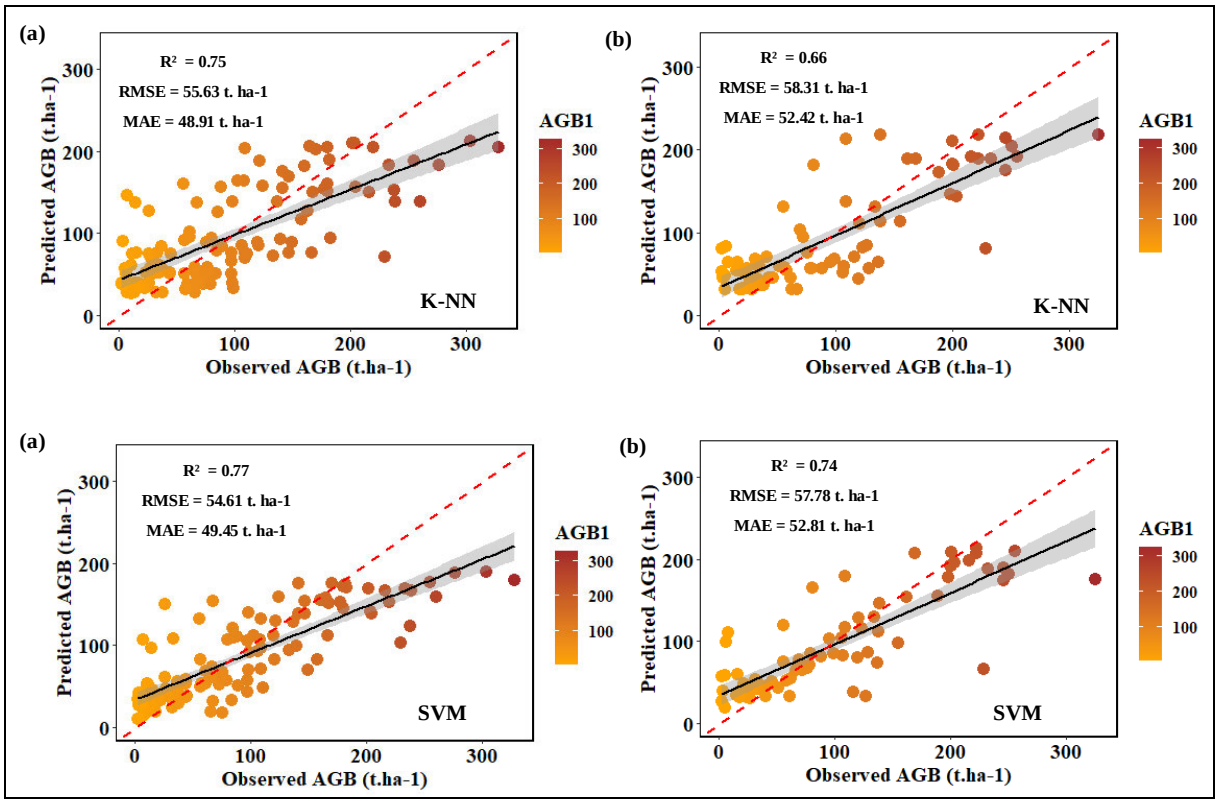


Figure 3.5 Relationship between predicted and observed AGB in a reforested urban environment for K-NN and SVM calibration (a) and validation (b).

3.3.5 Spatial Distribution of AGB

The study site was analyzed using four different algorithmic techniques, namely XGBoost, RF, SVM, and K-NN regression-based models, as well as SRTM-DEM, Sentinel-2, Planetscope, and Sentinel-1-SAR spectral bands in order to construct AGB distribution maps **(**Figure 3.6**).** The AGB values range from 4.1 t. ha-1 to 286.55 t. ha-1 and are in line with the observed field AGB values derived from allometric calculations, which range from 3.1 t. ha-1 to 327.68 t. ha-1. The XGBoost algorithm generated a map that showed high AGB values ranging from 100.44 t. ha-1 to 286.55 t. ha-1, with approximately half of the study site having low AGB values (<100.44 t. ha-1) and a mean AGB value of 96.84 t. ha-1. The southern part of the study site featured a mix of high and low-density forests as indicated on both the study area map and the modeled maps. The central sections of the map, where the landfill site belonging to the eThekwini municipality is located, consistently showed the lowest AGB values. Additionally, scattered small shrubs and trees were found in all directions, a result of existing land use practices that have been transformed into reforested sections within the study site, and the trees in this region are still young and growing. In the northern sections of

the map, the dark green color represents mature field plots with AGB values ranging from 150.54 t. ha-1 to 286.55 t. ha-1. The overall results showed a clear spatial distribution of AGB in the study site, with variations in AGB values based on the land use practices and forest density in different regions of the study site.

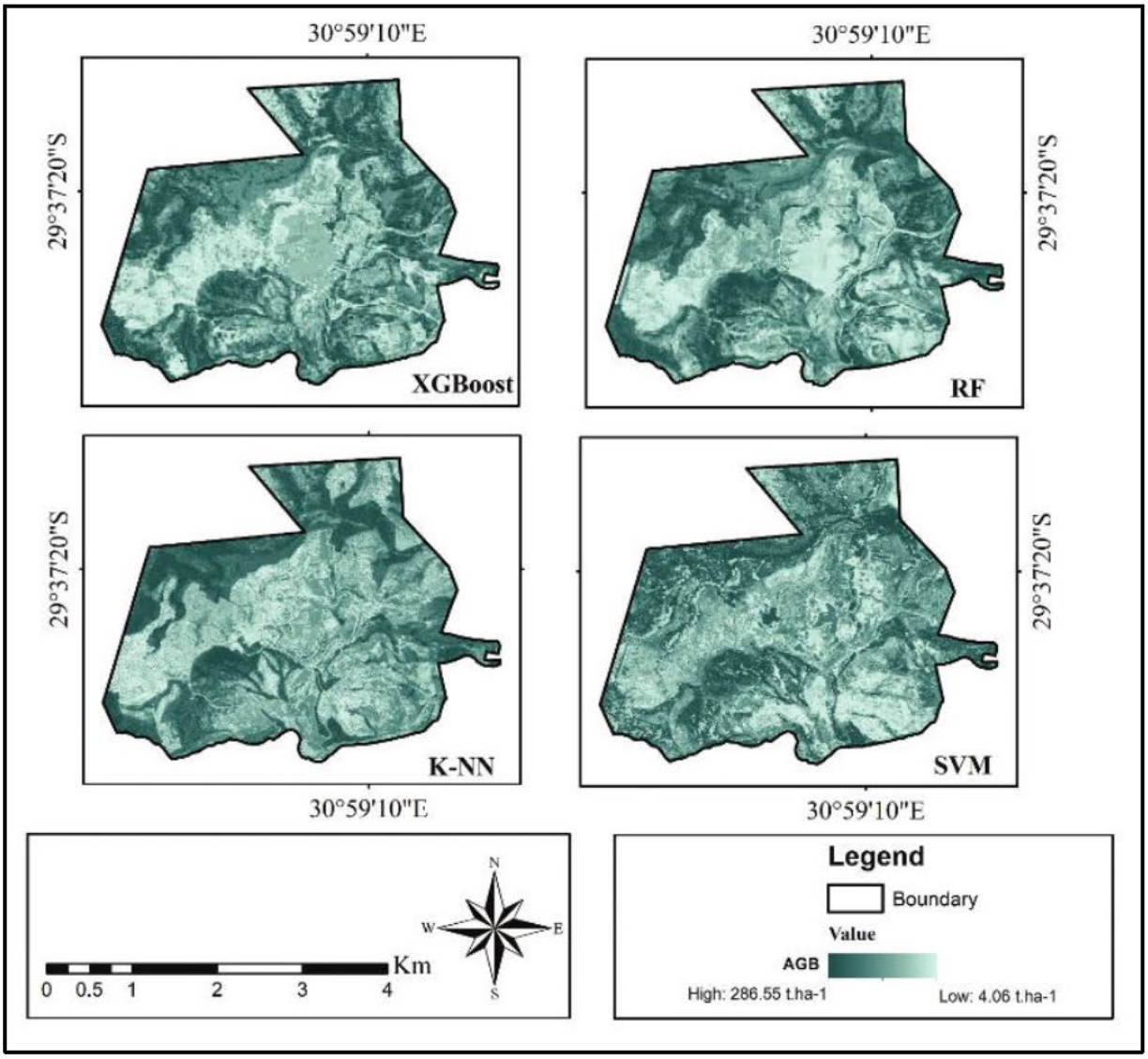


Figure 3.6. Maps comparing the predicted AGB in a reforested urban environment using four different machine learning algorithms.

3.4 Discussion

The storage of aboveground biomass (AGB) serves as a critical indicator for assessing the quality of urban ecosystems and their ability to absorb CO2 (Li et al., 2023). Estimating forest AGB in reforested urban areas using satellite remote sensing data presents several challenges, as multiple factors can impact the relationship between AGB and remote sensing variables. These factors include the choice of regression algorithm, sample size, topography, soil conditions, and forest structures (Lu et al., 2016, Li et al., 2021). Therefore, it is necessary to investigate machine learning algorithms on new remotely sensed data and their integration to enhance AGB estimation accuracy. In this study, we address these key issues by examining the effectiveness of ML algorithms in predicting AGB in reforested urban areas using a combination of remote sensing data from Sentinel-2A, Sentinel-1, Planetscope, and SRTM-DEM. R^2, RMSE and MAE were used to assess the quality of the forest AGB models.

In this study, several models (i.e., XGBoost, RF, SVM and K-NN) were tested to estimate forest aboveground biomass (AGB) in reforested urban landscapes, and their performance was evaluated, all of which exhibited satisfactory accuracy. The extreme gradient boosting (XGBoost) algorithm used in scenario (iv) achieved the highest accuracy among all models tested. This finding is consistent with previous studies that demonstrated the superior performance of XGBoost over other models such as CatBoost, GBRT, SVM, and RF in estimating AGB in different forest landscapes (Pham et al., 2020b). For example, Pham et al. (2020b)Pham et al. (2020b)Pham et al. (2020b)Pham et al. (2020b)Pham et al. (2020b) found that XGBoost outperformed these models in estimating mangrove biomass in the Red River Biosphere Reserve, Vietnam, while (Uniyal et al., 2022) achieved better results with XGBoost than RF, SVM, and K-NN in mapping the spatial distribution of AGB and carbon in urban forests in India. Luo et al. (2022)Luo et al. (2022)Luo et al. (2022)Luo et al. (2022)Luo et al. (2022) combined Landsat 8 OLI and Sentinel-2B data to estimate subalpine forest AGB using linear regression, and two ML approaches–random forest and extreme gradient boosting, with 54 inventory plots, XGBoost outperformed the two algorithms.

The XGBoost algorithm's strength lies in its ability to learn and extract features from aboveground biomass input data using a two-phase approach, as well as in its ability to approximate nonlinear relationships between forest AGB and covariates (Chen and Guestrin, 2016a). This enables XGBoost to capture any potential associations between them, making it a powerful algorithm for forest biomass/carbon estimation. Additionally, XGBoost can handle high-dimensional data, is fast and scalable, robust to outliers, and flexible (Chen and Guestrin, 2016a, Pham et al., 2020a). Overall XGBoost's exceptional performance can be

attributed to its ability to handle missing values, implement regularization techniques, use gradient-based optimization, provide feature importance insights, leverage parallel processing capabilities, and use ensemble learning techniques to combine the predictions of multiple decision trees into a highly accurate final prediction (Chen and Guestrin, 2016a, Uniyal et al., 2022, Pham et al., 2020a). RF also displayed good performance in the estimation of AGB in reforested urban landscapes and is consistent with existing literature predicting AGB in various forest landscapes (Mutanga et al., 2012, Mngadi et al., 2021a, Dube and Mutanga, 2015). Random Forest's exceptional performance can be explained by its capacity to mitigate overfitting, withstand noisy data, capture intricate nonlinear relationships, handle missing values, reveal feature importance, and harness parallel processing capabilities, rendering it a versatile and efficient choice for large-scale ML tasks, including regression (Breiman, 2001, Baines et al., 2020). Conversely, the K-NN model in scenario (iv) displayed the least performance, though its performance for estimating AGB in the study area was satisfactory. It is worth noting that RF and SVM algorithms are relatively good predictors of reforested AGB retrieval in this study area, where the amount of biomass is relatively low compared to mature forests. The results suggest that XGBoost and RF are powerful algorithms for estimating AGB in reforested urban landscapes, while K-NN should be applied with caution. However, it is important to note that each algorithm's performance varies according to the type of dataset and application, so it is always a good idea to carefully evaluate different algorithms beforehand. While all four models (XGBoost, Random Forest, KNN, and SVM) can handle complex non-linear relationships, XGBoost demonstrated superior performance, likely due to its ensemble learning approach that effectively captures intricate patterns, its ability to handle high-dimensional and sparse data common in remote sensing applications, and its regularization techniques that prevent overfitting. Random Forest's slightly lower performance could be attributed to its potential overfitting tendencies with correlated or redundant features. KNN and SVM models' performances may have been influenced by factors like the choice of distance metrics, kernel functions, and their sensitivity to high-dimensionality and noise (Ahmad et al., 2021). The relative performance can also depend on the specific features used, preprocessing steps, and hyperparameter tuning. Further experiments investigating the impact of different feature subsets, feature engineering techniques, and hyperparameter configurations could provide valuable insights into the observed performance differences and inform future applications in similar contexts (Ghorbanian et al., 2022).

A further evaluation of the effectiveness of integrating remote sensing data into biomass estimates was also carried out. Boruta feature selection was used to select predictor variables with the highest correlation with AGB, and variables with weak correlations and redundancy were excluded (Agjee et al., 2016). Based on the results from the variable of importance (Figure 3.3), estimation was primarily based on optical data from Sentinel-2A, including the red band, the near-infrared band, and the vegetation red edge band, with higher vegetation reflectance enhancing the estimation of reforested biomass as compared to radar remote sensing. This is primarily due to the rich information captured by optical multispectral sensors including Planetscope and Sentinel-2A. Moreover, the derived vegetation indices from the two sensors such as NDVI, MNDVI and GEMI can help enhance the prediction of biomass through their capability to differentiate vegetation types. Sentinel-2 and Planetscope data have higher spatial resolutions than radar Sentinel SAR, which means they can capture more detailed information about the vegetation structure and composition. This finding is consistent with Mutanga et al. (2012)Mutanga et al. (2012)Mutanga et al. (2012) who highlighted the role of red-edge bands in improving accuracy of estimating vegetation metrics such as biomass and carbon assimilation. This indicates that Sentinel-2 has good prediction performance for reforested urban ecosystems. Interestingly, literature has reported a similar result in other ecosystems (Li et al., 2020c, Mngadi et al., 2021a, Mutanga et al., 2012). In the remote sensing domain, it has been established that the NIR, SWIR, and the vegetation Red Edge spectral regions exhibit notably stronger correlations with forest biomass and carbon stock compared to the visible spectral range. These specific spectral bands have demonstrated a higher sensitivity to the vegetation properties and biochemical constituents, making them valuable indicators for assessing the biomass and carbon content of reforested urban forests using remote sensing techniques (Khan et al., 2020a, Mngadi et al., 2021a). In contrast, radar Sentinel SAR primarily captures data in the microwave portion of the electromagnetic spectrum, which may not provide the same level of discrimination between vegetation attributes as optical data. Optical data is better at detecting differences in vegetation greenness, while radar data is better at detecting stems and branches from high plants (Abudu et al., 2020, Ienco et al., 2019). This is because optical data captures light in the visible and near-infrared portions of the electromagnetic spectrum, while radar data captures light in the microwave portion of the spectrum. Performance of data for a particular study will depend on the specific characteristics of the study area. For instance, radar data may be more effective in areas with dense vegetation, while optical data may be more effective in areas with sparse vegetation (Pham et al., 2020a, Tamiminia et al., 2022, Gara et al., 2016, Smith et al., 1995).

Our findings suggest that the nature of our study site, which is a reforested urban forest with growing sparse remnant patches of the pioneer community may have contributed to the low performance of Sentinel-1 SAR data against Sentinel-2 MSI data and Planetscope data.

Vegetation indices provide radiometric measurements of the quantity, structure, and condition of vegetation (Gitelson et al., 1996). In this study, vegetation indices were developed from Sentinel-2A and Planetscope, which improved biomass estimates and the mapping of aboveground biomass (AGB) distribution in the reforested peri-urban landscape. The vegetation indices (i.e., NDVI, GNDVI, IRECI, MNDVI, EVI, ARVI, and GEMI) were computed from the spectral bands of Sentinel-2 (i.e., green, red-edge, NIR, and SWIR combinations) and Planetscope (i.e., red, green, and NIR), which were found to be a sensitive indicator of reforested AGB (Baloloy, 2018, Pandit et al., 2018, Luo et al., 2022, Zhang et al., 2022d). The red edge region was found to be particularly sensitive to vegetation parameters such as canopy biomass and chlorophyll content, which are intrinsic to forest carbon stocks, resulting in higher correlations with vegetation properties (Dube et al., 2016, Mngadi et al., 2021a, Gara et al., 2016, Mutanga et al., 2012). This finding emphasizes the contribution of the red-edge spectrum in estimating aboveground biomass in reforested forests, as well as the value of high-resolution bands in discriminating reforested trees. Therefore, it can be claimed that integrating Planetscope and Sentinel-2 MSI wavebands and vegetation indices can be used for accurate estimation of reforested aboveground biomass, based on the results of this study.

The results indicated that Sentinel-1 SAR data had a smaller contribution to the model compared to Sentinel-2A and Planetscope data, potentially due to its short wavelength (C-band). This limited Sentinel-1's performance in estimating aboveground biomass in reforested peri-urban sites because the C-band has a restricted ability to penetrate the forest canopy to gather essential forest structure information for enhanced aboveground biomass estimation, unlike longer wavelength SAR bands from other sensors (such as P and L) (García et al., 2018). Additionally, the absence of texture bands from the list of input features may be impacting Sentinel-1's performance. Recent research has shown that, due to the contribution of the Gray Level Co-occurrence Matrix (GLCM), Sentinel-1 outperforms Sentinel-2 and other sensors (Ronoud et al., 2021, Souza et al., 2019). Thus, integrating texture features and SAR data from longer wavelengths (P and L) with Sentinel-2, Sentinel-1, and Planetscope in future work should be considered to address the limitations of the SAR C-band and enhance accuracy. Nevertheless, the objective of the study was to map the distribution and density of aboveground biomass in a reforested urban landscape. The AGB distribution maps (Figure

3.6) are the predictions from the best performance (Model IV). The results demonstrate that the XGBoost model was robust, despite the wide range of site conditions represented in the field data. The aboveground biomass distribution maps show success in the identification of locations of high aboveground biomass and vice versa. Biomass values are estimated to be high in the densely populated vegetation areas of the site. As a result of remnant patches of pioneer forests before reforestation, the northwest and northeast sections of the study site have the highest concentrations of vegetation. Furthermore, the low-lying terrain in those areas favors growth. Young forests in the southeast and southwest sections of our study site have lower aboveground biomass densities because they have not yet reached their peak biomass capacity. As they mature, these forests will store more carbon and have a higher standing biomass. Furthermore, the central section of the study site is the bare landfill site that has been cleared for purposes of dumping waste. For future research, the model can be used to track changes in AGB density over time. This would help to understand the impact of climate change and other factors on carbon storage.

While providing valuable insights into estimating and mapping aboveground biomass (AGB) in reforested urban landscapes using machine learning and remote sensing data, this study has limitations that must be acknowledged. Inherent biases and uncertainties associated with remote sensing data, such as atmospheric effects, sensor calibration, and topographic influences, can propagate into model inputs and impact biomass estimate accuracy. Furthermore, the machine learning models' generalizability may be constrained by the study area's specific characteristics, like tree species composition, age distribution, and environmental conditions, necessitating cautious evaluation of their transferability. To address these limitations and advance research, future work should explore integrating complementary data sources (e.g., LiDAR, hyperspectral imagery) for improved accuracy and robustness, investigating transfer learning and domain adaptation techniques to enhance generalizability across varied urban reforestation contexts, conducting longitudinal monitoring and model updating to capture temporal biomass changes and growth dynamics, developing robust uncertainty quantification and error propagation techniques for reliable confidence intervals, and integrating ecological (e.g., disturbances, species interactions) and socioeconomic (e.g., urban development, management practices) factors to better understand the complex dynamics influencing biomass in reforested urban landscapes.

3.5 Conclusion

In this study, we demonstrate how integrating Sentinel-2A (optical), Sentinel-1A (radar), Planetscope MSI (high-resolution optical), SRTM-DEM (elevation) and image data transformations combined with ML algorithms, such as XGBoost, RF, K-NN and SVM, can be employed to estimate reforested aboveground biomass in urban environments. The main findings and conclusions of this study are as follows:

- The optimal XGBoost model achieved strong predictive performance (R2: 0.82, RMSE: 55.40 t. ha-1), demonstrating the promise of multi-source medium resolution optical, radar, and elevation data for wall-to-wall biomass mapping. Strategically positioned red, green, red-edge and near-infrared bands provided sensitivity to vegetation parameters.
- In data scarce regions, there is need for more applications of multi-source satellite sensors with strategically positioned bands including near infrared, red, red-edge, and green bands and a wide swath width (25 m) for wall-to-wall mapping of reforested biomass. Hence, these findings are valuable in sub-Saharan African countries, such as South Africa.
- Operational monitoring leveraging this methodology can support reforestation evaluation and evidence-based policymaking for urban forest carbon initiatives. The approach provides a valuable tool for data-scarce regions to quantify climate mitigation impacts.

Overall, harnessing integrated remote sensing and machine learning enables scalable, accurate modelling of reforestation aboveground carbon stocks to inform sustainable urban landscape planning and management.

3.6 Summary

This chapter presented a novel framework integrating multi-source satellite data and machine learning to map aboveground biomass (AGB) distribution across Buffelsdraai - an actively reforesting urban landscape. Specifically, an extreme gradient boosting (XGBoost) model outperformed other algorithms like random forest and support vector machines for AGB estimation using inputs from Sentinel-1, Sentinel-2, and PlanetScope. Variable importance analysis revealed key spectral bands and vegetation indices driving predictive performance enabling interpretability despite complexity. Findings demonstrated the superiority and operational utility of XGBoost for multi-sensor AGB modeling in heterogeneous urban forests.

Having established a robust method for mapping biomass using recent Earth observation data, the next phase involves applying this approach to quantify aboveground carbon stocks. Building on the framework developed here, the subsequent chapter leverages high-resolution PlanetScope spectral data and an optimized XGBoost model to map the distribution of carbon captured through extensive reforestation across Buffelsdraai.

4 CHAPTER FOUR: The Utility of Planetscope Spectral Data in Quantifying Above-Ground Carbon Stock in an Urban Reforested Landscape

This chapter is based on:

Matiza, C., Mutanga, O., Odindi, J., Mngadi, M., 2023. The Utility of Planetscope Spectral Data in Quantifying Above-Ground Carbon Stock in an Urban Reforested Landscape. *Ecological Informatics* 80 (2024): 102472.

Abstract

Urbanization, deforestation, and forest degradation significantly contribute to atmospheric carbon emissions and heightened climate change risks. Reforestation, a sustainable long-term land use strategy, offers mitigation by sequestering carbon dioxide. To assess reforestation efficacy within urban contexts, continuous carbon stock evaluation in reforested areas is essential for informed management and monitoring. Remote sensing techniques have gained traction in landscape analysis, driven by improved spatial-spectral data characteristics and novel indices. Notably, the Planetscope multispectral imagery, characterized by enhanced spatial and spectral attributes, has potential in enhancing carbon stock estimation. This study examines Planetscope's spectral, derived spectral features and terrain variables' effectiveness in estimating reforested urban landscape carbon stock. Employing extreme gradient boosting algorithm in Buffelsdraai, South Africa, the study's results are compared with an artificial neural network model to test the robustness of the model. Encouragingly, Planetscope spectral data accurately estimated reforested carbon stock with high R^2 (0.78 and 0.81) and low RMSE (27.33 and 29.75 t. ha-1) from calibration and validation datasets. Notably, the green normalized vegetation index (GNDVI), red-edge normalized difference vegetation index (NDVIRED), and red-edge simple ratio index (SRRED) are optimal predictors. These findings underscore the value of Planetscope spectral data and extreme gradient boosting for precise carbon stock predictions in reforested urban environments. This study's insights are pivotal for designing effective reforestation ecosystem management and monitoring strategies, with implications for larger-scale carbon sequestration projects and resilient urban landscapes.

Keywords: *ecosystem services; climate change; spectral indices, Planetscope, extreme gradient boosting*

4.1 Introduction

Urbanization is a major driver of environmental change, with significant implications such as deforestation, air pollution, and climate change (Odindi and Mhangara, 2012, Ren et al., 2011). This transformation affects the provision of valuable ecosystem services and processes that include mitigation of urban thermal heat and climate change events, reduction of noise and air pollution, provision of quality water, maintenance and enhancement of biodiversity within urban environments (Pechanec et al., 2018, Odindi and Mhangara, 2012, Mngadi et al., 2021a). Urban areas, despite occupying a small fraction of the global land surface, are responsible for a disproportionate share of atmospheric carbon emissions. This is due to their high energy and resource consumption, which are driven by a variety of anthropogenic activities (Yoro and Daramola, 2020, Mngadi et al., 2021a). Previous studies have shown that urbanization processes are often typified by extensive deforestation and forest degradation, resulting in substantial reductions in forest biomass and carbon sequestration potential, while accelerating greenhouse gas accumulation and climate change risks (Basyuni et al., 2023, Qi et al., 2022, Hanif, 2018, Ren et al., 2011, Shin et al., 2022, Fu et al., 2019). Thus, understanding terrestrial carbon sequestration and storage and its impact on urban landscapes can be helpful in addressing issues related to urban development, climate change mitigation and sustainable development. This necessitates establishing long-term alternative mechanisms for increasing carbon sinks and preventing potential climate change risks and impacts within urban landscapes.

Recently, the Reduction of Emissions from Deforestation and Degradation (REDD+) and the United Nations Framework Convention on Climate Change (UNFCC) have identified urban reforestation as a reliable and sustainable mechanism for long-term carbon sequestration and climate change mitigation (Angelsen et al., 2009, Kuyper et al., 2018, Teo et al., 2021). The emergence of reforestation initiatives in urban landscapes is anticipated to significantly influence local and regional carbon cycling, climate regulation, and environmental restoration (Mngadi et al., 2022b, Kuyper et al., 2018, Angelsen et al., 2009, Teo et al., 2021). Despite such expectations, information on carbon accumulation in reforested ecosystems and its contribution to the global and regional carbon dynamics remain largely scarce (Mngadi et al., 2022a, Wu et al., 2014). Hence, there is need for continuous mapping of aboveground carbon stock distribution in reforested urban landscapes to understand the value of reforestation initiative in the carbon fluxes and climate change regulation potential for informed management of urban landscapes. The absence of a sound methodology impedes the accurate quantification of aboveground carbon (AGC) stocks in reforested landscapes. Accurate

quantification of AGC is critical for protecting and increasing carbon stocks in these delicate ecosystems. Remarkably, this study pioneers a solution to this gap by utilizing high-resolution 8-band multispectral imagery from Planetscope to estimate AGC in urban reforested landscapes. Achieving this objective requires reliable, cost-effective and robust carbon mapping techniques and datasets. Although conventional approaches such as field survey and observation have proven highly accurate in mapping forest carbon stock, their shortcomings are widely reported in literature (Xu et al., 2016, Dube et al., 2018, Gara et al., 2016, Odebiri et al., 2022, Prakash et al., 2022). Meanwhile, remote sensing has emerged as a cost-effective and reliable source of primary data for concise wall-to-wall forest carbon modelling (Dube and Mutanga, 2015, Mutanga et al., 2012, Odebiri et al., 2020a).

Remote sensing offers cutting-edge spectral information at larger spatial coverage, necessary for local and regional estimation and monitoring of carbon accumulation in reforested urban landscapes (Mngadi et al., 2021a, Odebiri et al., 2020a, Mutanga et al., 2012). Recently, widely accessible medium-resolution satellite sensors like Landsat and Sentinel have been increasingly used for estimating aboveground forest carbon and biomass, especially in resource-limited areas like sub-Saharan Africa (Dube et al., 2016, Mngadi et al., 2021b, Mngadi et al., 2021a). These sensors provide numerous spectral-wavebands (located in the visible, near infrared and shortwave infrared regions) that are critical for precise detection of vegetation green-biomass (Mngadi et al., 2021a, Dube et al., 2016). The improvement in these sensor's spatial, spectral and radiometric characteristics has further prompted their adoption in mapping and monitoring forest aboveground biomass and carbon stock (Dube and Mutanga, 2015, Eckert, 2012, Odebiri et al., 2020a). However, saturation in dense forest canopies is a persistent challenge with these sensors, particularly in heterogeneous forest landscapes (Wang et al., 2022c, Dube and Mutanga, 2015). In this regard, accurate and concise carbon modelling and monitoring in a highly mixed forest ecosystem such as reforested landscapes require the utilization of highly detailed spatial resolution sensors (e.g., Worldview series, Quickbird and RapidEye). Finer spatial resolution of these sensors reduces the challenges associated with mixed pixel between canopy, shadow cast and soil-background, which could affect the performance of forest carbon mapping. Due to these tradeoffs, Worldview series, RapidEye and Quickbird have been widely exploited for mapping and monitoring forest carbon dynamics and biomass (Mugabowindekwe et al., 2022, Mngadi et al., 2022b).

The newly launched, high spatial resolution multispectral (MSI) Planetscope satellite sensor with cutting-edge sensing properties has not been explored for forest carbon stock estimation, especially in reforested urban landscapes. The sensor consists of eight strategically located

spectral wavebands within the visible (431 – 680 nm), near infrared (845 – 885 nm) and red-edge (697 – 713 nm) regions of the electromagnetic spectrum (Frazier and Hemingway, 2021). Despite its small spatial coverage (25 km), it acquires spectral information at a very high (3 m) spatial resolution (Frazier and Hemingway, 2021, Planet labs Inc, 2020). Furthermore, the sensor's high temporal resolution (1-day revisit) makes it ideal for frequent forest carbon modeling and mapping in reforested urban landscapes. The finer spatial detail could help delineate individual tree crowns and reduce mixing of signals from vegetation and non-vegetation components (Warwick-Champion et al., 2022). The additional red-edge band, known to be highly sensitive to vegetation characteristics, along with the visible and NIR bands, may enable better discrimination of vegetation types and retrieval of canopy biophysical properties linked to carbon stocks (Frazier and Hemingway, 2021). However, the potential of Planetscope's novel data characteristics for urban reforestation carbon mapping remains largely unexplored. The rich information in the sensor's indices and texture features, including the red edge band, has not been fully evaluated in an urban reforested landscape. These features are likely to be useful for optimizing forest carbon stock prediction due to their sensitivity to vegetation spectral response. Vegetation indices can minimize atmospheric effects, soil background, and senesced vegetation to provide pure spectral reflectance, which is critical for concise forest carbon stock estimation (Munyati, 2022, Taddeo et al., 2019, Poley et al., 2020). Furthermore, image texture has been used to aid in mapping forest biomass in dense tropical forests (Cutler et al., 2012), and in some regions texture is a better predictor of biomass than spectral vegetation indices (Hlatshwayo et al., 2019, Pandit et al., 2020). Because texture has been shown to be an effective method of mapping biomass in dense canopies, and can be calculated on widely available optical imagery, texture may be a useful technique for improving biomass and carbon maps at local and regional scales. Therefore, there is a need to evaluate Planescope’s derived spectral indices and texture features for precise estimation of reforestation carbon stock in urban landscapes.

Generally, remote sensing techniques provide spectral information that is highly correlated between the predictor variables (Baines et al., 2020, Zhang et al., 2019b, Zhang et al., 2022d, Odebiri et al., 2022). This could impede the performance of statistical techniques in analyzing remotely sensed data. Consequently, achieving optimal carbon mapping in reforested urban landscapes through remote sensing necessitates the application of robust and sophisticated statistical algorithms capable of mitigating multicollinearity and data overfitting. In this regard, non-parametric machine learning algorithms such as extreme gradient boosting (XGBoost) and artificial neural networks (ANN) have emerged as successful techniques in

estimating vegetation biomass and carbon stock across diverse landscapes (Pham et al., 2020a, Chan et al., 2022, Uniyal et al., 2022). The success of these models arises from their ability to select relevant covariates associated with spatially and temporally varying vegetation characteristics captured in remote sensing data, while avoiding overfitting (Pham et al., 2020a, Zhang et al., 2022d). Moreover, the XGBoost algorithm incorporates bootstrapping and bagging procedures, featuring iterative processes that facilitate effective data mining, crucial for optimizing vegetation modeling and mapping (Chan et al., 2022). The XGBoost constructs a series of decision trees using a stepwise model fitting technique to minimize bias and a model averaging approach to reduce variance (Tamiminia et al., 2022, Pham et al., 2020a).). It progressively increases the weight of weak learners (predictors) within each decision tree until model error is minimized and the response variable is accurately predicted. Conversely, artificial neural networks (ANN) typically consist of input, hidden, and output layers. ANNs require a substantial amount of training data to achieve effective learning. Among ANN architectures, the multi-layer perceptron (MLP) neural network stands out as the most used.

XGBoost is a powerful algorithm for estimating carbon stocks, but it is not immune to overfitting. By comparing its performance to ANN, we can gain deeper insights into its resilience against overfitting and other potential forms of model bias. This study provides a readily replicable methodology that uses field inventory and remote sensing multispectral data and machine learning. According to literature, a limited number of studies have tested the utility of the Planetscope eight band high resolution image under reforested urban landscapes. Hence, we examined the efficacy of Planetscope data coupled with vegetation indices, texture features and terrain variables in retrieving carbon stock estimates over a reforested urban landscape in South Africa using ML models i.e., XGBoost and ANN.

4.2 Materials and methods

4.2.1 Field data collection

In this study, field survey and data collection were carried out on the 4th of February 2020 during the summer season, favorable for maximum forest biomass. The study generated approximately 194 random sample points using a stratified random sampling method to ensure representative coverage of the study area. The sampling frame consisted of the entire reforested area of interest, which was first divided into non-overlapping strata based on topographic features (slope, aspect) and tree species composition derived from existing

vegetation maps. Within each stratum, a proportion of sample points was allocated based on the relative size of the stratum compared to the total study area. The sample points were then randomly generated within each stratum using the QGIS software's Random Points Inside Polygons tool, which utilizes a random number generator seeded with the computer clock.

The randomly generated point coordinates were inserted into a global positioning system and used to navigate to the sampling sites. At each random point, a plot-size window of 3 x 3 m was established, and the structural dimensions (e.g., height and diameter at breast height) of reforested trees were determined and recorded. The tree height was measured using a clinometer, while diameter at breast height (DBH) was measured using a caliper instrument. The location of sampled trees was recorded using a Trimble Global Positioning System with 0.5 m accuracy.

Despite limited labor and transportation resources, which may have constrained the number of sampling sites that could be visited and the time spent at each site, the sampling sites selection took into consideration many environmental conditions, such as a wide range of biomass, various topographic features, and various tree species. With the stratified random sampling strategy, we could reduce similarity between adjacent sampling locations, minimize spatial autocorrelation issues in statistical analyses, and enhance sample representativeness within the region. However, it is important to note that the limited resources may have introduced potential biases in the data collection process, such as incomplete coverage of certain areas or insufficient time spent at each sampling site, which could affect the accuracy and precision of the results.

4.2.2 Allometric computation of aboveground biomass and carbon

The Intergovernmental Panel on Climate Change (IPCC) has recently validated allometric models as a non-destructive technique that can be adopted to compute biomass (Shukla et al., 2019). Literature illustrates that the allometric relationship between tree diameter and height significantly affect tree biomass (Chave et al., 2014, Latifah et al., 2021), hence their measurements can be used to calculate in-situ biomass. Therefore, this study integrated tree height and DBH into allometric model (Eq. 4.1) to compute reforested aboveground biomass.

$$AGB = 0.0673 \times \left(p^{D^2H}\right)^{0.976} \quad \text{Eq. 4.1}$$

Where p represents wood density constant (1.2 g/cm^3), D for diameter (cm) and H indicates height (m).

Commonly, studies have affirmed that aboveground vegetation biomass holds approximately 50% of carbon stock, consequently, a factor of 0.5 can generally be used to convert dry biomass into carbon stock (Fang et al., 2001, Gara et al., 2016). However, it is important to note that this conversion factor can vary depending on the vegetation type and species composition. For instance, some studies have reported lower carbon fractions for certain coniferous species compared to broadleaf species.

Thus, this study used a fraction factor of 0.5 to convert allometric derived aboveground biomass into carbon stock based on the equation (Eq. 4.2) established by (Brown, 1997). However, we acknowledge that there may be some variation in carbon fraction across different tree species present, which could introduce a degree of uncertainty in the carbon stock estimates.

$$CDw = AGB \times 0.5 \qquad \text{Eq. 4.2}$$

Where CDw indicate carbon dry weight and AGB represents aboveground biomass (green weight biomass). The computed biomass and carbon stock using allometric equations were standardized to the same unit of measurement such as tons per hectare (t. ha-1). Future studies could consider quantifying species-specific carbon fractions or applying species-level adjustments to the conversion factor to further improve the accuracy of carbon stock estimation in heterogeneous forest stands.

4.2.3 Image acquisition and preprocessing

PlanetScope MSI level-3B captured on the 4th of February 2020 with 5% cloud cover was downloaded on the 15th of November 2022 on Planetscope Lab Explorer portal. PlanetScope image products are supplied already radiometrically and atmospherically corrected. Hence no image pre-processing was necessary. To ensure the suitability of the data for our specific study requirements, we took the following steps: 1) We overlaid the locations of our field sampling plots on the PlanetScope imagery to verify accurate geometric alignment, with the 3m spatial resolution allowing for precise co-registration with our plot coordinates collected using GPS; 2) We visually inspected the imagery for any obvious atmospheric contamination, cloud cover issues, or other anomalies that could impact data quality, and the 5% cloud cover reported was deemed acceptable; 3) We confirmed that the specific visible (coastal blue, blue,

green, yellow, red), near-infrared, and red-edge bands required for computing our target vegetation indices were present and of sufficient quality. Based on these checks, we determined that no further preprocessing was necessary beyond the radiometric and atmospheric corrections already applied by the data provider, as the sensor's 3m resolution and spectral band configuration aligned well with our objectives of extracting vegetation indices for biomass estimation from field plot data. However, we acknowledge that preprocessing needs can vary based on study-specific factors such as atmospheric conditions, terrain characteristics, and intended applications. Our decision was based on the suitability of the PlanetScope product for our particular study requirements, but future studies, especially those involving different sensor types, environmental conditions, or analytical approaches, should carefully evaluate the need for additional preprocessing steps to ensure data quality and reliability.

The sensor covers the visible (coast blue-b1, blue-b2, green-b3, green-b4, yellow-b5 and red-b6), near-infrared (b8) and red-edge (b7) regions of the electromagnetic spectrum, while capturing spectral information at a 3 m spatial resolution. This study extracted vegetation indices spectral data from a combination of Planetscope's bands. Such indices are critical for detecting vegetation green biomass, which include Enhanced Vegetation Index (EVI), Enhanced Vegetation Index-2 (EVI-2), Normalized Difference Vegetation Index (NDVI), Modified Normalized Difference Vegetation Index (MNDVI), Green Normalized Difference Vegetation Index (GNDVI), Soil Adjusted Vegetation Index (SAVI), Visible atmospherically Resistant Index (VARI), Chlorophyll Index Green (CIG) and Modified Soil Adjusted Vegetation Index (MSAVI). The red edge indices included Normalized Difference Vegetation $Index_{Red\text{-}edge}$ ($NDVI_{RE}$), Simple Red edge Ratio (SR_{RE}) and Red-Edge Triangulated Vegetation Index (RTVI) (Table 4.1). As a result, the derived vegetation indices were combined with spectral information extracted from individual bands.

Furthermore, Planetscope MSI texture metrics were derived using the popular gray level co-occurrence (GLCM) (Haralick et al., 1973) based on co-occurrence-based filter embedded in ENVI Classic 5.3 version. Eight texture variables (Table 4.1) including, mean, variance, homogeneity, contrast, dissimilarity, second moment, correlation and entropy were considered. Texture variables were calculated with two window sizes (i.e.,5 X 5 and 7 X 7) at 64 grayscale level quantization and co-occurrence shift 1X1. Selection of the appropriate window size and shift are a critical step in deriving texture features from remotely sensed

data. Thus, window sizes were chosen based on previous studies that applied texture to similar study sites (Pandit et al., 2020, Hlatshwayo et al., 2019) and recommended the two window sizes as optimum. For each window, band 2, 4, 5, 6,7, and 8 were extracted creating a total of 48 texture variables.

Previous studies have identified topographic variables as critical drivers to AGC stock distribution (Salinas-Melgoza et al., 2018, Smith et al., 2017). According Amatulli et al. (2018)Amatulli et al. (2018). spatial topographic variables are categorized into three main groups local, non-local and combined topographic features. Local topographic variables examine the surface geometry at a point on the land surface such as slope, curvatures and elevation, while non-local attributes portray relative locations of selected points such as catchment area, ruggedness and flow accumulation (Odebiri et al., 2020b). Combined attributes are an integration of both the local and non-local topographic variables such as slope length factor, topographic wetness index, topographic position index and channel-based network. In this study 10 different topographic variables (Table 4.1) that cut across the three classes (i.e., local, no-local, and combined) were selected. These variables were derived from a 30 m SRTM DEM in SAGA-GIS and ArcGIS PRO software.

Table 4.1 Planetscope derived variables, vegetation indices, texture features and terrain variables.

Variables	Formulae	Source
NDVI	$\frac{(NIR - Red)}{(NIR + Red)}$	(Pettorelli, 2013)
MSAVI	$\frac{2NIR + 1\sqrt{(2NIR + 1) - 8(NIR - Red)}}{2}$	(Qi et al., 1994)
$NDVI_{RE}$	$\frac{(RedEdgeIV - RedEdgeII)}{RedEdgeIV + RedEdgeII}$	(Imran et al., 2020)
MNDVI	$\frac{NIR - Red}{NIR + RED - SWIR}$	(Gitelson et al., 1996)
CIG	$\left(\frac{NIR}{GREEN}\right) - 1$	(Wu et al., 2012)
RTVI	$\left(100 * (NIR - RedEdge) - 10 * (NIR - Green)\right)$	(Kross et al., 2015)
VARI	$\frac{Green - Red}{Green + Red - Blue}$	(Stow et al., 2005)

GNDVI	$\frac{(NIR - GREEN)}{(NIR + GREEN)}$	(Imran et al., 2020)
SR_{RE}	$\frac{NIR}{RedEdge}$	(Shafri and Hamdan, 2009)
SAVI	$\left(\frac{NIR - Red}{NIR + Red + L}\right) \times (1 + L)$	(Clevers and Gitelson, 2013)
EVI	$2.5 \times \frac{NIR - Red}{(NIR + Red - (7.5 \times Blue)) + 1}$	(Bannari et al., 1995)
EVI_2	$2.5 \times \frac{NIR - Red}{NIR + (2.4 \times Red) + 1}$	
Texture features 5X5 & 7X7 for Band 2-8 (GLCM)	Gray-level co-occurrence matrix (Mean, Variance, Homogeneity, Contrast, Dissimilarity, Entropy, Second Moment, Correlation)	(Haralick et al., 1973)
Terrain variables (SRTM-DEM)	Slope, Topographic Wetness Index, Total Catchment Area, Plan Curvature, LS-factor, Roughness, elevation	

4.3 Modelling approach.

In our statistical analysis, each observation (sample) included carbon stock measurements at the plot level and spectral values extracted from corresponding plot locations in Planetscope images. We obtained this data from corresponding sample points using a combination of SAGA-GIS, ArcGIS, ENVI-Classic 5.3, and R-Studio. Despite noticeable plot clustering, we conducted a Moran's I spatial autocorrelation test on the study site samples to ensure data independence.

The dependent variable is the observed carbon stock, while independent variables comprise combined data from 8 single-band spectral information, 12 vegetation indices, 10 terrain variables, carbon stock data, and 48 textural features extracted from GLCM analysis of Planetscope remote sensing imagery. To identify the most relevant features and reduce dimensionality, we utilized the Boruta feature selection method, and 23 variables were selected. We further constructed carbon stock inversion models using the optimal features (23variables) identified, following Dang et al. (2019) approach . Specific variable information is provided in (table 1).

4.3.1 Extreme gradient boosting (XGBoost)

Extreme Gradient boosting (XGBoost) is known for its speed and accurate prediction process, especially when dealing with large and complex datasets. Its power lies in its ability to minimize bias error within and between spectral predictors. XGBoost builds sequential models to reduce errors of predictors that could potentially lower the estimation accuracy (Tamiminia et al., 2022, Zhang et al., 2019c, Huang et al., 2022b). In each sequent model, the weights of all weak learners (predictors) are increased (or boosted) and the weights for strong learners in the model are decreased (Huang et al., 2022b, Pham et al., 2021). This process is repeated until the error is reduced, and the data is correctly predicted. Furthermore, XGB permits model tuning through various hyper-parameters, namely: *ntree*, tree node size and gradient boosting iteration. *Ntree* builds a multitude of decision trees and select the most optimal tree subset for a better regression model (Jafarzadeh et al., 2021). Generally, the default value of *ntree* is 500 and can be adjusted to optimize the model (Jafarzadeh et al., 2021). Whereas node-size determines the smallest number of observations in a tree subset terminal node, and default value is always at 1 (Jafarzadeh et al., 2021, Pham et al., 2018). Gradient boosting iteration is a repetition process of increasing or decreasing weights in a training datasets (Huang et al., 2022b). All hyperparameters were tuned with a grid search method.

4.3.2 Artificial neural networks

Artificial Neural Network (ANN) algorithms simulate human learning by establishing and strengthening connections between input and output data (Campesato, 2020). These connections enable data linkage without the need for training data (Shen et al., 2021). Notable ANN algorithms include Radial Basis Function, Elman Recurrent, and Hopfield Neural Networks (Domingues et al., 2020, Campesato, 2020). However, Multilayer Perceptron Neural Networks (MLPNNs) using backpropagation have gained popularity and were employed in this study (Wang and Xing, 2008, Günlü and Ercanlı, 2020). MLPNNs consist of input, hidden, and output layers, each with interconnected nodes (neurons) that transform input into output data (Wang et al., 2017). In this study, the input layer had 23 neurons, correlating with predictors. The hidden layer's neuron count was determined using training and validation data, with the network featuring two hidden neurons and the lowest error being chosen. The output layer contained a single neuron representing carbon stock outputs.

During the MLPNN training phase, initial arbitrary connection weights were assigned. Inputs were forward-fed from the input to the hidden layer. Hidden neurons multiplied inputs by weights, summed products, and processed sums through a transfer function. Results propagated to the output layer, with output values compared to expected values for error computation. Iterative error back-propagation adjusted connection weights until reaching a target minimal error. The network then accurately estimated carbon stocks for both training and new input data without training data. This required tests to determine optimal learning rate (0.01), momentum (0.18), and training iterations (500). The trained network was subsequently used for feed-forward predictions on continuous spatial data.

4.3.3 Optimal predictor variable selection

Commonly, multicollinearity and image spectral noise impedes statistical analysis of remotely sensed data, leading to poor carbon estimation (Hall, 1999, Odebiri et al., 2020a). In this regard, rigorous selection of optimal subset of variables with smallest error rate is necessary for improving regression model performance and reducing spectral noise. The Boruta feature optimization method is a wrapper algorithm designed around the random forest algorithm (Kursa and Rudnicki, 2010). It considers the fluctuations in the average precision loss of trees in the forest and uses the average drop accuracy (i.e., Z_score) to measure the importance of each input feature. Using a mixed shadow feature set prior to selection, the correlation between the feature and the predicted value is eliminated, which is more advantageous when processing remotely sensed data that exhibits strong feature correlation. The *z-score* is defined as:

$$z \text{ score} = (I\ m)/s \qquad \text{Eq. 4.3}$$

where l is the importance score of the original feature, m is the mean importance score of the corresponding shadow features, and s is the standard deviation of the importance scores of the original feature. If the *z-score* is greater than a certain threshold, the feature is considered important. An overview of the execution steps and advantages of the Boruta algorithm can be found in Kursa and Rudnicki (2010) and (Ahmadpour Kasgari, 2018). There are ready-made packages in R and Python.

4.4 Model performance evaluation and validation

The effectiveness of the extreme gradient boosting model in estimating reforestation carbon stock in an urban landscape was tested using a 70-30 holdout validation approach. At first, the total input dataset (N = 194) was separated into 70% (n =118) calibration and 30% (n =76) validation datasets. The study used coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) to evaluate the performances of each machine learning algorithm in estimating carbon stock within a reforested urban landscape. Moreover, the importance of each predictor to the overall performance of the two models was assessed to aid in the interpretability of the models. The primary idea behind the variable of importance is to show the degree to which each variable contributed to the model's performance. Specifically, machine learning (ML) models often fail to achieve interpretability because they cannot quantify the importance of variables for regression or classification tasks, which is why they are called a black box (Campesato, 2020, Kruk, 2023). Recently, several techniques have been proposed to assist users in interpreting ML predictions. One of the methods is the shapely additive explanations (SHAP) used in this study. In SHAP, each variable is assigned an average value to demonstrate the magnitude of its impact on the output. For each type of machine learning model, SHAP has special functions, such as "modelExplainer" for ML and ANN, and "TreeExplainer" for XGBoost. As compared to other strategies, SHAP offers global and local interpretability (for more details on SHAP, see De Lucia et al. (2022), and (Chen et al., 2023).

4.5 Results

4.5.1 Aboveground carbon stock of reforested trees

The measured carbon stock's descriptive statistics revealed a range from 1.05 t. ha-1 to 163.84 t. t. ha-1, with a mean of 33.42 t. ha-1 and standard deviation of 28.73 t. ha-1. Initial examination exposed spatial clustering, followed by evidence of spatial autocorrelation. To address this, inverse distance weighting and row standardization were applied with a 10m weight. Subsequent retesting of the dataset, guided by the results of the Moran's I test shown in Figure 4.1, confirmed the suitability of the observed carbon data for further statistical analysis.

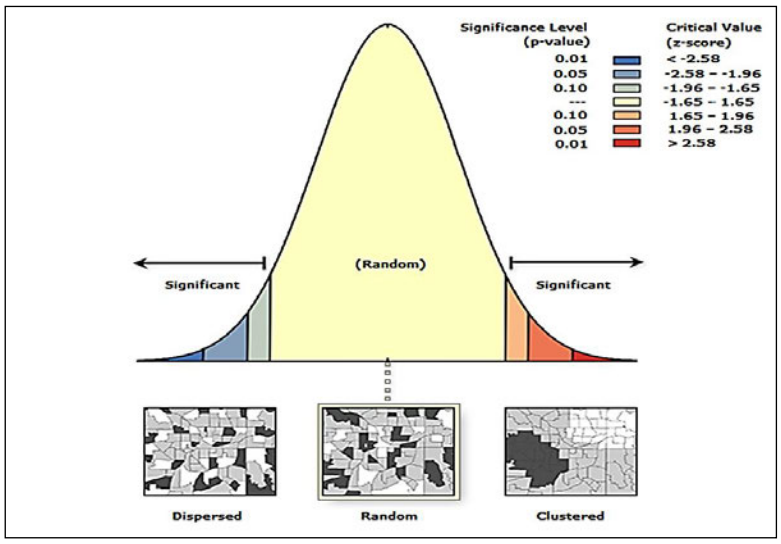


Figure 4.1 Moran's I test for spatial autocorrelation (Anselin, 2020).

The initial detection of spatial clustering and autocorrelation in the carbon stock data highlights the importance of considering spatial dependencies and accounting for them in analyses. The application of inverse distance weighting and row standardization effectively mitigated the spatial autocorrelation issue, ensuring that the carbon stock values at each location were appropriately adjusted based on the values of neighboring locations, meeting the independence assumption for reliable statistical analyses and modeling. The observed spatial patterns and the need to account for spatial autocorrelation also have implications for understanding the underlying processes and drivers influencing carbon stock distribution, as environmental factors, land-use histories, and ecological processes often exhibit spatial dependencies. By explicitly addressing spatial autocorrelation, subsequent analyses and models can more accurately capture the relationships between predictor variables and carbon stocks, leading to improved understanding, mapping, and prediction of carbon stock dynamics across the study area. The 10-meter distance applied for inverse distance weighting aligns with the spatial resolution of the available data and captures the local variations in reforestation and carbon stock. Secondly, it represents an ecologically meaningful scale at which environmental factors influence vegetation growth and carbon accumulation.

4.5.2 Feature optimization

Developing predictive models often requires the computation of the smallest possible number of the most influential variables. Through Boruta feature selection model, a significant reduction in dimensionality of the covariates was achieved. The results in (Figure 4.2) shows

ideal features selected for optimal prediction of carbon stock in a reforested landscape. Based on the Boruta feature optimization method, a subset of 23 predictor variables were selected for the final prediction model of carbon stock. Specifically, band 6 (red), band4 (green 2), band 8 (near-infrared), RENDVI, RE_SR, GNDVI, NDVI, CIG, SAVI, RTVI, band7 red edge, EVI1, EVI 2, band 4 (5X5 mean), band5 (5X5 mean), band5 (7x7 mean), band6 (7X7 mean), band7 (7x7 mean), band8 (7x7 mean), band8 (7X7 mean), elevation, topographic position index and Channel Network base. Thereafter, the selected variables identified by Boruta were then used to build separate models and their results were compared to an XGBoost and ANN model that included all 78 variables. The selection of these specific variables suggests that a combination of spectral information (bands and vegetation indices), spatial context (texture metrics), and environmental factors (topography) were important for accurately mapping carbon stocks across the study area.

Compared to using the full set of 78 original variables, the reduced 23-variable model likely benefited from reduced noise, multicollinearity, and overfitting issues, while still capturing the key predictive information.

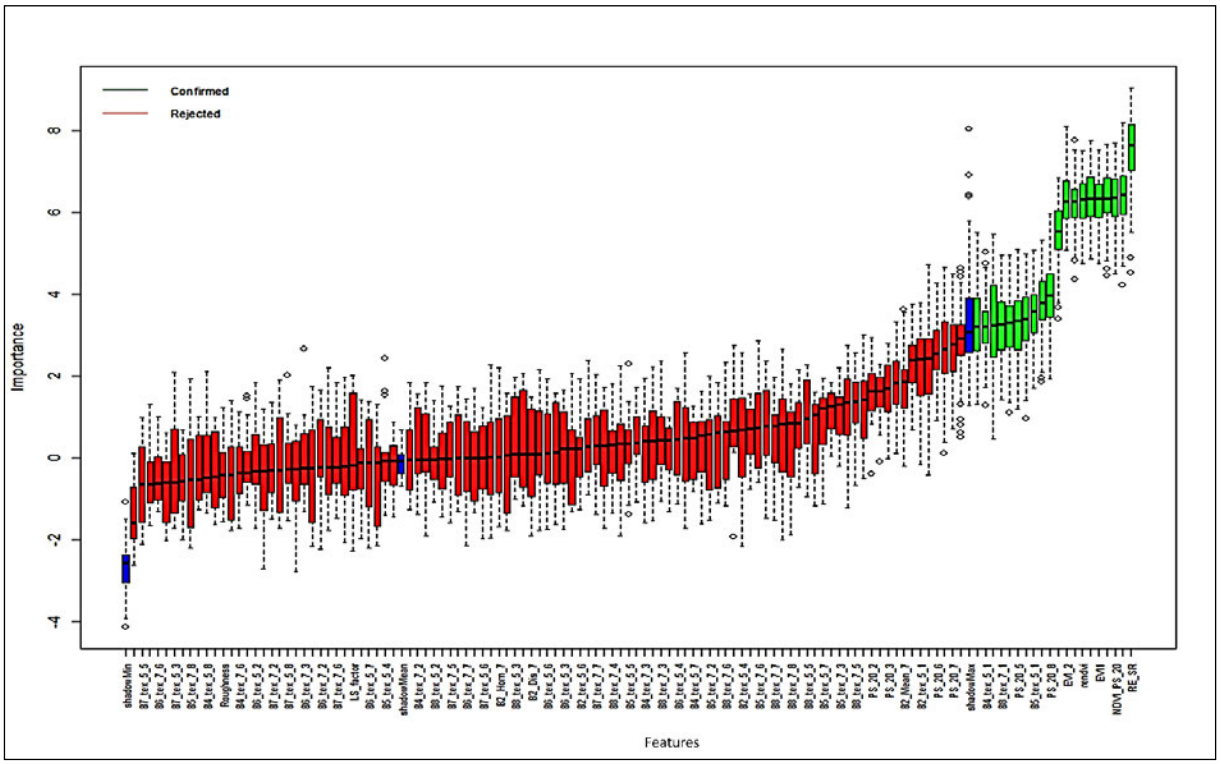


Figure 4.2 Boruta feature selection (n=78) the optimal variables are illustrated in green(confirmed) and the rejected in red. Note that half of the confirmed and rejected features aren't shown on the x-axis due to space.

4.5.3 Prediction performance of extreme gradient boosting and artificial neural network models

Results in (Table 4.2**)** show overall mean carbon stock and its predictive performance using Planetscope's spectral information extreme gradient boosting and artificial neural network model. Among the two models XGBoost showed the strongest robustness in predicting carbon stock. The consolidation of important predictor variables in XGBoost model produced an overall mean carbon stock of 30.42 t. ha^{-1} based on calibration data and 31.08 t. ha^{-1} using validation dataset. The XGBoost regression model obtained the highest coefficient of determination (R^2: 0.78 to 0.81) with lowest RMSE (29.75 to 27.33t. ha-1) and MAE (21.84 to 24.66 t. ha-1) when using integrated subset of selected spectral variables, compared to the utility of the entire set predictor variables in the model. The ANN model obtained a lower coefficient of determination (R^2: 0.77 to 0.73) with an RMSE (28.17 to 30.44t. ha-1) and MAE (22.39 to 24.43 t. ha-1) when using integrated subset of selected variables.

While both the XGBoost and ANN models demonstrated reasonably good performance in predicting carbon stock using Planetscope spectral information, it is important to assess the statistical significance of the differences in their performance metrics. To achieve this, we conducted a paired t-test to compare the R^2 values of the two models using both the calibration and validation datasets. The results showed that the XGBoost model significantly outperformed the ANN model in terms of R^2 values in both the calibration ($p = 0.028$) and validation ($p = 0.011$) datasets.

Furthermore, an analysis of variance (ANOVA) was performed to compare the RMSE and MAE values of the two models. The ANOVA revealed a statistically significant difference in RMSE values between XGBoost and ANN ($p = 0.037$ for calibration, $p = 0.049$ for validation), with XGBoost exhibiting lower RMSE values in both datasets. However, no significant difference was found in the MAE values between the two models ($p = 0.212$ for calibration, $p = 0.488$ for validation).

These statistical tests indicate that while both models performed reasonably well, the XGBoost model demonstrated a significantly higher explanatory power (as indicated by the

higher R^2 values) and lower prediction errors (as indicated by the lower RMSE values) compared to the ANN model. Therefore, the XGBoost model can be considered superior for predicting carbon stock in urban landscapes using Planetscope spectral data, although the differences in MAE values were not statistically significant.

Table 4.2 Prediction performance of reforestation carbon stock using XGBoost and ANN model and Planetscope's spectral information separated into calibration and validation datasets.

Algorithm model	**Prediction dataset**	**Mean C (t. ha-1)**	**R^2**	**RMSE (t. ha-1)**	**MAE (t. ha-1)**
XGB	Calibration	30.42	0.81	27.33	21.84
	Validation	31.08	0.78	29.75	24.66
ANN	Calibration	31.33	0.77	28.17	22.39
	Validation	32.44	0.73	30.44	24.43

The relationship between predicted carbon stock and allometric derived carbon stock (measured) is represented in Figure 4.3**.** The results in Figure 4.3**.**present a strong correlation coefficient (R: 0.83 to 0.85) between predicted and measured carbon stocks for XGBoost and (R:0.77 to 0.80) for ANN. The SHAP technique was employed to ascertain the relative significance of individual variables within the machine learning model. Notably, for the XGBoost model, the five most prominent variables encompassed red-edge simple ratio (RE_SR), green normalized vegetation index (GNDVI), channel network base (terrain), red band (band 6), and near-infrared band (band 8). In the case of artificial neural networks, the pivotal variables included RE_SR, normalized difference vegetation index (NDVI), red-edge normalized difference vegetation index (RENDVI), soil-adjusted vegetation index (SAVI), and enhanced vegetation index 2 (EVI2) (Figure 4.4).

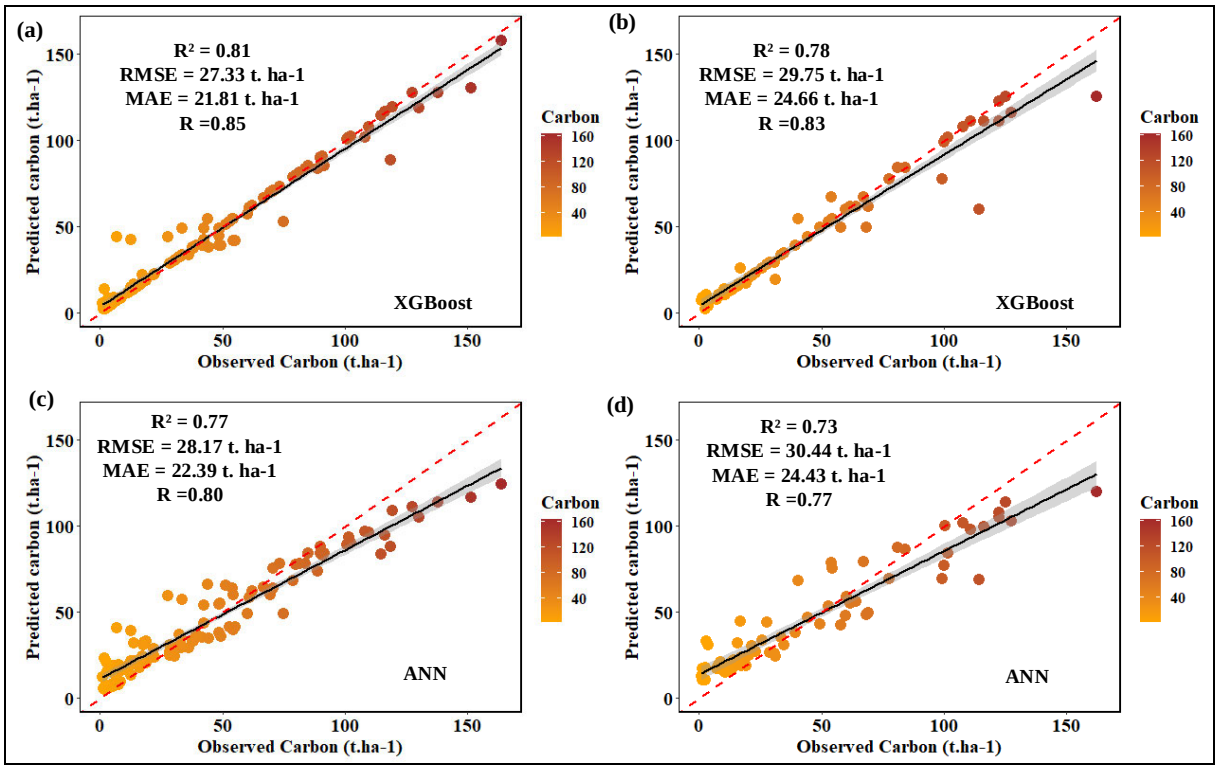


Figure 4.3 Relationship between predicted and observed aboveground carbon stock of a reforested urban landscape for (a) calibration and (b) validation datasets. The regression correlation between predicted and observed carbon stock was established using a subset of optimal input variables.

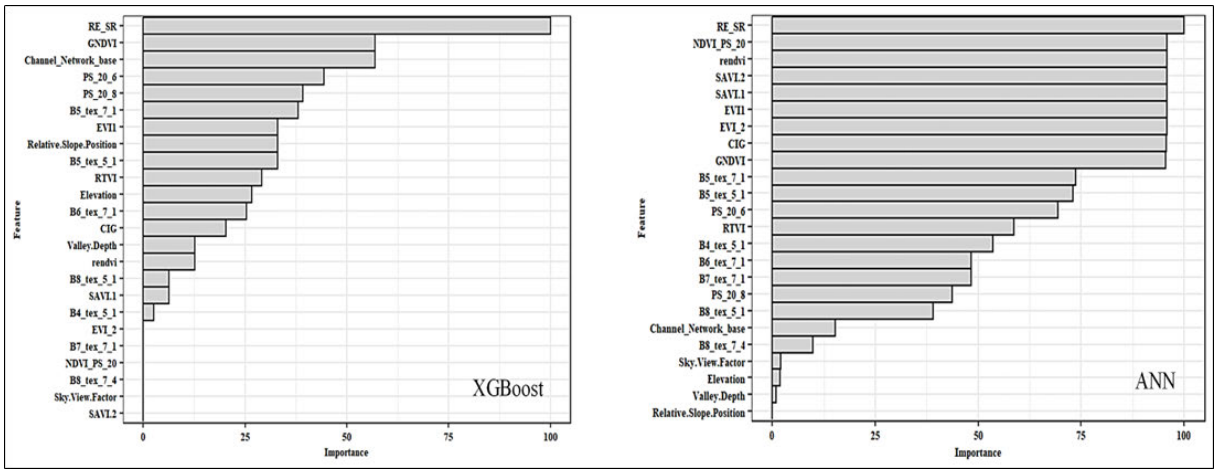


Figure 4.4 The importance ranking of individual variable for carbon stock estimation using XGBoost and ANN models.

In addition, Figure 4.5 demonstrates carbon stock spatial variability across reforested trees within an urban landscape. Commonly, carbon stock varies with an increase and/or decrease in green biomass and canopy cover.

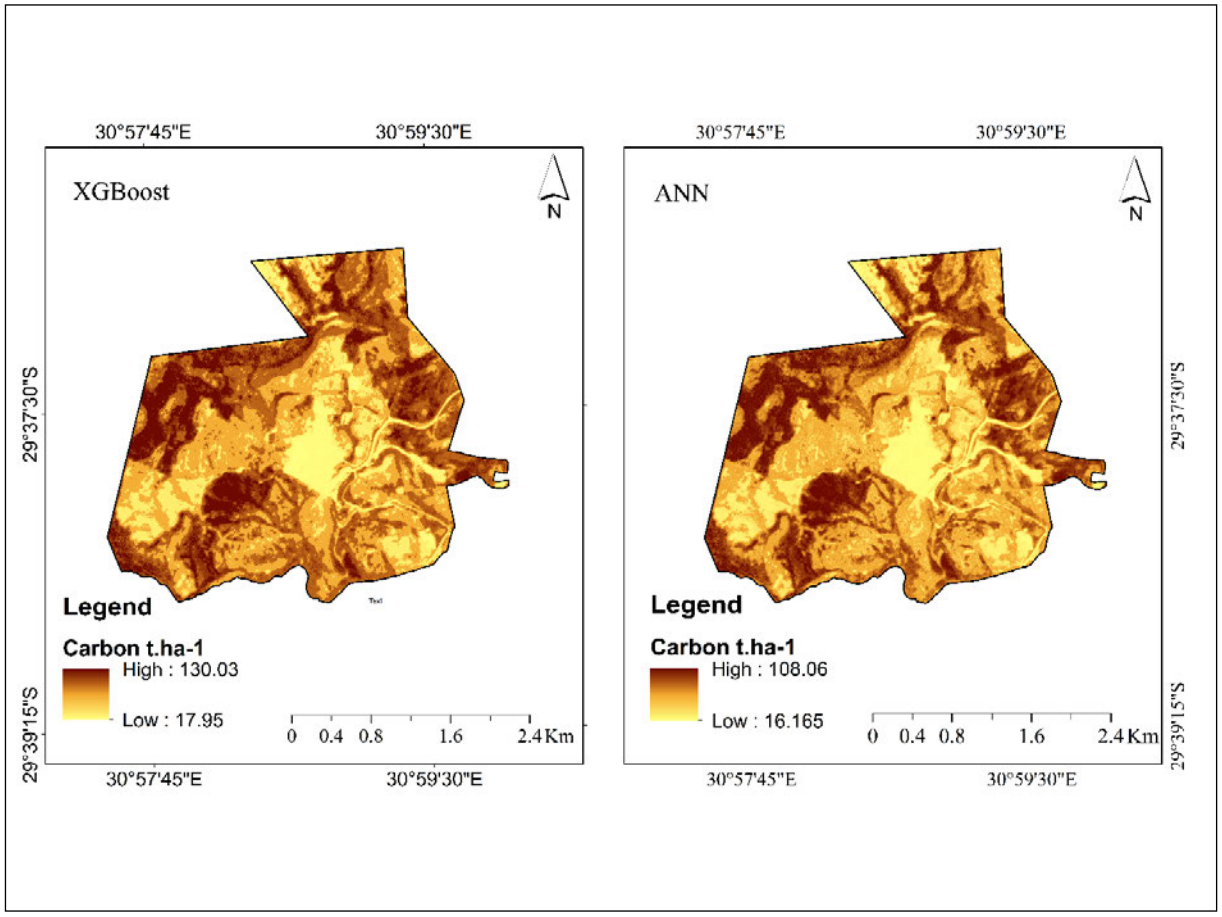


Figure 4.5. Aboveground carbon stock estimation map within a reforested landscape using extreme gradient boosting and artificial neural networks.

4.6 Discussion

4.6.1 Planetscope estimated Aboveground Carbon Stocks (AGC)

Accurate carbon assessments in urban reforestation are vital to comprehend reforestation's role in the global carbon budget. This study aimed to estimate aboveground carbon stock in reforested urban landscapes using high-resolution Planetscope MSI and machine learning. We established a predictive framework for reforested urban landscapes in South Africa, incorporating spectral data, vegetation indices, texture features, and terrain variables based on

XGBoost and ANN. Results indicated that the adoption of Planetscope spectral data, indices , texture features and terrain variables successfully predicted carbon stock in a reforested urban landscape with a high R^2 (0.78 and 0.81) and low RMSE (27.33 t. ha-1 and 29.75 t. ha-1) based on calibration and validation datasets of XGBoost, respectively. Notably, our estimates of carbon stock in a reforested area are consistent with previous studies that quantified the dynamics of forest carbon stocks using high resolution data in young and middle-aged reforested forests. Forest carbon stock ranged from 17.95 to 130.03 t. ha-1 in our study and 7.22 to 146.83 t. ha-1 in a reforested site in Thailand (Pibumrung et al., 2008). Similarly, Chinembiri et al. (2023), obtained 0.2-228 t. ha-1 for carbon stocks in a disturbed plantation in Zimbabwe using Sentinel-2. Liu and Nan (2018) obtained 16-276 t. ha-1 of carbon stock for a secondary coniferous forest along an altitudinal gradient which is slightly higher than what this study obtained, showing the strong resilience of secondary forests without further disturbance even under unfavorable conditions. Some researchers have reported that forest carbon stocks increased with stand age and maturity in other natural ecosystems and thus time dependent (Padalia et al., 2023, Coursolle et al., 2012). Biomass affects the amount of carbon stored in every part of the tree. Hence an increase in biomass will be followed by an increase in carbon content. (Li et al., 2016b). Accordingly, reforested carbon stocks from the study site were still less than that of old -growth forests both in quantity and proportion to the total carbon stock, such as those reported for old growth secondary forests indicating a potential for further accumulation (Tang et al., 2011).

Overall, the Planetscope multispectral 3m resolution image provided satisfactory results with the integration of other critical predictors such as mean texture, channel-based network and vegetation indices. This performance is consistent with Baloloy (2018) who noted that Planetscope produced a higher R^2 value than RapidEye, while RapidEye had a higher R^2 value than Sentinel-2-m in the estimation of aboveground biomass. This observation suggests that canopy and understory mixing was partially reduced by using higher resolution data, a characteristic that was also noted within our reforested urban landscape.(Andreatta et al., 2022) further validates the reliability of commonly employed Vegetation Indices (VIs) that rely on the distinction between Near-Infrared (NIR) and the red wavelength region. For instance, the Enhanced Vegetation Index (EVI) utilizing S2 imagery demonstrates a strong correlation ($R^2 = 0.91$). Additionally, our findings underscore the superior performance of VIs based on the red-edge spectral range for Planetscope (PS) imagery (with a correlation coefficient of $R^2 = 0.89$ for RECI). The accurate depiction of spatial patterns at the study site,

characterized by intricate grassland fractional vegetation cover (FVC) variations, was achieved exclusively through medium to high spatial resolution imagery sentinel 2 and Planetscope.

4.6.2 Model prediction performance and optimal variables.

The sensor's predictive success is credited to its strategically configured spectral-wavebands, such as red-edge and near infrared. Furthermore, variables such as red-edge simple ratio, enhance vegetation index 1 and2, soil adjusted vegetation index and normalized difference vegetation index were amongst the shapely ranked important variables. A strong correlation (R:0.85 and 0.83) was found between predicted and measured aboveground carbon stock using calibration and validation data. This correlation stems from optimal explanatory variables selected via Boruta feature selection. GNDVI is a critical indicator for estimating carbon stock, offering sensitivity to vegetation properties that impact photosynthesis and carbon stock conversion. The rich spectral data of GNDVI combines green and near-infrared bands for precise mapping of green biomass. Adoption of red-edge indices like NDVIRED also improved predictive performance for urban reforestation carbon stock, offering pure spectral reflectance with reduced influences.

The findings in this study are consistent with (Sibanda et al., 2017, Gara et al., 2016, Khan et al., 2020b) who discovered that red-edge derived spectral indices are highly robust in measuring or detecting important vegetation metrics such as carbon stock, biomass and leaf area index. For instance,Mutanga et al. (2012) found that red-edge indices can optimally increase the prediction performance of aboveground biomass in wetland ecosystems.Khan et al. (2020b) established that red-edge derived indices are critical for improving forest biomass prediction accuracy. All these studies recommended adoption of red-edge spectral indices as the effective explanatory variable for measuring and monitoring vegetation health and productivity (based on carbon uptake). Furthermore, red-edge derived spectral indices are less susceptible to saturation challenges common with traditional NDVI, hence can be effectively adopted in a dense forest canopy environment (David et al., 2022, Baloloy, 2018). Additionally, red-edge indices offer sensitive spectral reflectance as red-edge bands detect and record upright changes in plant leaf mesophyll and leaf structure, hence robust in monitoring spatio-temporal dynamics of vegetation productivity and health.

Wang et al. (2019a) found that climate is the main driver of biomass, but topography can have an indirect effect by influencing climate. Zhang et al. (2021) showed that elevation, a topographic factor, affects the spatial distribution of AGC by influencing light and water availability. Guerra-Hernández et al. (2022) used remotely sensed and terrain data to predict AGB in Mediterranean forests. They found that AGB estimates made with satellite data are affected by the edaphoclimatic and topographic characteristics of the study area. In this study, terrain variables were found to be important predictors of AGB. Stand development is influenced by environmental factors such as climate, site productivity, infiltration, and wind exposure. Topography and climate can also affect carbon stocks through their interactions with tree species development and diversity.

The accuracy of the predictive models based on Planetscope imagery surpasses similar models in comparable environments (Jiang et al., 2021, Ahmed et al., 2022). Notably, the XGBoost based carbon stock predictive model demonstrates the most precise predictions compared to ANN, as indicated by the lowest RMSE (27.33 t. ha-1 and 29.75 t. ha-1) based on calibration and validation. This highlights its superiority in estimating carbon stocks. However, ANN also demonstrated accuracy though its R^2 was lower (R^2: 0.77 to 0.73) than XGBoost and RMSE slightly higher RMSE (28.17 to 30.44t. ha-1) and MAE (22.39 to 24.43 t. ha-1). This is consistent with previous studies that have reported the superiority of XGBoost over ANN. For instance Ramdani and Furqon (2022) found that XGBoost had the lowest RMSE value (1.56) compared to ANN (4.33), Random Forest (6.81), and Support Vector Machine (7.45) in the classification of urban forests. Furthermore, Memon et al. (2019) noted that XGBoost outperforms the image classification task for CP RISAT-1 data than ANN in landcover classification. Finally, the neural network algorithm (overall accuracy of 73% and kappa index of 0.66) had the lowest accuracy in land cover mapping in the pilot area of the province (Eskandari and Sarab, 2022). However, despite the differences in accuracy, this study still recommends using ANNs in forest carbon stock estimation as its results were satifactory.

4.6.3 AGC maps

The context of carbon monitoring within terrestrial ecosystems has necessitated the development of high-resolution tree biomass maps, driven by the imperative to assess ecosystem responses to climate change (Huang et al., 2019). This has fueled the utilization of remote sensing (RS) technologies over the past few decades. As part of this effort, this study

presents location-based carbon stock maps derived from predictive models. These maps are pivotal in monitoring ecosystem health, identifying concerns in plantation forests, and guiding decisions during extreme weather events and disease outbreaks, enabling effective management strategies. Similar to Basyuni et al. (2023), this study's carbon stock maps from restored urban landscapes signify the feasibility of urban landscape recovery through rehabilitation, highlighting its role as a natural carbon removal mechanism. Nevertheless, it may take decades or even centuries to restore biomass carbon stocks to a level comparable to that of natural forests (Kanowski and Catterall, 2010) . The biomass and carbon stock of rehabilitated forests, especially those with higher species diversity, are similar to those of mature secondary forests (Liu and Nan, 2018).

The spatial distribution of carbon stock within the study area, as depicted in Figure 5, exhibits variability linked to canopy cover and green-biomass density. This variation can be attributed to topographic features significantly influencing green-biomass productivity and density. Xiong and Wang (2022) reveal that the local R-squared values concerning the Normalized Difference Vegetation Index (NDVI) and elevation, within smooth components, exhibit an augmentation with increasing decomposition scale. These findings underscore the substantial role of elevation at a broader scale, yet also suggest a muted influence at this level, indicative of the ongoing and gradual elevation shifts. Numerous studies underscore the impact of topographical attributes such as topographic position index, channel network base, and elevation on the spatial distribution of carbon stock within forest landscapes (Odebiri et al., 2020b, Chan et al., 2022, Salinas-Melgoza et al., 2018, Xiong and Wang, 2022). Furthermore, species composition, driven by biophysical (e.g., leaf area index, stomata) and biochemical (e.g., carotenoids, leaf pigments) properties, contributes to carbon stock variations. Deciduous trees like Acacia and Bridelia exhibit more significant stomatal properties, facilitating greater carbon uptake per unit of absorbed light compared to shrub trees (e.g., Artemisia), characterized by limited structural geometry and leaf biomass, leading to reduced carbon accumulation (Mngadi et al., 2022a, Miller, 2014).

4.6.4 Implications and applications

The study delving into the application of Planetscope spectral data for above-ground carbon stock estimation within urban reforested landscapes presents consequential implications and applications. The accurate quantification of carbon stocks through machine learning algorithms offers insights that extend across diverse domains. This includes providing

informed urban planning, enabling efficient restoration initiatives, quantifying the socio-economic value of reforestation projects, and facilitating ongoing carbon monitoring. Moreover, the study's findings resonate in policy circles, guiding land-use decisions by integrating precise carbon stock data. The research holds significance in promoting cross-disciplinary collaboration, as it aligns ecology, remote sensing, and data science, while also stimulating educational and public awareness campaigns. Beyond its immediate context, the study underscores the economic, environmental, and social benefits of reforestation, contributing to sustainable urban landscapes and climate change mitigation efforts. In essence, the study's comprehensive implications underscore its pivotal role in advancing urban sustainability, informed decision-making, and collaborative endeavors in pursuit of resilient and carbon-conscious urban ecosystems.

4.7 Conclusion

This study sought to explore the effectiveness of Planetscope MSI for estimating aboveground carbon stocks in reforested peri-urban landscape. Based on the results, it is deduced that:

- Planetscope derived spectral information can be effectively adopted to model aboveground carbon stock.
- The utility of optimal spectral indices such as GNDVI, $NDVI_{RED}$ and SR_{RED} can significantly enhance the predictive performance of aboveground carbon. The results presented in this study are valuable for understanding the contribution of reforestation initiative in the global carbon cycle and its potential in meeting the objectives of Kyoto Protocol and Reducing Emissions from Deforestation and Forest Degradation (REDD+), which are based on climate change mitigation, particularly in urban landscapes. Furthermore, the information provided in this study is critical for forest managers, decision and police-makers to adopt well-informed management and monitoring frameworks, including planning for larger-scale reforestation projects.

Acknowledgements

The authors would like to acknowledge the DST-NRF SARChI Chair in LandUse Planning and Management at UKZN (Grant No. 84157) for their financial support. A special acknowledgement to Durban municipality through the Durban WoodRights project for

allowing this research project to be carried out on their protected reforestation site. We further thank the reviewers for their constructive criticism which improved this research work.

4.8 Summary

This chapter demonstrates the efficacy of high-resolution PlanetScope data integrated with an optimized machine learning approach for quantifying and mapping aboveground carbon stocks across an actively reforesting urban landscape. The Extreme Gradient Boosting model framework leveraging image textures, vegetation indices and terrain variables provides localized insights into carbon densities to strengthen municipal planning and monitoring. However, the spatio-temporal variability in accumulation rates and trajectories resulting from extensive reforestation initiatives is unknown. Building on the carbon modeling foundations established here, the next chapter addresses this knowledge gap by quantifying impacts on forest carbon accumulation and storage. Specifically, it analyzes time-series RapidEye imagery from 2012-2019 using an enhanced Extreme Gradient Boosting model to assess changing carbon stocks across Buffelsdraai. Comparative assessment of spatial and temporal patterns provides a localized understanding of variability in accumulation rates based on site ecological histories. Findings equip urban forest policymakers with an adaptive, evidence-based methodology for continually tracking emergent climate regulation benefits towards maximizing and prioritizing reforestation investments. Overall, this pioneering application stretches the capabilities of integrated remote sensing and machine learning to strengthen quantifiable, transparent climate action policy.

5 CHAPTER FIVE: Quantifying the impact of urban reforestation on carbon accumulation using RapidEye satellite imagery and extreme gradient boosting regression.

This Chapter is based on:

Collins Matiza [1], Onisimo Mutanga[1], Kabir Peerbhay [1], Romano Lottering [1], John Odindi [1, *], and Mthembeni Mngadi [1] 2023. Quantifying the impact of urban reforestation on carbon accumulation using RapidEye satellite imagery and extreme gradient boosting regression. *Ecological Applications. (Under Review)*

Abstract: Urbanization contributes to environmental change through deforestation and forest degradation, subsequently increasing atmospheric carbon emissions leading to climate change. Hence, reforestation has been identified as a reliable long-term approach for carbon sequestration and climate change regulation in urban landscapes. Despite various studies quantifying and mapping the spatial distribution of urban forest carbon stocks resulting from reforestation efforts, there is a noticeable paucity in literature on evaluation of the spatio-temporal aspects of carbon accumulation within urban reforested landscapes. This study employed RapidEye spectral data, topographic and bio-climatic variables, land use, and forest inventory data to model spatio-temporal variations in carbon stock in a reforested urban landscape in Buffelsdraai, KwaZulu-Natal province, South Africa. Using the extreme gradient boosting algorithm, the study computed the aboveground carbon stock between 2012 and 2019 and extracted vegetation fractions to determine areal forest extents in the respective years. Findings of the study show a net increase of 132 hectares in forest cover and an annual rate of forest cover change of approximately 18.9 ha per year, leading to a 7192-tonne corresponding increase in aboveground carbon stocks during the study period. These findings, achieved by integrating remote sensing and field inventory methods, provide valuable insights and a basis for policy formulation, planning and implementation of urban reforestation projects. Furthermore, the findings of the study can be used to facilitate sustainable urban forest management (SFM), necessary for provision of the valuable socio-ecological goods and services, CO_2 assimilation and climate change mitigation.

Keywords: Ecological balance, Urban vegetation, Nature-based solutions; Rapid-Eye imagery, Extreme Gradient Boosting

5.1 Introduction

Urbanization, typified by the transformation of natural landscapes into impervious built-up surfaces has been identified as a major driver of environmental change (Ren et al., 2012, Mngadi et al., 2022b). Such transformation deteriorates the integrity of natural ecosystems, leading to among others degradation of water quality, increased thermal heat and air pollution, and habitat fragmentation and loss (Zhao et al., 2020, Li et al., 2016a, Zhang, 2022). Despite covering small land surface, urban areas contribute the most significant volumes of atmospheric carbon emissions due to high energy and resource consumption (Cheng and Xiao, 2022, Pouyat et al., 2002). Consequently, urbanization is characterized by significant deforestation and forest degradation that lead to decreased carbon assimilation potential and accelerated climate change risks (Grimm et al., 2008, Mngadi et al., 2021a, Li et al., 2016a). Recently, urban reforestation (UR) has emerged as a long-term and reliable alternative approach to atmospheric carbon sequestration and climate system regulation within urban landscapes (Pasher et al., 2014, Nowak et al., 2013, Sun and Chen, 2017). Specifically, reforesting diverse indigenous trees within urban landscapes offer great potential to local and regional carbon budget and climate systems and improve environmental quality (Ogunbode and Asifat, 2021, Scholz et al., 2018).

Whereas numerous studies have assessed the spatial distribution of carbon stocks in reforested urban landscapes (Huang et al., 2014, Lee et al., 2019, Myeong et al., 2006), multi-temporal analysis of reforested carbon stocks for effective monitoring of its accumulation remains unavailable. Such spatiotemporal analysis is essential in evaluating the successful implementation of among others Reducing Emissions from Deforestation and Forest Degradation (REDD+) initiatives (Nakakaawa et al., 2011, Bhattarai et al., 2015). Furthermore, knowledge on spatio-temporal distributions is critical for monitoring and managing disturbances of forest carbon sinks. However, achieving this objective requires a cost-effective and reliable dataset analyzed within robust carbon modeling techniques. Whereas traditional approaches are known to be highly accurate, they are costly, laborious and impractical in remote areas (Safari et al., 2017, Sibanda et al., 2015, Gara et al., 2016). Hence, the emergence of remote sensing datasets and techniques has proven instrumental in providing reliable and cost-effective primary data for precise carbon estimation (Dahy et al., 2020, Odebiri et al., 2023, Mngadi et al., 2022b).

Remote sensing provides robust spectral information at larger spatial coverage, important for local and regional modeling of forest carbon stock (Yao et al., 2015, Mngadi et al., 2021a, Dube and Mutanga, 2015). Furthermore, the recent emergence of freely and readily available multispectral sensors such as Sentinel and Landsat series have become popular in estimating forest carbon stock, especially in financially constrained and resource scarce regions such as Southern Africa (Dube and Mutanga, 2015, Gara et al., 2016, Muhe and Argaw, 2021). These sensors cover sensitive regions (such as visible, near-infrared and shortwave near-infrared) in the electromagnetic spectrum necessary for concise modeling of vegetation green-biomass and carbon sinks (Mngadi et al., 2022b, Gara et al., 2016). Furthermore, the sensors improvements in spatial, spectral and radiometric properties have further facilitated their adoption in modeling spatio-temporal patterns of forest biomass and carbon stock (Dube and Mutanga, 2015, Szatmári et al., 2019, Ren et al., 2019). However, despite the ability of freely available sensors to estimate forest carbon stocks with acceptable accuracy, spectral saturation continues to pose a significant challenge when mapping dense forest cover and heterogenous landscapes characterized by highly diverse trees such as mixed species reforestation sites. Consequently, concise and reliable wall-to-wall forest carbon modeling particularly in a highly mixed forest landscape such as urban reforestation sites necessitate adoption of high spatial resolution satellite sensors.

The RapidEye is a high spatial resolution (5m) sensor that provides spectral data using fewer (5 bands) but robust wavebands that are strategically positioned within the electromagnetic spectrum (Dube et al., 2014a). RapidEye's fine spatial resolution reduces challenges related to mixed pixel between tree canopy, soil background and shadow cast that could impede prediction performance of forest carbon stocks modelling (Asrat et al., 2018, Dube et al., 2014a). Furthermore, RapidEye has a relatively rich spatial data archive that effectively map and monitor the temporal variability of forest carbon stock (Otunga et al., 2019). Also, the rich information that can be derived from the RapidEye image data, especially indices, has not been fully explored in understanding spatio-temporal variations in carbon accumulation within a reforested urban landscape (Hojas-Gascon and Hugh, 2014, Su et al., 2016). Adopting RapidEye's vegetation indices is particularly valuable due to their sensitivity in vegetation green-biomass spectral response. Furthermore, vegetation indices alleviate the effects of senesced vegetation, atmospheric effects, and soil background associated with satellite images, hence providing refined spectral data that accurately model and map forest carbon stock (Grinand et al., 2017, Mngadi et al., 2021a). In this regard, there is a need to

explore RapidEye derived indices in understanding spatio-temporal variations in carbon accumulation within a reforested urban landscape.

Nonetheless, spatial data alone are often insufficient to describe the patterns of aboveground carbon stocks that often manifest in three-dimensional structure. Several studies have proposed the use of multiple predictors as covariates to account for vertical properties variation (Shen et al., 2018b, Salinas-Melgoza et al., 2018). The storage and long-term maintenance of aboveground carbon (AGC) in diverse ecosystems exhibit notable variations (Shen et al., 2018a, Salinas-Melgoza et al., 2018). These disparities are not solely determined by intrinsic tree structural characteristics, such as volume, stem density, leaf area index, and height but are also influenced by dynamic factors such as bio-climatic conditions, plant density, topographic properties, and historical disturbances (Baines et al., 2020). Precipitation and temperature are bioclimatic parameters contributing to biomass accumulation and positively correlate with biomass density. However, predominant factors affecting aboveground carbon in urban reforested landscapes across varying geographic scales remains unresolved (Mitchell et al., 2018a, Mngadi et al., 2021a). Therefore, it is essential to continuously map and monitor the spatial variations in AGC stocks within urban reforested landscapes using field observations, remote sensing and machine learning techniques. These ongoing efforts play a critical role in informing comprehensive policies and initiatives aimed at preserving existing carbon stocks and mitigating carbon emissions.

Literature indicates that remote sensing provides highly correlated spectral information between and within predictor variables (Wulder and Boots, 1998, Landgrebe, 2002, Zhang et al., 2019a). This challenge could hinder an accurate statistical analysis of remotely sensed data. Thus, it is necessary to adopt robust and advanced statistical algorithms capable of dealing with data overfitting or multicollinearity. For instance, a non-parametric algorithm such as extreme gradient boosting (XGBoost) is powerful in mapping and modeling vegetation green-biomass and carbon stock (Pham et al., 2020a, Uniyal et al., 2022, Liu et al., 2021). XGBoost uses bootstrapping and bagging techniques, and repeated iteration procedure that enable precise data mining critical for enhancing forest carbon mapping and modeling (Bort Escabias, 2017, Chen and Guestrin, 2016b). The algorithm builds a series of decision trees based on stepwise model fitting approach, which is critical for reducing error of bias and averaging technique for minimizing model variance (Bort Escabias, 2017, Chen and Guestrin, 2016b). Weak learners are given a higher weighting in each decision tree. In contrast, strong learners are reduced, and the process is repeated until the error in prediction is reduced to a minimum, and predictions of explanatory variables are accurate. The XGBoost further

integrates cross-validation and elimination techniques to select the optimal number of predictor variables and enhance the prediction performance of forest carbon stock (Li et al., 2019b, Uniyal et al., 2022). These optimization techniques have prompted the application of XGBoost in modeling and mapping vegetation biomass and carbon stock using remotely sensed spectral data. Therefore, this study sought to develop a framework for estimating and quantifying the impact of urban reforestation on carbon accumulation leveraging a combination of RapidEye satellite imagery, forest cover data, topographic data, bioclimatic data, vegetation indices, and forest inventory data within an extreme gradient boosting (XGBoost) algorithm environment. This framework will be employed to precisely estimate the dynamic spatial and multi-temporal distribution maps of carbon stock in a reforested urban landscape.

5.2 Methods

5.2.1 Field data collection and inventory data

Data collection and field-survey were carried out from the 4th of February to the 24th of August 2019 during the summer season when biomass productivity was at its peak. In this study, approximately 194 predetermined randomly selected samples were uploaded into a Global Positioning System (GPS) and used to navigate the site. At each random point, a 5 m x 5 m plot-size window was implemented and numerous tree structural and volumetric attributes, such as the tree's height and diameter at breast height (DBH), were measured and documented. For measuring the height of trees, a clinometer was used, while the diameter of the trunk was measured with a caliper. Garmin Trimble GPS was used to record the location of the sampled trees with a precision of 0.5 m. An increment borer was used to measure wood density for the dominant species. Increment borers are long, hollow drill bits that are specifically designed for extracting small, cylindrical cores from trees. Archival inventory records (i.e., 2012 to 2018) of existing plots were also used for the multi-temporal analysis.

Despite challenges relating to labor and transportation, the location of sampling plots took into consideration many environmental considerations, such as a wide range of biomass, various topographic characteristics and tree species (Table 5.1), for full details, see supplementary material). Effort was made during the sampling process to reduce similarity between adjacent sampling sites to minimize spatial autocorrelation issues in statistical analyses and ensure that the samples within the region were representative.

Table 5.1 Summary of biotic and topographic characteristics of sample plots.

Sample plot	Aboveground biomass (t.ha-1)	Tree species (botanical name)	Slope	Aspect	Wood Density g/cm^3
170	108.77	*Brachylaena discolor*	20	SE (135°)	0.6- 0.9
85	210.08	*Acacia robusta*	11	SE (135°)	0.6- 0.9
166	100.9	*Antidesma venosum*	11	SE (135°)	0.5- 0.9
135	75.77	*Bridelia micrantha*	11	SE (135°)	0.5- 0.9
96	50.34	*Ekebergia capensis*	11	SE (135°)	0.6 – 0.8
151	78.88	*Milletia grandis*	11	SE (135°)	0.6 – 0.8
149	29.54	*Harpephyllum caffrum*	11	SE (135°)	0.6 – 0.85
14	63.75	*Protorhus longifolia*	12	NE (45°)	0.6 – 0.8
159	43.33	*Bauhinia natalensis*	12	SW (225°)	0.6 – 0.8
11	190.88	*Ficus natalensis*	12	SE (135°)	0.5 – 0.8
60	17.63	*Protorhus longifolia*	12	SE (135°)	0.6 – 0.8
78	8.53	*Tabernaemontana ventricosa*	12	E (90°)	0.6 – 0.8
66	25.46	*Erythrine caffrum*	8	NE (45°)	0.6 – 1.2
145	17.55	*Harpephyllum caffrum*	8	NW (315°)	0.6 – 1.2

5.2.2 Allometric computation of aboveground carbon stock

Numerous studies have shown that the relationship between tree height and diameter at breast height (DBH) significantly influences the production of aboveground biomass (Ali and Mattsson, 2017, Xiao and Ceulemans, 2004, Vahedi et al., 2014). Therefore, measuring these tree characteristics can be useful in calculating in-situ biomass (Chave et al., 2014)). In this study, we used an allometric equation (Eq. 1) developed by Chave et al. (2014),developed to determine the aboveground biomass (AGB) for general tropical tree species Eq. 5.1. This equation takes into consideration wood density, tree height, and DBH as predictive factors. It is worth noting that wood density plays a crucial role in AGB calculations and varies among different tree species, regardless of their height and DBH (Baker et al., 2004).

$$AGB = 0.0673 \times \left(p^{D^2H}\right)^{0.976} \quad \text{Eq. 5.2}$$

Where *p* represents wood density constant (1.2 g/cm^3), D is the diameter (cm) and H is the height (m). Commonly, studies have affirmed that aboveground vegetation biomass is responsible for about 50% of carbon storage, consequently, a generic value of 0.5 can be used as a conversion factor of wet biomass into dry carbon stock (Fang et al., 2001, Gara et al., 2016). Following established practices, a conversion factor of 0.5 was employed to convert dry biomass into carbon stock.

Spatial autocorrelation

The Global Moran's I proposed by Cliff and Ord (1972) is a widely employed spatial autocorrelation index. In this study, Global Moran's I was applied to annual carbon estimates obtained from sample plots (allometric derived carbon) to determine the relationship between sampled plots. This index yields values ranging from 1 to -1, where a value of 1 signifies positive spatial autocorrelation (i.e., cluster patterns), while a value of -1 indicates negative spatial autocorrelation (i.e., checkerboard patterns). A value of 0 indicates no spatial autocorrelation, i.e. random patterns (Koenig, 1999, Hu et al., 2020). In accordance with the queen's case of spatial contiguity, we computed the Global Moran's I using the Moran function within the R Spdep package. This calculation yielded a single spatial autocorrelation value for each year throughout the study period. The Global Moran's I Eq. 5.3 (denoted as "I") is expressed as follows:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(X_i - \hat{X}\right)\left(X_j - \hat{X}\right)}{\left(n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right) n\sum_{i=1}^{n}\left(X_j - \widehat{X}\right)^2}\,(i \neq j) \qquad \text{Eq. 5.3}$$

In our analysis, 'n' denotes the total number of distinct locations within each annual measurement of carbon stock. The variables 'xi' and 'xj' correspond to the carbon stock values measured at specific locations 'i' and 'j,' respectively. The variable 'x' signifies the mean carbon stock value across all sampled locations, while 'wij' represents the measure of spatial proximity or connectivity between these locations. In years where a positive spatial autocorrelation was observed, a spatial weights matrix was constructed. This matrix assigns specific weights to individual observations, effectively rendering the spatial relationships between these observations stochastic and unrelated.

5.2.3 Image acquisition and data preprocessing

In this study, a series of Rapid-Eye multispectral image scenes captured from 18th December 2012 to 14th December 2019 were downloaded on the 22nd of February 2023 from Planet Labs image explorer website Planet Explorer. These image scenes were acquired during summer season favorable for maximum biomass production and when cloud cover was minimal, i.e., between 0 and 5% (Table 2). A RapidEye image comprises 5 spectral wavebands that capture spectral information at a spatial resolution of 5 meters, a ground sampling distance (nadir) of 6.5 meters, and a swath width of 77 kilometers (Jung-Rothenhäusler et al., 2007, Krischke et al., 2000). The Rapid Eye wavebands are located in critical regions of the electromagnetic spectrum such as visible (440-685 µm), red edge (690-730 µm), and near-infrared (760-850 µm). The Planet Labs Explorer provides Rapid-Eye Imagery and offers radiometric and geometric corrections as part of its Level 3A (RapidEye, 2012) calibration that is calibrated to the at-sensor radiances. Furthermore, this study extracted spectral data based on different waveband combinations representative of vegetation indices (Table 3). Based on previous studies on forest carbon stock prediction (Safari et al., 2017, Rijal et al., 2023, Mngadi et al., 2021a), normalized difference vegetation index (NDVI), green normalized vegetation index (GNDVI), enhanced vegetation index (EVI) and red-edge normalized vegetation index (RENDVI) have been proven optimal in predicting forest carbon stock, hence the prediction models adopted these vegetation indices (Table 5.2).

Table 5.2. Sensor type and image acquisition date of Rapid-Eye data.

Cloud%	Acquisition date	Satellite Sensor	Specifications
0	18 December 2012	Rapid-Eye-3	5 spectral bands, 6.5m resolution
0	18 December 2013	Rapid-Eye-3	5 spectral bands, 6.5m resolution
2	18 December 2014	Rapid-Eye -4	5 spectral bands, 6.5m resolution
3	21 December 2015	Rapid-Eye-3	5 spectral bands, 6.5m resolution
0	22 December 2016	Rapid-Eye-4	5 spectral bands, 6.5m resolution
0	23 December 2017	Rapid-Eye-4	5 spectral bands, 6.5m resolution
0	16 December 2018	Rapid-Eye -4	5 spectral bands, 6.5m resolution
5	14 December 2019	Rapid-Eye-5	5 spectral bands, 6.5m resolution

5.2.4 Topographic metrics

In this study, topographic variables were integrated into the prediction model to understand forest carbon stock variability over-time (Lin et al., 2012, Soriano-Luna et al., 2018, Odebiri et al., 2023). Literature has demonstrated that slope, elevation and aspect can significantly influence carbon stock spatio-temporal distribution (Lin et al., 2012, Soriano-Luna et al., 2018). Hence, this study consolidated slope, elevation, roughness and Topographic Wetness Index (TWI) (Table 2) into the prediction model. To obtain these topographic variables, we used the Quantum Geographic Information System (QGIS) (version 3.24.4) software to extract them from the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM).

5.2.5 Bio-climatic variables

The Modern Era Retrospective-analysis for Research and Applications climate (MERRAclim) consists of three datasets, each encompassing 19 bioclimatic variables generated over the past three decades (C Vega et al., 2017, Bazzato et al., 2021, Waltari et al., 2014). These datasets are constructed using hourly data from January 1, 1981, to December 31, 2010. The bioclimatic variables within MERRAclim are derived from geographically consistent temperature and specific humidity gridded data, benefiting from a standardized assimilation technique applied on a global scale, including Antarctica. It is worth noting that MERRAclim datasets are built upon the foundation of MERRA data, a dataset that has undergone extensive validation in scientific literature (C Vega et al., 2017). MERRAclim provision of hourly data offers a high temporal resolution, which is of paramount importance for capturing the nuanced variations in climate variables over time. This level of granularity is particularly valuable for assessing the impact of climatic factors on ecological processes and carbon stocks (Bazzato et al., 2021). In our machine learning model, we incorporated five MERRAclim climatic variables shown in Table 5.3 from bio-climatic variable 1, 2,4,5, and 8 with a 10 m resolution, harnessing the richness of the dataset to comprehensively analyze and understand the complex interactions between climate and ecological dynamics.

Table 5.3 Vegetation indices, topographic features and land use classification maps used in the construction of extreme gradient boosting model.

Remote sensing data	Indices or Parameters	Formula/description	Source
Vegetation Indices	NDVI	$\frac{NIR - Red}{NIR + Red}$	(Pettorelli et al.,

			2005)
	EVI	$2.4\frac{NIR - Red}{NIR + 2.4 * Red + 1}$	(Jiang et al., 2008)
	GNDVI	$\frac{NIR - Green}{NIR + Green}$	(Shaver et al., 2006)
	RENDVI	$\frac{NIR - RedEdge}{NIR + RedEdge}$	(Dhumal et al., 2017)
Terrain	Elevation		(Sharma et al., 2020)
	Slope		(Sharma et al., 2020)
	Roughness		(Sharma et al., 2020)
	TWI		(Sharma et al., 2020)
Landuse and landcover	Classification maps		(Hernández-Guzmán et al., 2019)
MERRA-CLIM Bio-climatic data	Temperature and rainfall data for 2000s		(C Vega et al., 2017)
Bio-1	Temperature	Annual mean temperature (10m res)	(C Vega et al., 2017)
Bio-8	Temperature	Mean temperature of the most humid quarter	(C Vega et al., 2017)
Bio-4	Temperature	Temperature seasonality (10m res)	(C Vega et al., 2017)
Bio-5	Temperature	Maximum temperature of the warmest month (10m res)	(C Vega et al., 2017)
Bio-2	Precipitation	Annual mean precipitation (10mres)	(C Vega et al., 2017)

5.3 Statistical analysis

5.3.1 Correlation analysis of variables

Correlation analysis is used to examine the correlation matrix of variables to identify pairs with strong correlations. The correlation matrix provides a numerical measure indicating the strength and direction of the linear relationship between two variables, ranging from -1 to 1. A correlation coefficient of 1 signifies a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 implies no correlation (Figure 5.1).

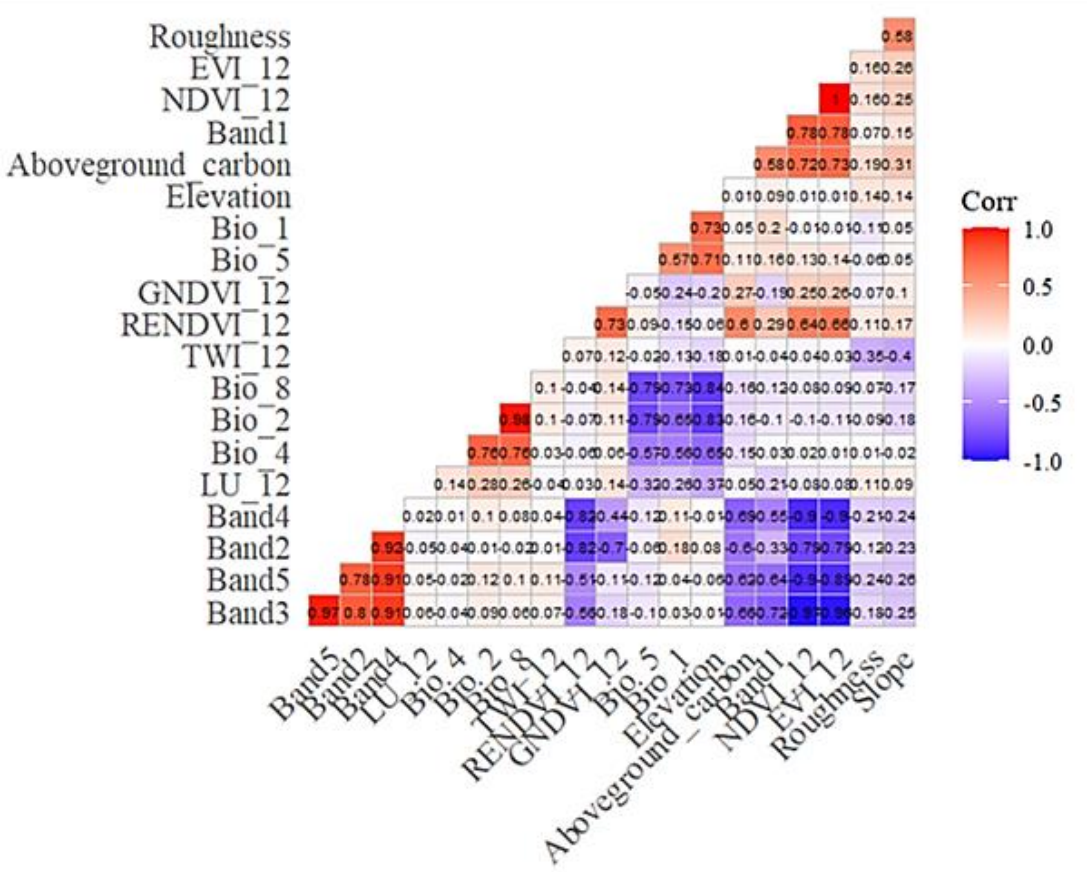


Figure 5.1.Correlation analysis of input variables.

In this study, Pearson correlation analysis was conducted using R software to investigate the relationships among RapidEye waveband features, derived vegetation indices, topographic features, landuse features, bioclimatic features, and aboveground carbon data. Features from the training dataset exhibiting significant correlation (correlation coefficient (r) absolute value greater than 0.1) with control data (derived carbon) were selected for subsequent estimation of carbon stocks. Since correlations amongst variables were not as strong, all the 19 input variables were incorporated into the model.

5.3.2 Forest cover change

Change detection encompasses the utilization of multi-temporal datasets to discriminate regions exhibiting alterations in forest cover over distinct imaging dates (Coppin and Bauer, 1994). This methodological approach serves to ascertain the locations and magnitudes of observed changes. Simultaneously, it addresses three pivotal aspects of forest cover change detection, which encompass the discernment of change occurrences, characterization of the nature of these alterations, and quantification of the spatial extent of such transformations (Coppin and Bauer, 1994). Furthermore, a change detection matrix was generated to elucidate broader trends and patterns pertaining to land use and land cover alterations, particularly with regard to changes in forest cover. The rate of forest cover change in this study was calculated using the formula Eq. 5.4:

$$R = \frac{(t_2 - t_1)}{a} \qquad \text{Eq. 5.4}$$

There are 3 components to this equation: r = rate of change in forest cover, t_2 = recent forest cover in ha, t_1 = initial forest cover in ha, and a = number of years between t_2 and t_1.

5.3.3 Above ground carbon stock estimation

To predict and map aboveground carbon stocks, the study integrated multiple data sources and analytical techniques. This involved calculating aboveground biomass using established allometric equations that were tailored to the major tree species within our study site intended to comply with guidelines provided by the Intergovernmental Panel on Climate Change (IPCC) in 2004 (IPCC, 2006). To determine the aboveground biomass for the entire study area, we applied these equations to the data collected from field sampling plots, ensuring that each plot was matched with the allometric equation.

Following the recommendations outlined by IPCC in 1995 (Carter et al., 1994), we consistently applied a carbon conversion factor to estimate carbon stocks for each of the sampled plots, leveraging the calculated allometric biomass. This plot-level carbon stock data was then employed to calculate the overall carbon density across the entire study site using the extreme gradient boosting regression model. This modeling process was repeated for each year under study from 2012 to 2019, with calculations relying on annual forest inventory data coinciding with RapidEye satellite imagery and auxiliary data (bioclimatic data, topographic variables derived from SRTM-DEM, forest cover maps and derived vegetation indices). This

comprehensive approach enabled the generation of detailed temporal maps that reveal the distribution of aboveground carbon stocks within the study area over the studied time frame.

This study adopted extreme gradient boosting (XGBoost) algorithm in estimating spatio-temporal variability of carbon stock in reforested urban landscapes, known for its high processing speed and accurate predictions, especially with complex datasets (Gašparović and Dobrinić, 2020, Ahirwal et al., 2021, Zhang, 2022). XGBoost excels in reducing bias error among spectral predictors by identifying the model with the lowest error rate (Chen and Guestrin, 2016b, Valavi et al., 2022). This iterative process involves incrementally weighting weak learners and diminishing strong learners until the error level is minimized, ensuring precise estimation of explanatory variables. Furthermore, the model tuning hyperparameters (e.g., *ntree*, node-size and iteration) were adjusted to optimize the prediction performance of spatio-temporal variability of carbon stock over time. *Ntree* is critical for creating a series of decision trees and choose the most precise tree subset based on the lowest root mean square error for optimal regression analysis (Valavi et al., 2022, Chen and Guestrin, 2016b). Meanwhile, a node-size identifies tree subset terminal node with the smallest number of observations, and the default value is 1 (Kim et al., 2022, Chen and Guestrin, 2016b). The iteration techniques are a repetition procedure of decreasing or increasing the weight of predictors in a training dataset to reduce potential errors in the prediction model. In this study, the *ntree* was set at 200, 300 and 500, with the 0.1, 1 interval, gradient boosting iterations from 100 to 1000, and the node-size kept at default value of 1.

5.4 Model evaluation and validation.

To test the effectiveness of the XGBoost regression model in estimating the spatio-temporal distribution of carbon stock in a reforested urban landscape, a 70/30 hold-out validation technique was used (Shahani et al., 2021). The total data sample (n = 194) was initially partitioned into a 70% training set (n = 118) and a 30% testing set (n = 76). The final model was validated using the coefficient of determination (R^2) Eq. 5.5 and root mean square error (RMSE).

$$R^2 = 1 - \sum \frac{(y_i - \bar{y}_i)^2}{\sum (y_i - \hat{y}_i)^2} \qquad \text{Eq. 5.5}$$

Where: y_i is the actual value of the dependent variable for the ith observation, $\bar{y}$_i is the mean of the actual values of the dependent variable,$\hat{y}$_i is the predicted value of the dependent

variable for the ith observation. The R-squared value ranges from 0 to 1, with a value of 0 indicating that the independent variable(s) has no explanatory power for the dependent variable, and a value of 1 indicating that the independent variable(s) completely explain the variation in the dependent variable. In general, an R^2 value of 0.7 or higher is considered a good fit, while an R^2 value of 0.9 or higher is considered a very good fit Eq. 5.6.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{Observed} - X_{Predicted})^2}{n}} \qquad \text{Eq. 5.6}$$

Where *Xobserved* is the measured carbon values, *Xpredicted* is the predicted carbon stock values, i represent each of the predictor variables included in the [summation process and *n* represents the number of observations (number of collected plot samples) it. *∑(i=1) ^n* specifies the range over which the summation is performed starting from i = 1 up to n, representing each data point in the dataset.

The Shapely Additive Explanations (SHAP) methodology was applied to assess the relative importance of individual predictor variables in determining the overall performance of the XGBoost model(Movsessian et al., 2021) . The SHAP technique, facilitated through the 'ModelExplainer' feature embedded in the DALEX and ModelStudio packages, assigns specific magnitudes of importance to each predictor variable based on the XGBoost output model. The rationale behind adopting the SHAP method lies in its demonstrated robustness in both global and local interpretability, distinguishing it as a superior choice in comparison to alternative techniques (see (Movsessian et al., 2021, Bifarin, 2022) for a comprehensive exploration of the SHAP method).

5.5 Results

5.5.1 Field measured reforested carbon stocks.

Based on the descriptive statistics (Table 5.4), the mean carbon stock value from 2012 to 2019 ranges from 53.87 to 84.42 t. ha^{-1}, with a standard deviation of 28.14 to 38.73 t. ha^{-1}.

Table 5.4. Descriptive statistics of the measured above ground carbon stock (in t·ha^{-1}).

Year	Mean	Standard deviation	Minimum	Maximum
2012	57.82	26.75	1.05	116.33
2013	58.97	26.75	4.05	115.23
2014	56.71	28.14	4.85	114.64
2015	57.19	32.90	7.05	151.70
2016	53.87	30.37	7.05	124.65
2017	78.42	38.73	10.05	163.84
2018	82.42	38.73	13.05	163.84
2019	84.42	38.73	14.05	163.84

The annual carbon plot data was subjected to a spatial autocorrelation analysis. In cases where the analysis identified spatial clustering in certain years (2014, 2016, 2017,2018 and 2019), they were assigned weights based on inverse distance measurements of 10 meters. Subsequently, a follow-up assessment was conducted using Global Moran's I to validate the randomness and spatial independence of the entire dataset.

5.5.2 Reforested Area Change Analysis

Major forest loss occurred in 2013-2014 and 2015-2016, reducing forest cover from 38% to 36% and 42% to 41%, respectively. The highest gain in forest area was 51 hectares, while the smallest gain recorded was 6 hectares, with an average change of 18.9 hectares over the study period (Table 5.5).

Table 5.5 Calculated Forest area and change in forest area.

Period	Total forest area (ha)	Change in forest area	% change
2012	260		
2012-2013	310	50	+19.2
2013-2014	292	-18	-5.8
2014-2015	343	51	+17.5
2015-2016	333	-10	-2.9

2016-2017	350	17	+5.1
2017-2018	386	36	+10.3
2018-2019	392	6	+1.6

5.5.3 XGBoost model performance evaluation

The results in Table 5.6 represent the overall mean of predicted carbon stock and prediction performance of the multi-source data and XGBoost regression model. The model shows that mean carbon stock increased from 2012 to 2013 (35. 73t.ha-1 to 36. 77t.ha-1), while a slight decrease in predicted carbon was observed in 2014 and 2016, i.e., 34.76 t. ha-1 and 34.33 t. ha-1, respectively. Between 2017 and 2019, the mean carbon stock increased consistently. Furthermore, the XGBoost regression model achieved high prediction performance (R^2: 0.76-to-0.83) with low RMSE (16.44-to-31.39 t. ha^{-1}) in estimating the spatio-temporal distribution of carbon stock between 2012 and 2019. The results in Figure 3 represent the relationship between measured and predicted carbon stock using RapidEye spectral information from 2012 to 2019. The regression analysis shows a strong correlation coefficient (r: 0.66 to 0.90) between measured and predicted carbon from 2012 to 2019 (Figure 5.2).

Furthermore, Figure 4 summarizes the relative contribution of the 19 predictor variables to the prediction of AGC stocks using the SHAPley method in XGBoost models. The top 5 most influential predictors of AGC using the XGBoost model across the study period included Bio-2 (precipitation), GNDVI, band 4 and SLOPE. In contrast topographic, landuse and other bioclimatic variables like elevation, Bio-8 (temperature), Bio-1(temperature), and TWI were not as influential to the model's performance. The combined use of bands, indices, bioclimatic variables, landuse data and topographic variables enhanced the precision of AGC estimation.

Table 5.6 Predictive accuracies using Rapid-Eye, land use and topographic variables for estimating carbon stock overtime.

Year	**Mean carbon stock (t.ha-1)**	**Calibration**		**Validation**	
		R^2	RMSE (t.ha-1)	R^2	RMSE (t.ha-1)
2012	35.73	0.83	16.44	0.81	17.52
2013	36.77	0.76	23.13	0.71	24.03
2014	34.76	0.79	22.33	0.72	25.70
2015	35.62	0.80	23.24	0.77	26.82
2016	34.33	0.83	22.82	0.80	16.11
2017	38.43	0.83	16.95	0.80	19.95
2018	38.5	0.77	27.52	0.72	30.52
2019	40.69	0.86	31.39	0.80	32.80

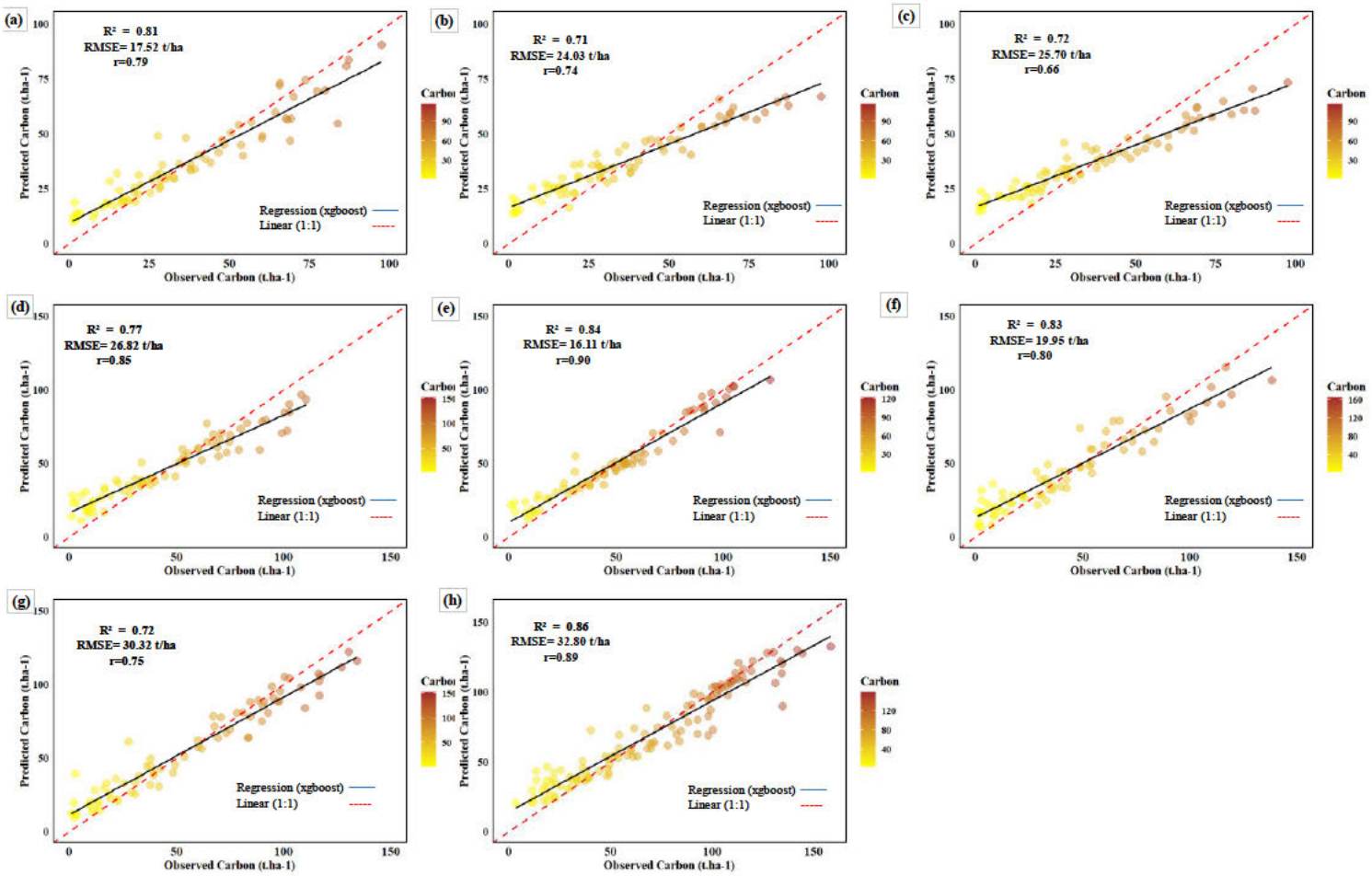


Figure 5.2 Relationship between predicted and observed carbon stock in reforested urban landscapes, 2012-2019 (a-h). Regression analysis was used to establish the relationship using the best variables.

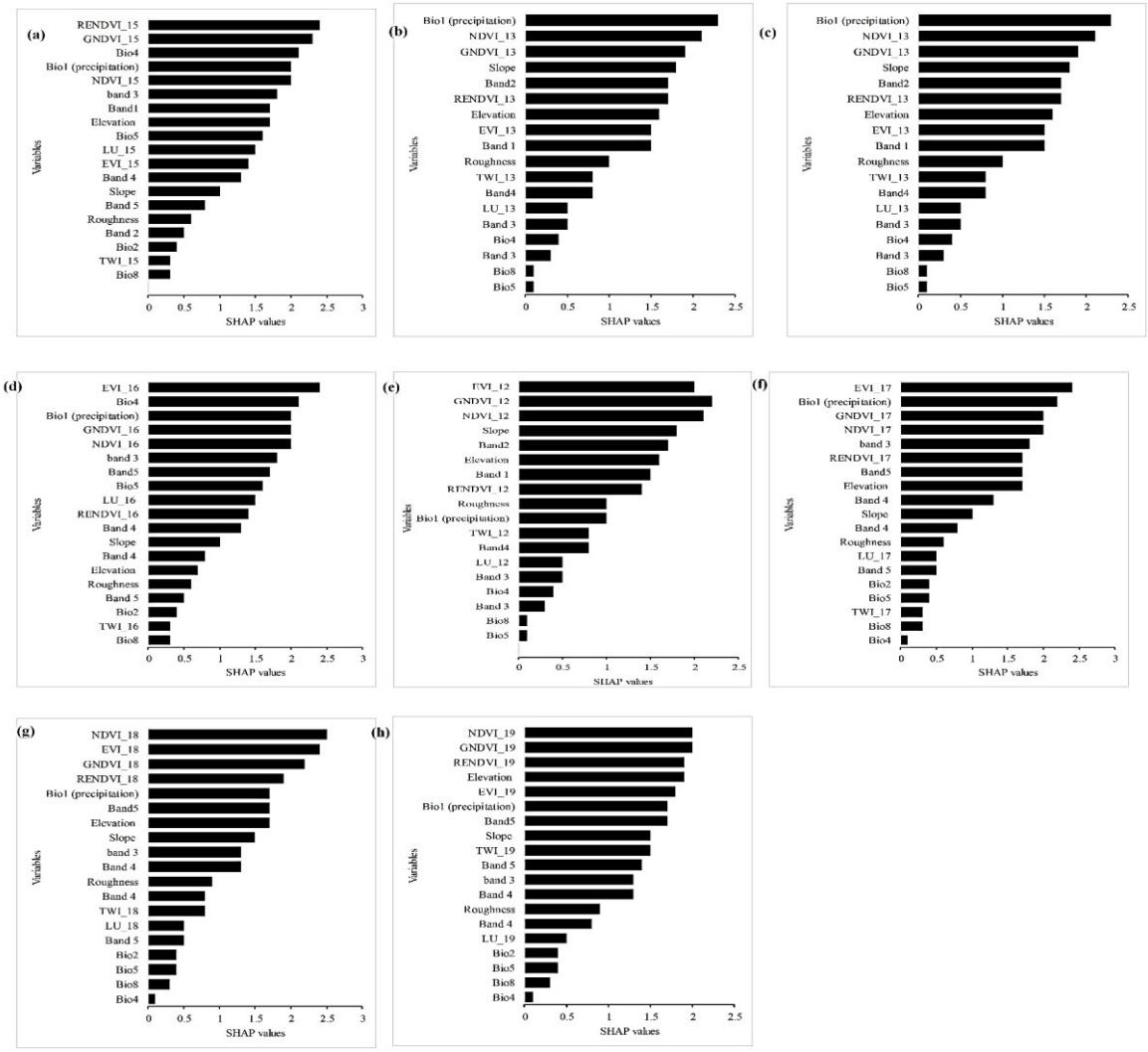


Figure 5.3. Importance ranking of predictors used for the estimation of AGC over the period, 2012-2019 (a-h) respectively.

5.5.4 Temporal changes in reforested carbon stocks

Table 6 provides year-by-year change regarding the carbon stock while Figure 5 illustrates the trend of total forest carbon over the study period. Forest carbon stock decreased by 842 tonnes in 2013-2014 due to forest loss but increased by 2029 tonnes in 2016-2017 due to a 3% forest cover gain (Table 5.7). The changing trend in forest carbon stock resulting from change in forest area is depicted in Figure 5.4. Overall, from 2012 to 2019, the increase in forest area accounted for 132 ha gain, resulting in 7192t increase in corresponding aboveground carbon stocks (Figure 5). Based on the p-value 0.00821, a statistically significant relationship exists between the time interval and the cumulative carbon change (t). Similarly, there was a statistically significant relationship between the total carbon (t) and the period of study signified by a p-value of 0.005 using the t-test.

Table 5.7 Calculated total carbon stock and stock.

Period	Total Carbon stock (t)	Stock difference
2012	9290	
2012-2013	11399	2109
2013-2014	10132	-1267
2014-2015	12210	2078
2015-2016	11421	-789
2016-2017	13450	2029
2017-2018	14861	1411
2018-2019	16482	1621

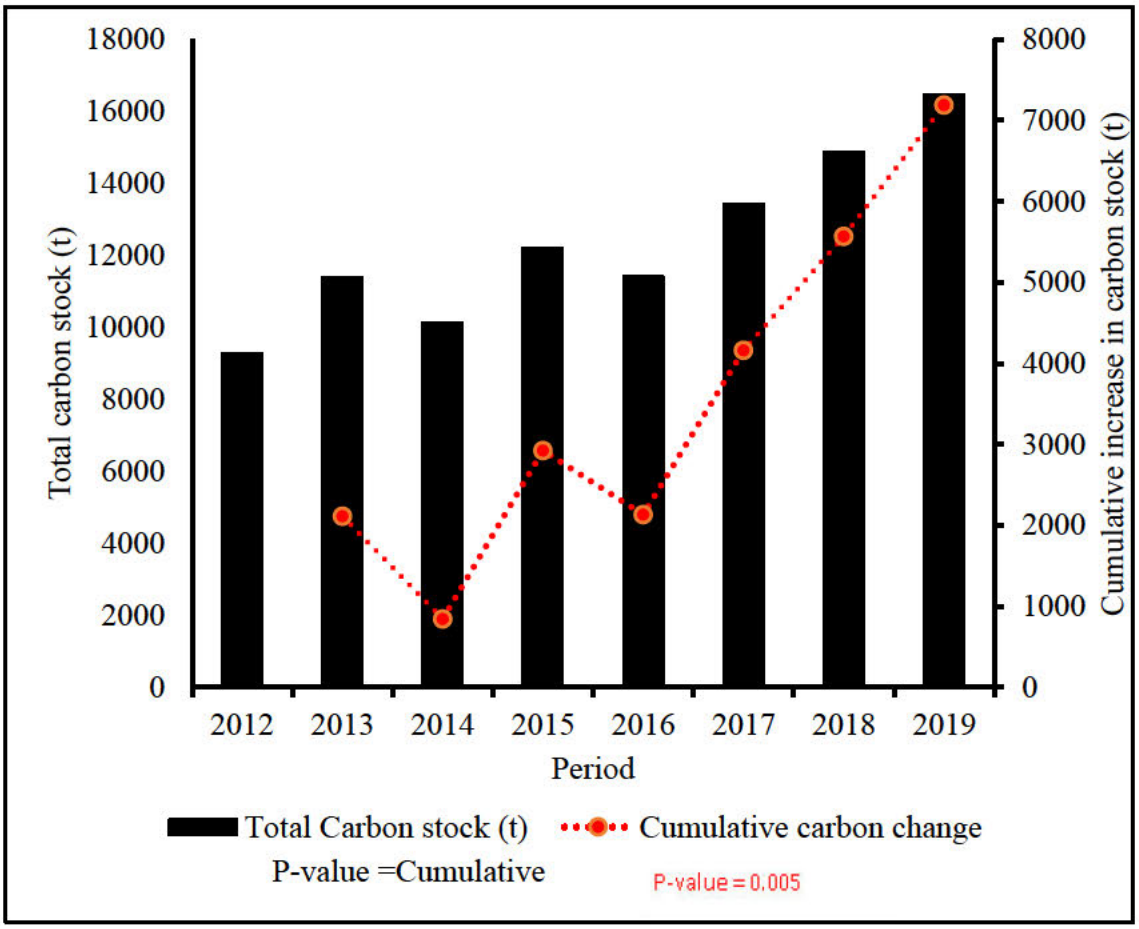


Figure 5.4. Cumulative graph of carbon stock change from 2012-2019.XGBoost algorithm's temporal predictive performance.

5.5.5 Spatio-temporal distribution and variability of AGC stocks

Figure 5.5 illustrates the variation in carbon stock over the seven-year period i.e., from 2012 to 2019. The figures show that in 2012, the carbon stock distribution across the study site was relatively low. However, from 2017 to 2019, there was a significant increase in carbon distribution, indicating higher levels of carbon stock. The prediction maps also highlight a noticeable decline in carbon stock during the years 2014 and 2016. These periods experienced a decrease in carbon stock across the study site.

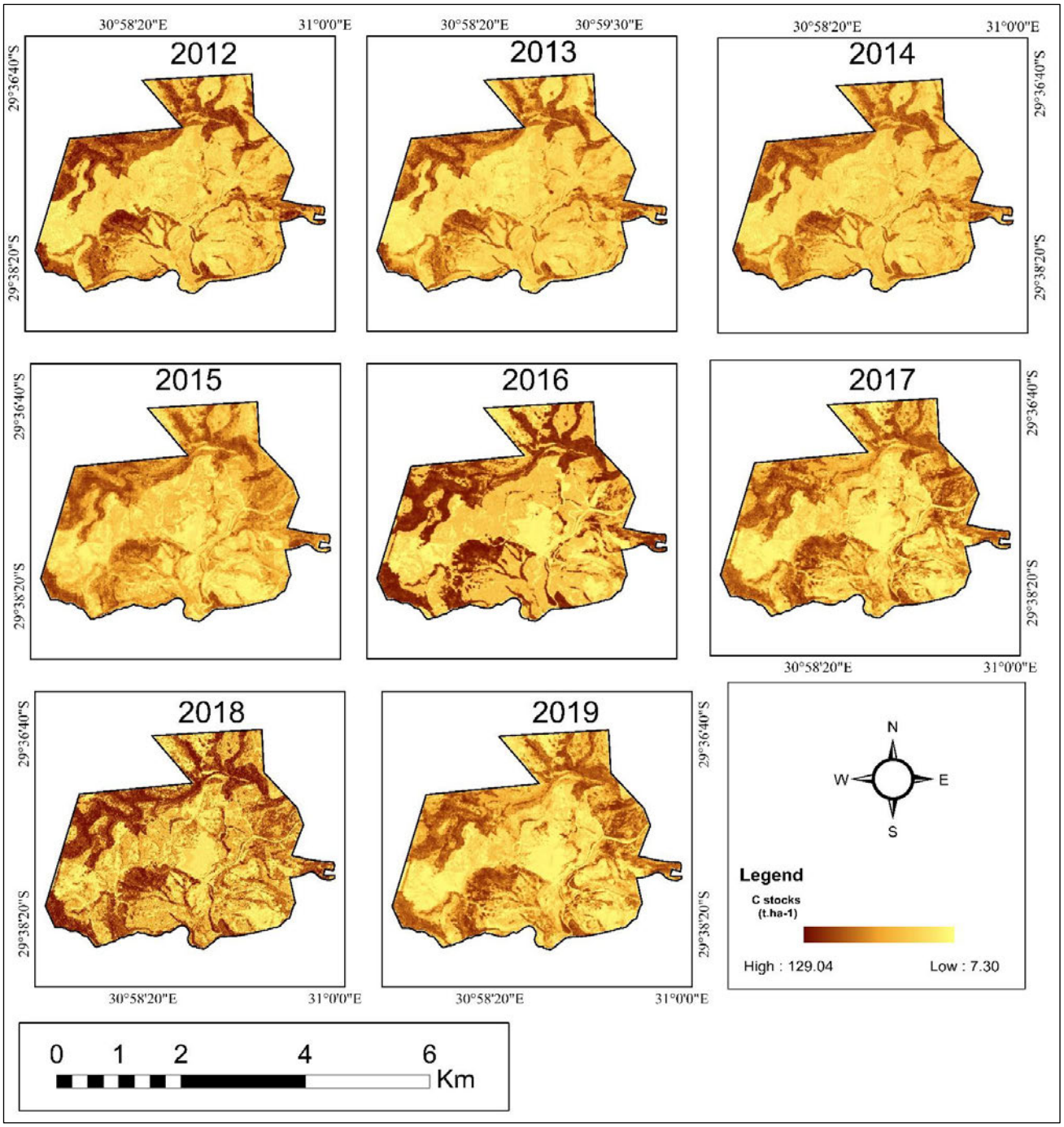


Figure 5.5. Distribution of carbon stock in Buffelsdraai reforestation site using extreme gradient boosting.

5.6 Discussion

5.6.1 XGBoost algorithm's predictive performance of carbon stock

This study underscores the significance of assessing multi-temporal changes in carbon stocks to evaluate the success of urban reforestation projects in improving carbon sequestration and ecosystem health maintenance(Lindsell and Klop, 2013, Li et al., 2022, Di Sacco et al., 2021). It also provides valuable information for decision-making as well as tracking and accounting for the carbon reduction targets. Despite its significance, few studies have explored the spatio and multi-temporal patterns of carbon stocks in reforested urban systems, creating a notable knowledge gap (Di Sacco et al., 2021, Jindal et al., 2008, Wu et al., 2023, Dobbs et al., 2018). Hence, to address this knowledge gap, a robust framework designed for the precise estimation of forest carbon was formulated, subsequently applying it for multi-temporal estimation and analysis of carbon stocks. Utilizing a combination of Rapid-eye spectral data, derived vegetation indices, land use and land cover data, topographic and bioclimatic data and allometry derived carbon data in an extreme gradient boosting algorithm environment.

Whereas individual machine learning algorithms have their limitations, the XGBoost excels in handling densely distributed datasets, uncovering higher-order relationships between features, and providing a more comprehensive description of variable relationships. These qualities contribute to an overall improvement in AGC modeling accuracy. Our study, despite using multi-source data and the XGBoost algorithm, shows that combining spectral data, vegetation indices, bioclimatic variables, topographic variables, and land-use data is superior to using a single dataset and other machine learning approaches (Dantas et al., 2021, Safari et al., 2017). This superiority can be attributed to the utilization of an extreme gradient boosting modeling approach in our study, which effectively incorporates multiple sources of uncertainty into the prediction modeling framework (Cui et al., 2021, Uniyal et al., 2022). This finding is consistent with Zhang et al. (2022c) who leveraged XGBoost model in combination with energy supply data to determine the critical factors influencing the prediction of carbon dioxide emissions of expanding megacities. Their results revealed that using the XGBoost algorithm with energy supply data yielded more accurate carbon emission predictions.

By combining bioclimatic parameters, topographic variables, and optical datasets, the extreme gradient boosting algorithm eliminated saturation issues at higher spatial resolutions and provided accurate estimates of carbon stocks (Fararoda et al., 2021b). Most biomass models

based on remote sensing data are site-specific as they do not account for climatic variables that influence aboveground carbon accumulation. The study used a single model that included bioclimatic variables and topographic covariates, which enabled it to predict carbon stocks with similar accuracy or even better than regional models(Fararoda et al., 2021a).

5.6.2 Evaluating carbon stock trends in peri-urban reforested areas

Our results demonstrate the potential of remote sensing techniques and machine learning algorithms, such as extreme gradient boosting, in monitoring reforestation projects. Furthermore, this study adds to the growing body of literature on carbon stock monitoring in urban landscapes and provides a foundation for future research in the urban reforestation field (Mngadi et al., 2022b, Teo et al., 2021). Over the seven-year period (2012 to 2019), the total forest area (in hectares) exhibited fluctuations. The lowest recorded forest area was 260 hectares in 2012, while the highest was 392 hectares in 2018-2019. This period witnessed an overall net increase of 132 hectares in forest coverage. On average, the annual change in forest area was approximately 18.9 hectares per year, although this rate showed year to year variability as a result of increased tree planting and forest loss due to drought witnessed from 2013-2016.

The results of our study reveal an overall increasing trend in carbon stocks, totaling 9290 t in 2012, 11421 t in 2016, and 16482 t in 2019. Notably, a sharp decline occurred from 12210 t in 2015 to 11421 t in 2016. The positive trend in carbon stocks observed between 2012 and 2019 can be attributed to the successful collaboration between the eThekwini Municipality and the local community in restoring the ecosystem. Planting diverse, fast-growing native tree species i.e. *Syzygium cordatum* (water berry), *Bridelia micrantha* (coastal golden-leaf), *Dalbergia obovate* (climbing flat-bean) and *Erythrina caffra* (African coral tree) (Douwes et al., 2015b) played a crucial role in increasing forest cover and carbon stocks (Teo et al., 2021, Ponce-Hernandez et al., 2004). Moreover, as noted in literature, the restoration efforts led to improved soil health, reduced soil erosion, and enhanced water retention, which helped to improve the growth and health of the trees (Odebiri et al., 2023, Dietrich et al., 2014). This restoration effort, documented in literature (Odebiri et al., 2023, Ren et al., 2019, Sun and Chen, 2017), enhanced soil health, reduced erosion, and improved water retention, fostering tree growth and carbon sequestration. Conversely, the decline in aboveground carbon observed during 2015-2016 can be attributed to wildfires and drought conditions prevalent in South Africa from 2013-2016. Specifically, extreme dryness and heat in December 2013, exacerbated by an El Niño induced drought spanning 2013 to 2016 which reduced rainfall,

likely contributed to moisture stress and increased fire susceptibility thereby leading to lower biomass accumulation in subsequent years. Due to the malfunctioning of the project's water pump (Buffelsdraai reforestation), high levels of tree mortality were reported. (Donnenfeld et al., 2018, Douwes et al., 2015a).

To assess the success of our study in predicting carbon estimates and densities in South African urban reforested ecosystems, we compared our results with recent and existing studies on similar landscapes and terrestrial ecosystems. It is important to note that these comparisons offer valuable insights into the unique characteristics of our study area. Forestry types in the eThekwini municipality exhibited mean carbon densities ranging from 53 to 121 tC. ha-1, which generally align with our study's findings, that showed mean carbon densities ranging between 7.30 to 124.05 tC .ha-1 (Glenday, 2007). In contrast, New York's natural urban forested areas showed a higher mean carbon stock of 263.04 Mg C. ha-1, as reported by Pregitzer et al. (2022) conducted an estimation of urban forest aboveground carbon using machine learning in China. Their findings indicated a slight increase from 24.90 Mg. ha-1 in 2016 to 25.61 Mg. ha-1 in 2019, which is consistent with the results of our study, that showed an average of 32.05 t. ha-1 over a seven-year period (2012-2019).

5.6.3 Spatio-temporal variations in carbon stocks

The objective of this study was to analyze changes in carbon stocks in reforested urban landscapes over time using a multi-source dataset comprising Rapid-Eye spectral data, land-use data, and topographic and bio-climatic data. The results demonstrate that these covariates are sensitive to tree morphological, physiological, and phenological properties, which improve the accuracy of aboveground carbon stock prediction and distribution over time. Spatial and temporal variations in carbon stocks were observed to be influenced by stand age, composition, soil properties, terrain, and climate. The study site showed lower carbon density in reforested areas in 2012 due to the recent initiation of the reforestation program, while the northwestern and southwestern sections of the site exhibited higher carbon stocks due to the presence of a pioneer community. A decline in carbon stocks was observed between 2013 and 2015, following a fire occurrence and drought, but the ecosystem gradually recovered to reach its peak in 2019. Carbon accumulation and storage were found to be influenced by climatic, land-use, and topographic conditions, and the spatial distribution of carbon stocks coincided with the site's land-use characteristics and terrain. From 2018 to 2019, an overall high carbon accumulation and storage incidence was observed, with the northwestern and northeastern sections having the highest carbon stocks. It is worth noting that the changes in forest carbon

stock are closely related to variations in green-biomass density and canopy cover. The decline in carbon stock during 2014 and 2016 underscores the dynamic nature of carbon storage, which is influenced by changes in green-biomass density and canopy cover.

Continued use of RapidEye imagery: The RapidEye satellite constellation has proven to be a valuable data source for our analyses, offering a good balance of spatial resolution, spectral bands, and coverage. Given its effectiveness, I recommend continuing to utilize RapidEye imagery for future studies in similar contexts or applications.

1. Integration of drone-based imagery: As you rightly pointed out, drone-based imagery can provide even higher spatial resolutions, potentially up to 20 cm depending on the flight altitude. Such high-resolution data could be particularly useful for applications that require fine-scale detail, such as precision agriculture, urban mapping, or monitoring of small-scale features. I recommend exploring the integration of drone-based imagery, either as a standalone data source or in combination with RapidEye data, to potentially increase the accuracy and level of detail in our models and analyses.
2. Multi-sensor data fusion: Building upon the previous recommendation, I propose investigating the fusion of data from multiple sources, including RapidEye, drone-based imagery, and potentially other satellite sensors (e.g., Sentinel, Planet, etc.). By combining the strengths of different data sources, we may be able to enhance the accuracy, robustness, and comprehensiveness of our analyses, leveraging the complementary information provided by each data source.
3. Improvement of processing techniques: In addition to exploring new data sources, I recommend continuously evaluating and improving our data processing techniques, such as image preprocessing, feature extraction, and machine learning algorithms. This could involve implementing advanced techniques like deep learning, ensemble methods, or domain-specific algorithms to better handle the complexities of the data and improve the overall performance and accuracy of our models.
4. Collaboration and knowledge sharing: Finally, I strongly recommend fostering collaborations with other research groups, industry partners, and stakeholders working in related fields. Such collaborations can facilitate knowledge sharing, provide access to diverse datasets and resources, and potentially lead to interdisciplinary insights and innovative solutions.

The study concludes that carbon accumulation capacity varies with forest age and species composition, with younger forests still growing and yet to reach maximum carbon accumulation capacity.

5.7 Conclusions

The integration of remote sensing and machine learning can provide rapid cost-effective and reliable information for the spatiotemporal assessment, planning, and management of complex natural resources, including forests. This study estimated annual forest cover and carbon stock changes in a reforested urban landscape, with findings showing an average biomass of 36.85 t. ha-1 and an average forest cover change of 18.9 ha from 2012 to 2019. For a reforested urban landscape, the study estimated annual forest cover and carbon stock changes and found an average carbon of 36.85 t. ha-1 and 18.9 ha average in forest cover change. Based on our assessment, forest areas decreased by 28 ha between 2014 and 2016, attributed to fire outbreaks and drought. However, a total net gain of 7192t of carbon was observed during the study period (2012–2019) attributed to mass awareness and tree planting under the reforestation programmes/projects. Overall, the collaborative efforts towards ecosystem restoration and urban reforestation have proved successful in sequestering carbon, mitigating climate change, and improving the health of the urban ecosystem. We recommend that the local eThekwini Municipality incorporate carbon stock information into future forest inventories using remote sensing data archives, such as Rapid-Eye, and machine learning algorithms, such as extreme gradient boosting. This information would be valuable for planning, implementation, and monitoring reforestation projects that aim to protect and conserve forests, enhance forest carbon stocks, reduce emissions and increase carbon sequestration. Additionally, we emphasize the importance of promoting new reforestation sites as well as social forestry projects and reducing communities' reliance on forests through the provision of alternative wood products for construction materials and fuelwood.

Acknowledgments: This study was supported by the SARChi Chair in Landuse Planning and Management (Grant No. 84157). A special acknowledgement to the Durban Municipality through the Woodrights Project for allowing the study to be conducted at the Buffelsdraai Reforestation Site. Furthermore, a special thanks to Planet Labs Inc for allowing us to access their data. Appreciation goes out to the reviewers for their constructive criticism and comments that greatly improved the manuscript.

5.8 Summary

This chapter quantified the impacts of extensive reforestation initiatives on aboveground carbon sequestration and storage across the Buffelsdraai landscape overtime, revealing gradual increases over 2012-2019 with fluctuations linked to tree mortality events. Findings highlighted the significant role of strategic urban greening in strengthening local carbon sinks and climate regulation functionality. However, the future trajectories of these emergent stocks under accelerating land development and climate change remains unspecified.

Therefore, leveraging the historical quantification foundations from previous work, the next phase involves forecasting potential shifts in reforestation carbon under varying scenarios. Specifically, Chapter 6 develops an artificial neural network model integrating past Rapideye imagery and current Sentinel-2 imagery with CMIP6 climate projections and shared socioeconomic pathways to simulate changing carbon densities up to 2040. Comparative assessment of gains and losses under diverging global trajectories equips municipalities to better account for, enhance and protect critical green infrastructure in alignment with carbon neutrality commitments. Overall, this work stretches remote sensing and machine learning capabilities for dynamic assessment and adaptive evidence-based planning of resilient urban forests realizing mitigation ambitions.

6 CHAPTER SIX: Mapping aboveground carbon stocks under past, present and future climate scenarios in urban reforested Landscapes.

This chapter is based on:

Collins Matiza[1*], Onisimo Mutanga[1], Kabir Peerbhay [1], and Romano Lottering[1.] 2024. Mapping aboveground carbon stocks under past, present and future climate scenarios in urban reforested Landscapes. *Beyond the Surface: Revealing Ecosystem Services through Remote Sensing: Chapter 9.*

Abstract

Understanding the role of reforestation over time is essential for preserving forest ecosystem health, maintaining carbon cycle balance, and addressing climate change. However, little is known about the contribution of reforestation to aboveground carbon (AGC) accumulation and ultimately, its contribution to climate change mitigation in urban landscapes. This limitation hinders the ability to develop effective strategies for managing urban forests in order to maximise their carbon sequestration potential. Therefore, this study aims to understand the past, present and future aboveground carbon stocks in reforested urban landscapes using forest inventory observation data, combined with land use, topographic and global climate circulation model data (CMIP6). An artificial neural network regression model was selected to construct the spatial relationship between AGC and the ancillary datasets. Analyzing the spatio-temporal dynamics of aboveground carbon (AGC) from 2012 to 2040, this study employed a resolution of 1 km and a 20-year interval, focusing on three distinct climate scenarios outlined in CMIP6. The cross-validation outcomes substantiate the reliability of AGC predictions in reforested urban landscapes, albeit acknowledging uncertainties in specific study site sections. The mean AGC values recorded over the past three decades were observed at 55.94 and 84.02 t. ha^{-1}. In contrast to historical AGC (2012-2022), future AGC exhibited notable fluctuations under diverse climate scenarios. Specifically, AGC showcased a gradual increasing trend under SSP1-2.6 low emissions, a diminishing trajectory under SSP2-4.5, and a markedly rapid decline under SSP5-8.5. This work enhances the management of urban forest carbon dynamics to support informed climate change mitigation policies and sustainability strategies.

Keywords: Aboveground carbon, reforestation, climate change, artificial neural networks, spatio-temporal evolution

6.1 Introduction

Over the past two centuries, rapid urbanization and human activity have led to global deforestation, resulting in a significant loss of forest cover, emerging as a major driver of environmental issues like landscape fragmentation, pollution, climate change, biodiversity loss, and thermal stress (He et al., 2014, Laurance, 2010, Birdsey and Pan, 2015). Urban landscapes account for the majority of global energy consumption and CO_2 emissions, figures projected to rise further by 2030 (United Nations, 2023, NRDC, 2018, Yue et al., 2023, IEA, 2008). Tropical forest loss has also accelerated, reducing carbon storage (Cai and Zhang, 2014). Given these challenges, urban reforestation has been proposed as a sustainable climate change mitigation solution by increasing carbon capture through strategies like planting trees and vegetation in cities (Ogunbode and Asifat, 2021). Despite the challenges of urbanisation, urban reforestation (UR) provides a sustainable solution through carbon sequestration and storage (Teo et al., 2021, Nave et al., 2019, Anjali et al., 2020). Therefore, assessing the dynamics of AGC stocks over time is critical to quantify the effectiveness of urban greening efforts and guide future investments. Furthermore, projecting climate and land use change influences on future reforested carbon also allows cities to prepare and prioritize interventions that maximize carbon gains.

Climate change and land use change are expected to significantly influence aboveground carbon (AGC) stocks and fluxes in urban reforestation projects. Changing precipitation, temperature, and extreme weather events alter AGC dynamics (Feng et al., 2022). Climatic factors like warming temperatures can suppress plant carbon uptake while accelerating losses through enhanced respiration (Zavaleta et al., 2003, Wan et al., 2005).Topography also modifies carbon density and spatial patterns (Raich et al., 2006). However, land use change poses the greatest threat, contributing substantially to global CO_2 emissions (Hong et al., 2021). Over 1.38 peta-grams of potential AGC loss is projected across the tropics by 2030 due to land use changes (Seto et al., 2012). While reforestation offers climate mitigation potential, substantial knowledge gaps persist regarding precise carbon stock quantification and projection under environmental change (Mitchell et al., 2018b, Rodríguez-Veiga et al., 2017).

The advent of remote sensing has significantly advanced carbon mapping, by integrating field data with quantitative methods to map ecosystems and predict tree properties (Xiao et al., 2019, Odebiri et al., 2023, Kowe et al., 2021). Leveraging high resolution satellite imagery, derived vegetation indices and machine learning, enables insights into carbon accumulation patterns (Matiza et al., 2023b).. For instance, Eckert (2012) utilized WorldView-2 satellite imagery to estimate aboveground biomass density and carbon stocks across landscapes in northeastern Madagascar. Recently Odebiri et al. (2022) combined Sentinel data, climate projections General Circulation Models (GCM) (i.e. CNRM-CM61-1, CanESM5, GFDL-ESM4, ACCESS-ESM1-5, and INM-CM5-0), and soil datasets to model future soil organic carbon across South Africa under land use and climate change. As remote sensing data quality improves, analytical innovations are required to fully exploit the rich details provided.

While statistical models have proven useful for aboveground carbon (AGC) mapping, their predefined forms limit flexibility in capturing intricate multivariate relationships in heterogeneous datasets (Kanevski et al., 2008, Matiza et al., 2023b, Koldasbayeva et al., 2023). Machine learning provides distinct advantages, with techniques such as neural networks identifying inherent nonlinear patterns without assumptions, enhancing predictions (López-Serrano et al., 2020, Wang et al., 2022d). Further, complex AGC variability can overwhelm simplistic models. In contrast, ML algorithms such as artificial neural networks (ANNs) excel at extracting robust features invariant to noise through hyperparameter tuning (Pannakkong et al., 2022, Abiodun et al., 2019). This enables more accurate learning from large, noisy datasets. Despite these benefits, applying advanced ML techniques for remote sensing-based AGC simulations remains underexplored, presenting immense research potential.

Therefore, this research leverages remote sensing technology, topo-climatic data and machine learning, specifically artificial neural networks, to predict and analyze dynamics of reforested aboveground carbon (AGC) stocks and distributions in urban areas. The study utilizes a robust multi-source dataset to quantify AGC variability across three key time periods – historical (2012), current (2022) and projected (2040). This temporal analysis of urban forest carbon enables assessment of AGC variability considering evolving land use and climate. Integrating both remotely sensed data, topo-climatic data and machine learning facilitates significant exploration of urban reforestation carbon patterns. Ultimately, these insights can

inform sustainable strategies and policies to harness urban forests for enhanced climate change mitigation at local and global scales.

6.2 Methods and Materials

6.2.1 Observed reforested aboveground carbon stock data collection.

A random field sampling strategy was implemented in the months of February and May 2012 and 2022 across 97 distributed sites spanning the Buffelsdraai reforestation site for data collection. This approach enabled consistent spatial coverage and comparison across the study period. Precise sampling locations were randomly predefined using ArcGIS Pro. At each plot, two 3m x 3m quadrats were delineated and situated within larger 10m x 10m plots to facilitate multi-scale data collection. These boundary dimensions were chosen based on recommended guidelines for balancing sampling effort with representation of species-level variability (Heiskanen et al., 2013). Every tree within each quadrat was surveyed, recording species, height using a Leica clinometer (accuracy ±0.5m @ 100m), and diameter at breast height (dbh at 1.3m) using a tape measure. Additionally, a Trimble global positioning system GPS receiver (accuracy <50 cm) recorded coordinates. Concurrently, land slope, aspect, and land cover were recorded given their known influence on biomass accumulation (Tang et al., 2016). A total of 97 plots measuring 10m by 10m with a minimum of 2 samples per plot giving a total of 194 AGB samples was used.

The collected field measurements served to estimate standing aboveground tree biomass by applying allometric models that mathematically convert structural attributes into biomass estimates. The IPCC (2006) endorses these allometric equations as they enable indirect determination of forest carbon stocks. An extensive review by (Djomo et al., 2016) of allometric models across Africa concluded that generalized species-specific equations provide robust regional estimates in heterogeneous forests. Hence, measured tree height (H) and diameter (D) data were integrated into the pantropical model Eq. 6.1to estimate plot-level biomass (Chave et al., 2014).

$$AGB = 0.0673 \times \left(p^{D^2H}\right)^{0.976} \quad \text{Eq. 6.1}$$

Where AGB represents aboveground oven-dry biomass (kg), D is diameter-at-breast height (cm), H is tree height (m), and *p* represents wood density constant (1.2 g/cm^3). Lastly, aboveground biomass values were converted to carbon stock by applying a 0.5 conversion factor, consistent with forest inventory guidelines 0.5 was used to compute the actual carbon stock from above ground biomass (Gara et al., 2016, Solomon et al., 2017, MacDicken, 1997).

6.3 Data acquisition

6.3.1 Multispectral Image collection and preprocessing

Historical (2012) and current (2022) aboveground carbon (AGC) estimation models were first developed to map the distribution and density of AGC stocks across the reforestation. The analysis integrated multispectral imagery from two sources - Planet Labs' RapidEye and the European Space Agency's (Fassnacht et al.) Sentinel-2 platforms. RapidEye and Sentinel-2 were selected as the best available data sources for the 2012 and 2022 time points. Cloud-free imagery (<10% cover) was acquired for both sensors during peak productivity months (February to May) in 2012 and 2022. Specifically, the study utilized Sentinel-2 Level 1-C top-of-atmosphere reflectance data, retrieved from ESA's SciHub portal (https://scihub.copernicus.eu). Sentinel-2 provides 10–60-meter resolution imagery across 13 spectral bands (443 - 2190 nm), with a 5-day revisit cycle enabling regular capture (Languille et al., 2015, Mngadi et al., 2021a). Preprocessing via ESA's SNAP software included geometric, radiometric and atmospheric corrections. Images were then resampled to a 10-meter spatial resolution (Mngadi et al., 2022b). Vegetation indices which combine Sentinel-2 bands to monitor parameters like biomass, leaf area, and chlorophyll content -are all proxies for AGC stocks (Dube and Mutanga, 2016, Gara et al., 2016). Consequently, using information derived from Mngadi et al. (2021a) and Suardana et al. (2023), vegetation indices including normalized difference vegetation index (NDVI), green normalized vegetation index (GNDVI), enhanced vegetation index (EVI), red-edge normalized vegetation index (RENDVI) and chlorophyll index (red-edge) together with raw sentinel wavebands were used to construct the current AGC estimation 2022 model.

Similarly, Planet Labs' (https://www.planet.com) RapidEye images offer a 5-meter pixel size across five bands (440 - 850 nm), with a daily revisit time. Level 3A products retrieved incorporate radiometric, geometric, and sensor corrections (Anderson et al., 2011). This data

covers an orthorectified area of 25 x 25 km per scene (Jung-Rothenhäusler et al., 2007). As with Sentinel-2 data similar vegetation indices were derived from the RapidEye bands to help explain observed AGC variability, as demonstrated in other studies (Gara et al., 2016, Mngadi et al., 2022a). Similarly, vegetation indices and Rapideye wavebands were integrated to form a model to estimate the past AGC for the year 2012. All remote sensing datasets were clipped and resampled to a standardized 10-meter resolution.

6.3.2 Topographic variables

Topographic covariates were derived from a regional Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) obtained from the NASA-USGS portal (Cowan and Cooper, 2004). Using SAGA GIS software v8.2.3 (https://saga-gis.org), multiple factors were processed: slope, aspect, wetness index, curvature, roughness and elevation. The topographic data were resampled to 10 m for standardized analysis with other input predictors.

6.3.3 Bioclimatic variables and future climate scenarios

Gridded historical and future climate data were obtained from the WorldClim portal (WorldClim) at approximately 1 square km resolution (Fick and Hijmans, 2017). Variables include monthly precipitation, temperatures, and derived bioclimatic factors representing annual/seasonal trends. Bioclimatic descriptors condense complex climatic factors (e.g. seasonality) into localized metrics useful for ecological modeling. Future climate projections were accessed from three CMIP6 general circulation models (GCMs) - ACCESS-ESM1-5, MIROC6, and EC-EARTH2 - under three different greenhouse gas emissions scenarios. Using multiple climate models captures a range of sensitivity and better constrains uncertainty (Wang et al., 2022a). Representative concentration pathway shared socio-economic scenarios (RCP-SSP) were selected to represent low (SSP1-26), medium (SSP2-4.5), and high (SSP5-8.5) emission trajectories (Gao et al., 2013). These three RCPs were chosen to capture a range of plausible futures. The raw climate model outputs for these scenarios were then statistically downscaled using statistical downscaling model (SDSM) to generate higher resolution projections better aligned with the spatial and temporal scale of the other predictors in this case study.

6.3.4 Projected landcover

This analysis integrated multi-temporal LULC maps obtained from the ESRI landcover classifications based on Sentinel 2 archive (2017-2021) encompassing 12 cover class types at 10m resolution (>70% accuracy) (Venter et al., 2022). The South African Landcover map (accessed through the specified portal (https://egis.environment.gov.za/gis_data_downloads/) for 2012 served as the basis for the 2012 land cover classification (SANLC, 2020). Maps were reclassified into three major categories based on vegetation descriptions.

An artificial neural network approach within the MOLUSCE QGIS plugin was implemented to project LULC forward based on historical changes and relationships with slope, aspect and other spatial factors. Projected maps for 2022 and 2040 were generated and validated against 2021 reference data, achieving 81% correct match and 0.774 kappa statistic. All LULC and environmental data outputs were resampled to a 60 m grid for consistent integration.Table 6.1summarizes all the input predictor variables used in the models.

Table 6.1 Covariates used for past (2012), current (2021) and future (2040) AGC simulation.

Predictor variable	**Description/ formulae**	**Source**
Sentinel-2 bands 2, 3,4,5,6,7,8,8A, 11, and 12	Wavebands of sentinel ranging from 490nm-2190nm	(Drusch et al., 2012)
RapidEye bands 1,2,3,4, and 5	Waveband s of RapidEye ranging from 440-850nm	(Stoll et al., 2012)
Normalized difference vegetation index (NDVI)	$\frac{NIR - RED}{NIR + RED}$	(Pettorelli, 2013)
Green normalized difference vegetation index (GNDVI)	$\frac{NIR - Green}{NIR + Green}$	(Jiang et al., 2007)
Enhanced vegetation index (EVI)	$2.5 \times \frac{NIR - RED}{(\text{NIR} + 6 \times \text{RED} - 7.5 \times \text{BLUE} + 1)}$	(Jiang et al., 2007)
Red-edge normal difference vegetation index (RENDVI)	$\frac{NIR - RedEdge}{NIR + RedEdge}$	(Evangelides and Nobajas, 2020)
Chlorophyll index red-edge (CIRE) Sentinel-2A	$\frac{Red\ edge(763nm)}{Red\ edge\ (705nm)}^{(-1)}$	(Evangelides and Nobajas, 2020)

Chlorophyll index red-edge (CIRE) RapidEye	$\frac{Near\ Infrared}{Red\ edge}^{(-1)}$	(Clevers and Gitelson, 2012)
Elevation	Height above sea level in meters	(Jacobsen, 2005)
Topographic Wetness index (TWI)	Relative wetness due to terrain factors within catchments a measure of position on the slope.	(Languille et al., 2015)
Surface Roughness	the irregularities and variations in the elevation of a land surface.	(Lang et al., 2013)
Aspect	Slope direction approximation of the amount of radiation received by a site.	(Lang et al., 2013)
Slope	The gradient of a specific site	(Lang et al., 2013)
Landcover classification map 2012	South Africa National Land use Land cover map (2012)	(SANLC, 2020)
Landcover classification map 2022	Projected Land use Land cover map (2022)	(Venter et al., 2022)
Landcover classification map 2040	Projected Land use Land cover map (2040)	
Bioclimatic variable 12	Mean Annual Precipitation	(Fick and Hijmans, 2017)
Bioclimatic variable 1	Mean annual temperature	(Fick and Hijmans, 2017)
Bioclimatic variable 4		(Fick and Hijmans, 2017)
Bioclimatic variable 5	Maximum temperature of the warmest month	(Fick and Hijmans, 2017)
Bioclimatic variable 13	Precipitation of wettest month	(Fick and Hijmans, 2017)
Average of 3 general circulation models (GCM)	Mean annual precipitation/ temperature	(McBride et al., 2021)

6.4 Spatial Prediction model

Artificial neural networks (ANNs) provide flexible machine learning frameworks adept at capturing complex multivariate relationships. Their multi-layer structure enables both linear

and non-linear feature transformations, making ANNs well-suited for the intricate connections between aboveground carbon stocks, climate factors, topography, land use, vegetation indices and other covariates (Pannakkong et al., 2022, Abiodun et al., 2019). The core architecture of an Artificial Neural Network (ANN) model commonly utilizes a Multi-Layer Perceptron (MLP) structure. This structure features an input layer, one or more hidden layers, and an output layer. Interconnected neurons within each layer transmit information unidirectionally from input to output, enabling the model to capture both linear and non-linear relationships between input and output variables (Abiodun et al., 2019, Ghorbanian et al., 2022). Overfitting, a concern across modeling approaches, also affects neural networks with inaccurate predictions on new data. Mitigation requires careful hyperparameter tuning and training techniques.

The initial 2012 ANN model was used to predict 2022 carbon stocks without incorporation of actual 2022 field data. The 2022 predictions were validated against in situ measurements, serving to evaluate model accuracy. Adjustments were made based on this validation to improve performance. With the model refined and validated on 2022 data, it was then leveraged in a two-step process to predict 2040 carbon stocks. In stage 1, the ANN utilizes embedded regression relationships to estimate 2040 carbon levels based on climate and land use projections. Stage 2 further refines these projections through additional training on error residuals.

The same model architecture and covariates were maintained from 2012 to 2040 projections, with only the climate, terrain, and land cover input data updated to represent projected 2040 conditions. Three shared socioeconomic pathway (SSP) scenarios were simulated to evaluate a range of outcomes. The ANN flexibility enables re-training if additional future data becomes available, further improving accuracy. Validating the 2022 predictions demonstrates the model's capability for projecting future carbon stocks.

The available data was split into training and testing datasets using a stratified random sampling approach. This ensures that both datasets maintain a similar distribution of aboveground carbon stock values across different forest types, climate zones, and topographic regions. A common split ratio could be 80% for training and 20% for testing. Model Selection Criteria: Several ANN architectures with varying numbers of hidden layers and neurons were explored during the model development phase. The models were evaluated based on

performance metrics such as root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R-squared) on the testing dataset. The model with the optimal balance between performance and complexity was selected, considering factors like overfitting and computational efficiency.

Cross-Validation Techniques: To further assess the model's generalization ability and prevent overfitting, cross-validation techniques were employed. A common approach is K-fold cross-validation, where the training data is divided into K subsets or folds. The model is trained K times, using K-1 folds for training and the remaining fold for validation in each iteration. The performance metrics are then averaged across all K iterations, providing a robust estimate of the model's performance on unseen data. Additionally, techniques like early stopping and dropout regularization could be used to mitigate overfitting during the training process. Early stopping involves monitoring the model's performance on a validation set and halting the training process when performance stops improving, preventing further overfitting. Dropout regularization randomly drops out or ignores a fraction of neurons during training, encouraging the network to learn more robust and redundant representations.

6.5 Model evaluation

We evaluated model performance using complementary goodness-of-fit metrics, widely used in remote sensing studies (Odebiri et al., 2022, Matiza et al., 2023b). The coefficient of determination (R^2) was leveraged to quantify the proportion of variance explained by the model. Additionally, we considered the mean absolute error (MAE) as a crucial indicator of the average magnitude of discrepancies between observed and predicted values. Lin's concordance correlation coefficient (LCCC) provided insight into both precision and accuracy, offering a comprehensive assessment of agreement. Furthermore, the root mean square error (RMSE) was employed to gauge the square root of the mean of squared errors, providing a measure of the model's overall accuracy. A higher R^2 and LCCC along with lower MAE and RMSE indicate better model fit (Morley et al., 2018).

To evaluate model performance, we adopted a rigorous validation approach. The ANN model was developed on a training set comprising 70% of the 2012 data (135 samples). The remaining 30% (59 samples) were reserved for testing model predictions. We developed an

initial model on just the 2012 data, then projected the 2022 carbon stocks and validated its predictions on the actual 2022 data not used for training. Comparing these 2022 predictions to actual observations would assess generalization ability before finalizing the model. This phased validation strategy would further verify accurate generalization to new time periods, enhancing reliability evaluation, especially for long-term forecasts. Our use of training/testing splits across multiple years, and prospective testing on future data provides robust validation of model performance over time.

SHapley Additive exPlanations (SHAP) is a model interpretation method grounded in cooperative game theory that explains the output of machine learning models by quantifying the contribution of each feature to an individual prediction (Liben-Nowell et al., 2012). The algorithm is derived from Shapley values, a solution concept in coalitional game theory that distributes payouts among players according to their marginal contribution. In the SHAP methodology, each feature is conceptualized as a "player" in the model (Liben-Nowell et al., 2012). The Shapley value φi for feature is computed by comparing the model output to the expected output over all possible feature coalitions. The SHAP method has also been shown to enhance model transparency and interpretability in remote sensing applications (Temenos et al., 2023). Huang et al. (2022a) applied SHAP to a random forest model for grassland aboveground biomass prediction in China, identifying precipitation, temperature, and slope as key predictors. Due to its documented strengths in providing granular insights into feature importance and individual predictions, the SHAP method was chosen as the preferred tool for interpreting the model's performance (Movsessian et al., 2021).

6.6 Results

6.6.1 Descriptive statistics of AGC stock measurements

The descriptive analysis of the data across (Table 6.2) various periods indicates that observed AGC exhibited a moderate range and variation throughout the study period. The distributions were found to be asymmetrical, with skewness values reaching 1.05. In terms of variance homogeneity, Levene's test produced p-values below 0.05 for all pairwise comparisons of periods, suggesting statistically significant differences between samples across different time frames.

Table 6.2 Descriptive statistics of measured aboveground carbon stock (in t. ha^{-1})

Year	Mean	Standard deviation	Minimum	Maximum
2012	57.82	26.75	1.05	116.33
2022	87.45	32.09	17.43	177.84

6.7 Model performance

Table 6.3presents validation results for the artificial neural network (ANN) model developed to estimate aboveground carbon (AGC) stock changes over time. The ANN regression approach explained 78% of the observed variation in AGC within the validation dataset, assessed using a 10-fold cross-validation. Accuracy was evaluated using root mean square error (RMSE), mean absolute error (MAE), and the Lin Concordance Correlation Coefficient (LCCC). Analysis of both spatial and temporal prediction accuracy revealed 2012 had the lowest precision (R^2 0.68, RMSE 27.55 t. ha^{-1}, MAE 21.70 t. ha^{1} and LCCC 0.78) based on these metrics, with accuracy improving in 2022 (R^2 0.80, RMSE 25.68 t. ha^{-1}, MAE 22.16 t. ha^{1} and LCCC 0.83). The model achieved an overall mean carbon stock of 55.94 t. ha^{-1}, slightly lower than the reference data's mean carbon stock of 57.75 t. ha^{-1}, indicating acceptable AGC predictions. Assessment by year shows the model captured interannual AGC fluctuations. The relationship between observed and predicted stocks **(**Figure 6.1) strengthened over time as biomass accumulated within the reforestation area as supported by the model's 2022 mean carbon stock of 87.45 t. ha^{-1} for the reference data and 84.02 t. ha^{-1} achieved by the model. By robustly validating both the spatial and temporal performance of the ANN approach, this study demonstrates reliable landscape-scale mapping of carbon stock accumulation over decades within the regenerating forest.

Table 6.3 Model validation results based on the 10-fold cross validation for ANN regression model.

Year	R^2	RMSE (t. ha^{-1})	MAE (t. ha^{-1})	LCCC	Mean Carbon (t. ha^{-1})
2012	0.68	27.5546	21.70	0.78	55.94
2022	0.80	25.6899	22.1698	0.83	84.02

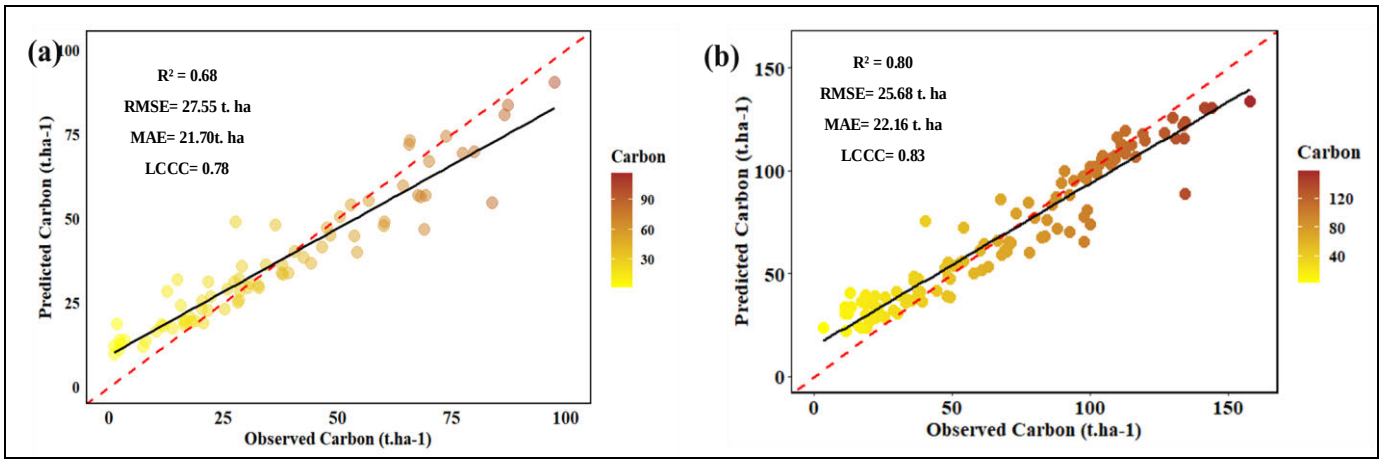


Figure 6.1 Correlation between the observed and predicted aboveground carbon stocks (a) validation scatterplot for 2012 (b) validation plot for 2022.

The SHAPley method identified 10 key predictors with high importance (SHAP value ≥ 3.0) for modeling aboveground carbon (AGC) in urban reforested landscapes. Precipitation was the most influential variable, followed by NDVI, band 8, land use type, temperature, EVI, RENDVI, elevation, topographic wetness index (TWI), and band 4 (Figure 6.2). Notably, while elevation and TWI ranked near the bottom of the 10 most important variables at the plot level, together with other topographic factors like slope and aspect, they still contributed significantly to landscape-level AGC stock predictions. This suggests complex interactions between climate, vegetation and terrain in determining carbon storage across this heterogeneous urban ecosystems (Timilsina et al., 2014).

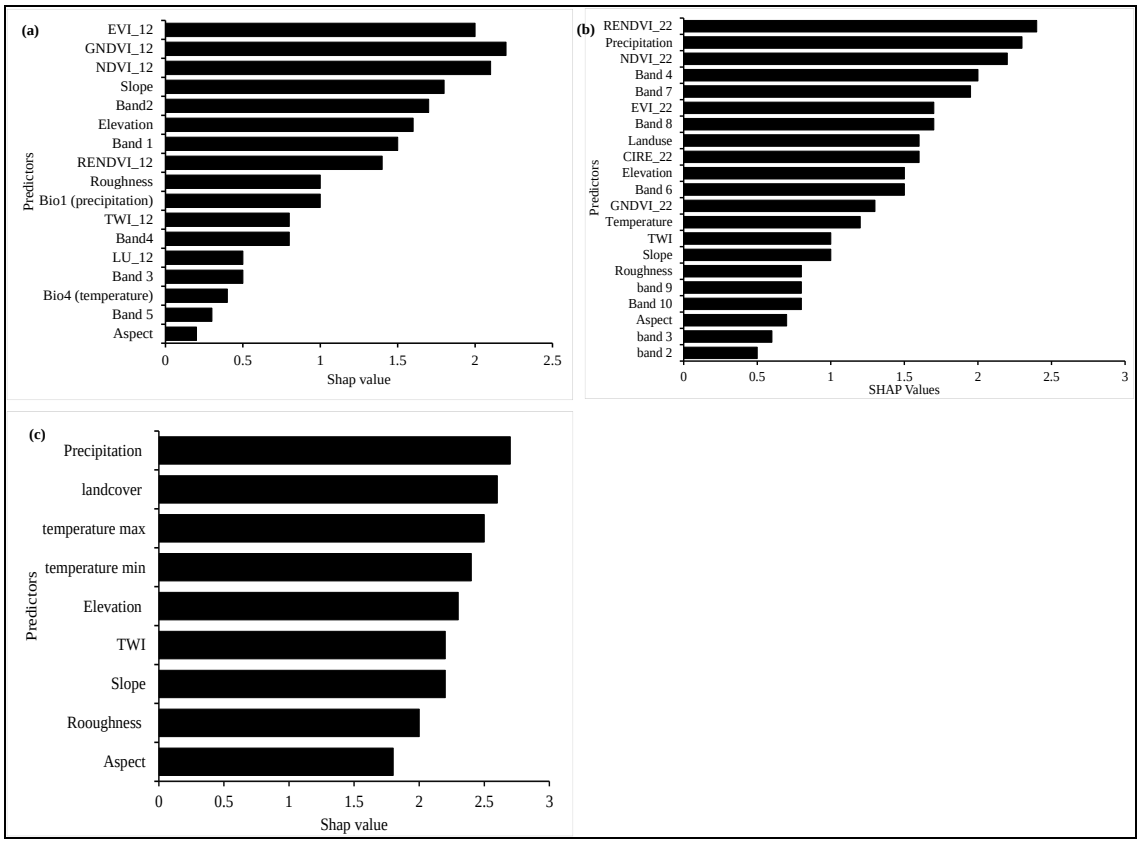


Figure 6.2 Shapley Importance ranking of predictors used for the estimation and simulation of past (a), current (b), and projected (c) future aboveground carbon stock.

6.8 Mapping of AGC stocks

The landscape-level maps of AGC stocks were generated using 10-fold cross-validation for each time interval in Python 3.9.1. Spatially, the maps revealed an increase in mean AGC stocks from 55.94 t. ha^{-1} in 2012 to 84.02 t. ha^{-1} in 2022. Additionally, the total AGC stocks were calculated. Notably, low AGC values were concentrated in the central sections of the study area, corresponding to the current land use (2022 map) of a landfill. The lack of vegetation in the central section explains the low aboveground carbon reserves (Figure 6.3). The maps show a gradual increase in AGC from 2012 to 2022 with an approximate total increase of 13 557t of carbon over the last decade. This is evident on the generated maps with a gradual increase in the northern, southern and western sections of the map. Aboveground carbon density also significantly rose over time.

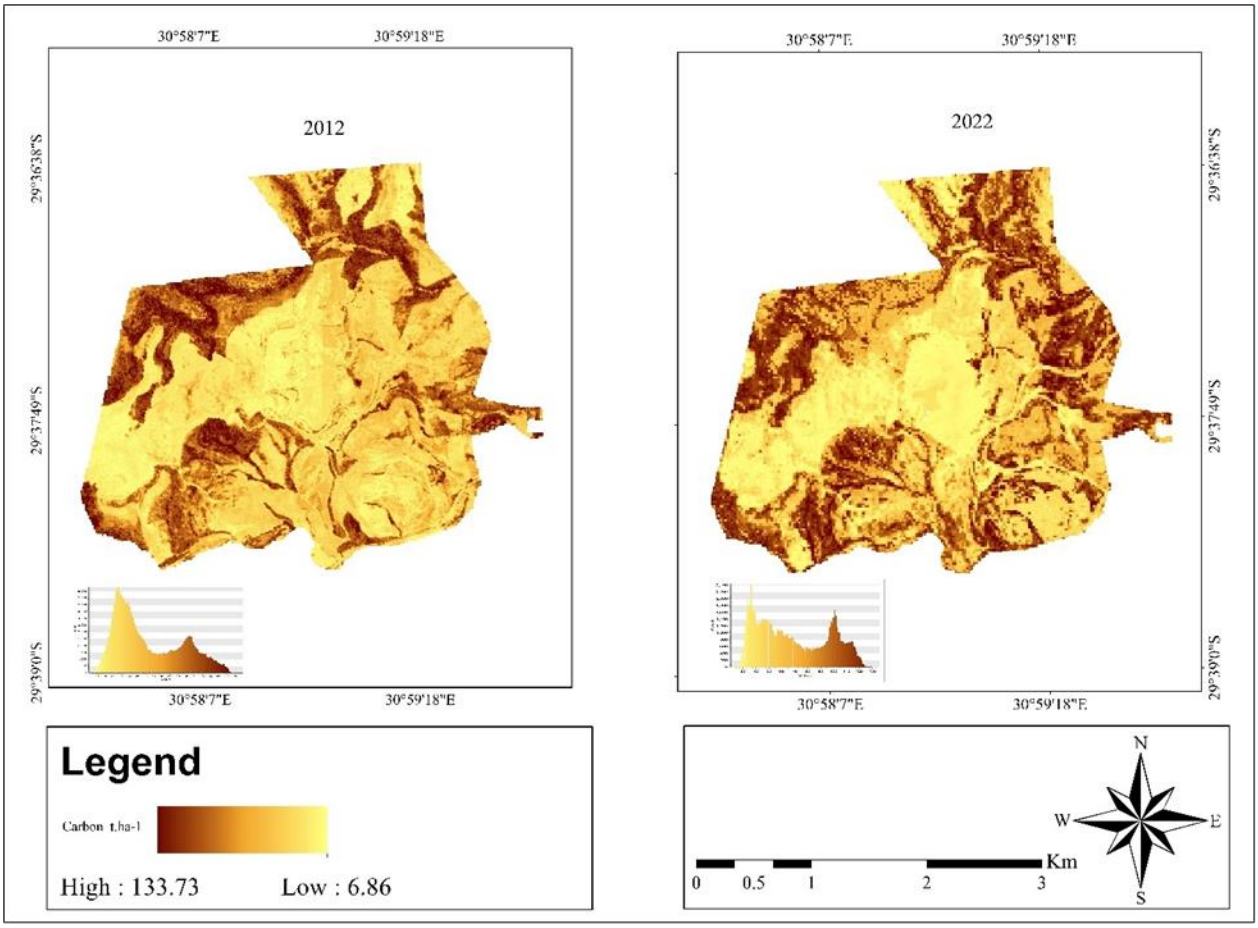


Figure 6.3 Spatial distribution of past (2012) and current (2022) *AGC* stocks estimated using ANN.

The predicted future AGC stocks differed amongst the three shared socio-economic pathways including SSP 126, SSP245. And SSP585. The lowest emission shared socio-economic pathway (SSP126) produced the highest accuracy with the following metrics (R^2= 0.63, RMSE= 27.44 t. ha^{-1}, MAE= 25.54 t. ha^{-1} and LCCC =0.75). SSP 245 (medium to high emissions) followed with metrics of (R^2= 0.61, RMSE= 28.44 t. ha^{-1}, MAE= 26.54t. ha^{-1} and LCCC =0.74) and lastly SSP 585 (high emissions) with the following metrics (R^2= 0.60, RMSE= 29.23 t. ha^{-1}, MAE= 27.54 t. ha^{-1} and LCCC =0.70). Table 6.4 shows the decrease of AGC stocks across the different SSPs, with an approximate total aboveground carbon stocks estimated at 39206 t (SSP126), 26400 t (SSP245), and 25200 t (SSP585) respectively. A notable diminishing trend in AGC is evident across the different SSP scenarios, with the lowest accumulation of AGC under SSP585, which coincides with the high emissions

scenario.

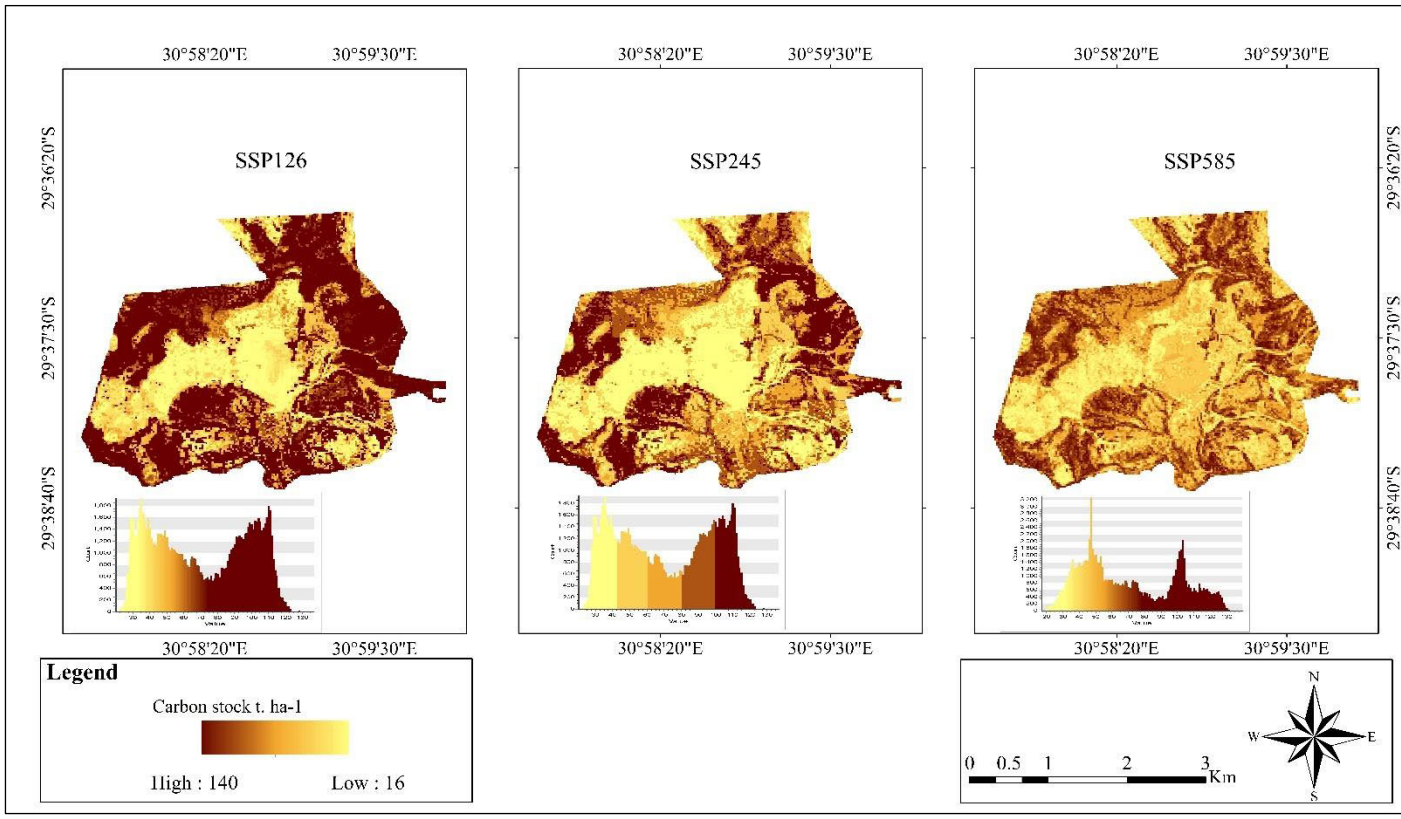


Figure 6.4 maps this diminishing spatial pattern in stocks over time as climate change intensifies, especially in northern and southern sections. The central landfill area persists as lowest AGC due to lack of vegetation limiting carbon accumulation. CMIP6 climate projections of rising temperatures and declining precipitation, driving extreme weather (Table 6.4), help explain modelled landscape AGC losses. As Figure 6.2 shows, climate factors exhibited the strongest influence, followed by land use change and terrain, on predicted future carbon stocks.

Table 6.4 Changes between past (2012), current (2022) and future (2040) aboveground carbon stocks expressed in tonnes the numbers in parenthesis show gain or losses in carbon stocks across the two landcover classes. The negative sign represents a loss.

Land use type	**Past Carbon Stock (t) (2012)**	**Current Carbon Stock (t) (2022)**	**Future Carbon Stock (t) (2040) SSP126**	**Future Carbon Stock (t) (2040) SSP245**	**Future Carbon Stock (t) (2040) SSP585**
Forest cover	9290	19257	31 956	20400 (-900)	19100 (-1300)

Non-forest	3200	6700	7340	6000 (-1340)	6100 (-100)
Total	**12 400**	**25957**	**39206**	**26400**	**25200**

Most of the predictors significantly influenced projected future Aboveground Carbon (AGC) stocks, with climate models exhibiting the strongest impact, closely followed by land-use change and terrain variables (Figure 6.2).

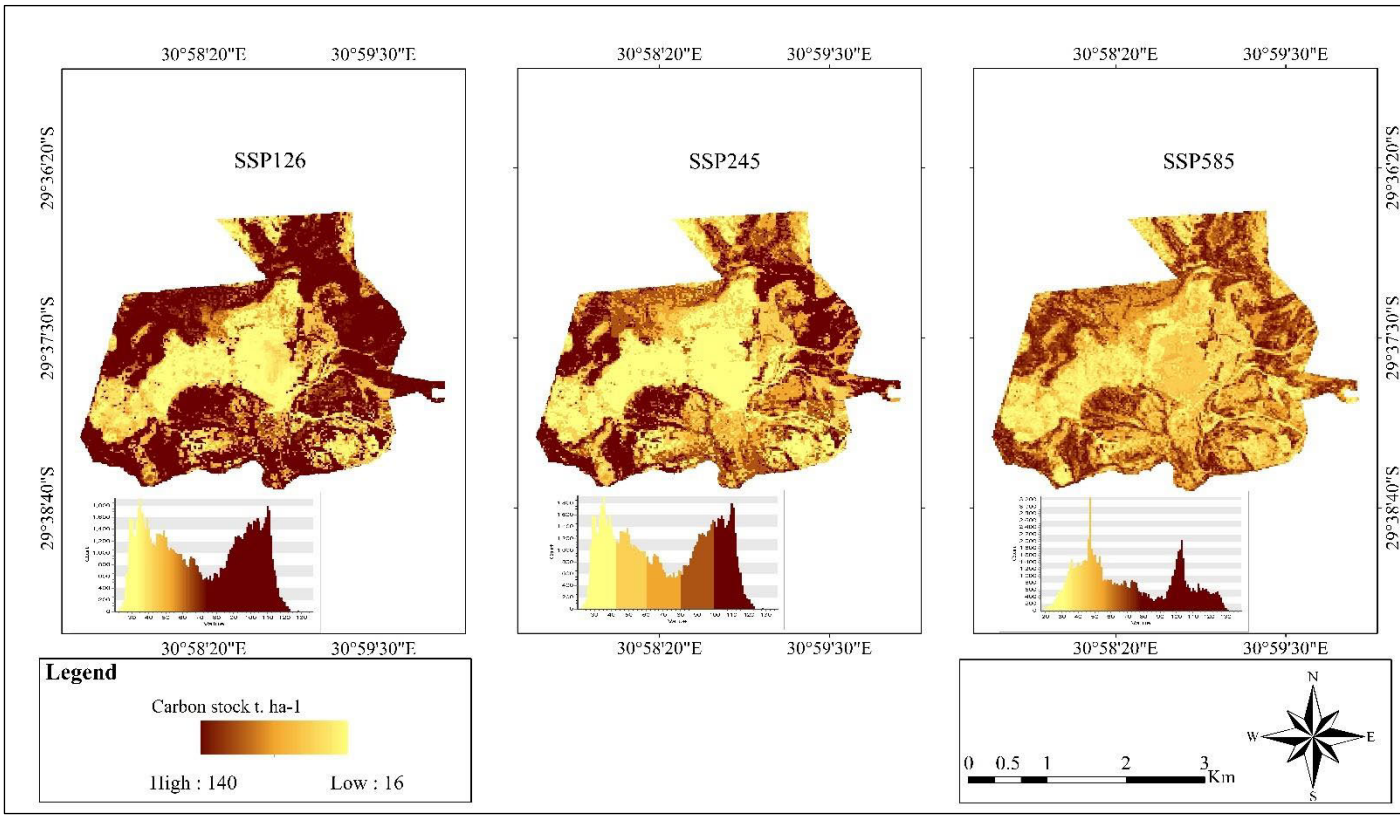


Figure 6.4 Spatial distribution of projected (2040) AGC stocks estimated using ANN under the three share socio-economic pathways.

6.8.1 Evaluation of past, current and future AGC stocks.

Table 6.5 shows changes in forest cover and non-forest cover in hectares (ha) over the past, present, and future projected time periods. Currently, non-forest land cover dominates the landscape at 697.19 ha compared to 350.19 ha of forest cover. However, forests are projected to increase substantially from 350.19 ha in 2022 to 526.30 ha by 2040. This expanded forest cover will occur concurrently with declines in non-forest land area over time.

Table 6.5 Projected Forest cover and non-forest cover change amongst the three time periods expressed in hectares (ha) and % as well as change (Δ) with negative sign showing a decline in land area.

Land use type	**2012 (ha)**	**2022 (ha)**	**2040 (ha)**	**Δ (ha)**	**2012 (%)**	**2022(%)**	**2040 (%)**	**Δ (%)**
Forest cover	260.45	350.19	526.30	266.85	24	33.43	50.27	32.27
Non-forest	787.93	697.19	521.08	-266.85	76	66.67	49.73	-32.27
		(-90.74)	(-176.11)					

Table 6.4 andTable 6.6present the estimated total aboveground carbon stocks (tonnes) and mean aboveground carbon stock density (tonnes per hectare) for the two studied land cover types under the three shared socio-economic pathways, projected over time. There is a notable increase in AGC stocks over time from 12 400 tonnes in 2012 to 39206 tonnes after two decades under the low emission SSP 126 pathway. This aligns with substantial projected forest cover gains (Table 6.5). Non-forest cover type maintained lower density (4.06-14.08 t. ha^{-1}) and stocks throughout from 2012 to 2040. Future variations in AGC vary by emission scenario and period. For instance, Table 6.6 Forest carbon density varied widely, with the 2040 SSP126 projection doubling the 2012 baseline (60.22 vs 35.66 t. ha^{-1}). Negligible changes occurred in stocks/density under high-emission SSP585, in contrast to SSP126.The mechanisms underlying these changes need further investigation in the light of internal processes, human activities and other external forces thereof.

In general, modeled carbon stocks demonstrated continued carbon uptake across the landscape over time, with sink strength magnified under sustainable pathways allowing forest recovery. Table 6.4 andTable 6.6 quantify these gradient effects on density and total stocks under contrasting scenarios.

Table 6.6 Mean aboveground carbon stock density t. ha^{-1} for past, current and future predictions, the value in parenthesis shows differences in mean stock density.

Land use type	Past (2012)	Current (2022)	Future (SSP 126)	Future (SSP 245)	Future (SSP 585)
Forest cover	35.66 t. ha^{-1}	54.99	60.72	56.34	55..66
Non-forest cover	4.06 t. ha^{-1}	9.61	14.08	11.51	11.23

6.9 Discussion

6.9.1 Model performance

The results demonstrate future climate and land use changes substantially impact urban reforestation AGC stocks, with differing trajectories across emission scenarios. Integrating remote sensing and machine learning gives valuable insights to inform land management and climate mitigation strategies, highlighting the carbon sequestration potential of urban reforestation. The results of the ANN model performance and visual fit indicate that the model performed well in estimating the dynamic changes in AGC in 2022 compared to 2012 and 2040 (table 2 and figure 2) with high metrics $R^{2\ =\ 0.80}$ on the validation dataset. Our estimates showed less bias than previous landscape models Verly et al. (2023), and an R^2 comparable to recent similar studies Datta et al. (2023), The strength of ANN lies in handling complex relationships between numerous environmental factors, often outperforming traditional methods. By combining historical data, remote sensing, and future projections, ANNs provide valuable insights into how urban development and climate change might impact carbon storage in urban forests. This is especially relevant in the context of eThekwini Municipality, where rapid urbanization and climate variability present ongoing challenges (Gharaibeh et al., 2020, Douwes et al., 2015a). The performance of ANNs in this study aligns with previous research (Dou and Yang, 2018, Zhang et al., 2022b) that has underscored the efficacy of machine learning approaches in environmental forecasting (Gharaibeh et al., 2020).

The future AGC simulations, however, displayed a reduction in the accuracies (from R^2 = 0.60-0.63, across all three shared socio-economic pathways (SSP126, SSP245, and SSP585) developed from the average of the three CMIP6 models including - ACCESS-ESM1-5, MIROC6, and EC-EARTH2 (van Noije et al., 2021, Hajima et al., 2020). Furthermore, the

absence of Sentinel-2 data and derived indices hindered accurate modeling of future aboveground carbon stocks. Literature highlights the importance of spectral data for AGC mapping, utilizing specific spectral bands and combinations (Mngadi et al., 2022a, Odebiri et al., 2023, Gara et al., 2016). The NIR band reveals canopy structure and density, the red band assesses photosynthetic activity, and the SWIR band enables vegetation differentiation. Indices such as NDVI and EVI integrate NIR and red bands for robust vegetation density measures (Baldocchi et al., 2020, Dube and Mutanga, 2016, Gara et al., 2016). The red-edge band, sensitive to chlorophyll, provides insights into plant stress and biomass (Gara et al., 2016).). In this study, Sentinel-2 and Rapid-eye spectral metrics, along with derived vegetation indices, including red band 4, red-edge bands 7 and 6, near-infrared band 8, NDVI, RENDVI, and GNDVI, were identified as critical variables for AGC estimation in both 2012 and 2022 (Figure 3). However, the reduced accuracy of the future projections is understandable given the data limitations, but these initial simulations still offer valuable insights into potential reforestation contributions and aboveground carbon accumulation moving forward.

6.9.2 Mapping of AGC stocks

Key findings show pronounced northern, eastern and southern AGC density increases over the past decade, with various patches seeing gains exceeding 60 t. ha-^{1} contrasting with more modest accruals averaging 20-30 t. ha-^{1} in central zones. These accumulation patterns likely stem from recent concentrated reforestation efforts alongside supportive localized climate and soil conditions enabling higher survival rates and carbon sequestration capacities in thriving young forests (Matiza et al., 2024a).

The AGC projections under SSP 245 and 585 pathways reveal minimal variances, with density estimates ranging from 55.66-56.34 t. ha-^{1} and total stocks of 25,200-26,400 tonnes between scenarios. This is because shared socioeconomic pathways may incorporate comparable narratives around aspects like population growth, economic development, technological innovations, and environmental policies between scenarios. This could lead to limited divergence in model outputs. These small AGC differences align with other studies applying similar socio-economic pathways (Gao et al., 2023, Ter-Mikaelian et al., 2021, Wang et al., 2022e). However, intermittent rainfall and temperature extremes from intensifying climate change could transform recovering forests into neutral or negative carbon

sinks over time. Adopting a low-emissions pathway appears to support balanced, eco-friendly urban ecosystems where forests significantly thrive. For example, the SSP 126 scenario showed a 102.71 million Mg carbon stock increase by 2050, reflecting substantial sequestration potential given proper conditions (Gao et al., 2023). Overall, quantifying spatial variability in historic AGC densities and projecting future stocks under varying pathways provides insights into targeted reforestation and climate mitigation needs.

6.9.3 Past, current and future changes in AGC stocks

Future projections reveal a divergence in reforestation success and carbon storage potential under varying emissions scenarios. The SSP126 low-emission pathway enables forest cover to expand to 50% by 2040 as trees thrive under reduced climate stress (Velasco et al., 2016). Improved growth conditions reinforce carbon uptake, with reforested areas continuing to effectively sequester CO2 per SSP126 goals. However, under SSP245 and SSP585, forest cover declines to only 35% by 2040, with carbon accumulation plateauing or diminishing over time. Rising emissions severely threaten reforestation efforts and carbon capture viability (Yan et al., 2023). Contrasting outcomes highlight the detrimental impacts of unrestrained emissions compared to an SSP126 trajectory where changes expanding reforested areas and biodiversity conservation unlock vegetation carbon storage and resilience (Lewis et al., 2019). Aligned climate and land use policies therefore hold the key to sustainable carbon management via urban reforestation.

6.9.4 Implications.

The findings underscore the need to incorporate urban reforestation initiatives into city planning and development strategies to maximize carbon uptake and ecosystem resilience to climate change. The study highlights the importance of adaptive forest management, emphasizing sustainable land use and biodiversity conservation to support reforestation success and long-term carbon sequestration. Quantifying the role of urban forests in meeting global carbon goals can inform climate policies across governance levels and incentivize their inclusion in climate action plans.

However, uncertainties remain in the initial carbon mapping and future projections, which rely on modeled nonlinear responses. Abnormal precipitation, increased wildfires, and land-use changes could impact actual carbon stocks differently than predicted. Ongoing monitoring and model refinement are therefore needed to track the complex dynamics of reforested

carbon in the urban environment. Management practices must adapt to environmental influences like precipitation, wildfires, and land-use changes affecting carbon stock dynamics. Sustainable land-use and biodiversity conservation create favorable conditions for successful reforestation and carbon storage. Further research should refine initial carbon mapping and projections, improve carbon stock estimates, and account for abnormal environmental conditions through continued monitoring and data collection. Interdisciplinary collaborations involving urban planners, ecologists, climate scientists, and policymakers can comprehensively understand the socio-ecological dynamics influencing urban reforestation and carbon sequestration. Addressing uncertainties and knowledge gaps through ongoing research and adaptive management is crucial while integrating these insights into comprehensive strategies for sustainable and climate-resilient urban development.

6.10 Conclusion

The study concludes that:

- Urban reforestation's contribution to increasing AGC stocks diminishes in the future, especially under high emission scenarios.
- Future climate change and land-use changes significantly impact urban reforested AGC stocks.
- Remote sensing and machine learning are complementary tools that can improve forest monitoring and inform policy and planning.

These projections provide valuable data to inform land use management and climate mitigation strategies by stakeholders and city authorities. However, fully unlocking the carbon storage capacity of cities requires integrated policy making and land management strategies. Overall, the analysis not only improves understanding of reforestation in urban zones, but also demonstrates the need for cross-sectoral coordination on climate action plans and adaptive ecosystem governance. Targeted monitoring, sustainable development policies, and public-private collaborations will be essential to overcoming uncertainties and steering cities toward more sustainable, climate-resilient trajectories.

6.11 Summary

This chapter demonstrates the variability of reforested carbon stocks under different land use change and climate scenarios. Using artificial neural networks and remote sensing data, past, current, and future carbon accumulation is modeled. The results reveal shifts in carbon stocks

across shared socioeconomic pathway scenarios, offering valuable insights. These findings allow for strategic planning and mitigation around how climate and land cover changes may impact reforestation efforts moving forward. Chapter 7 synthesizes the overall findings of this research study and provides a concise conclusion that ties together the central themes and results. Additionally, Chapter 7 proposes avenues for future research to build upon the knowledge gained, particularly within the context of quantifying and monitoring carbon dynamics in urban reforested areas over time. Recommendations are made to address gaps and limitations identified over the course of this work. Expanding data collection, model ensemble approaches, and scenario-based projection methods are highlighted to further improve understanding of this complex topic at the nexus of reforestation, climate change, land use, and carbon management.

7 CHAPTER SEVEN: Synthesis

This chapter presents the key findings, answering the research questions to address the thesis aim and objectives. It synthesizes recent developments and future prospects for remote sensing and machine learning to assess reforestation impacts on biomass and carbon in urban landscapes.

7.1 Main Findings

This research demonstrates persistent gaps in quantifying and tracking carbon mitigation from urban reforestation projects, despite their surge in global policy support. While cities exert outsized climate impacts, localized evidence remains limited on actual carbon sequestration rates achieved "on the ground" to inform management prioritization.

This dissertation helps address this gap by delivering an accurate, scalable methodology for mapping biomass and carbon stocks across a rapidly developing urban landscape using remote sensing and machine learning. Comparative spatiotemporal analysis enables robust accounting of reforestation impacts to strengthen localized climate action plans through targeted, evidence-based decision-making.

Overall, key contributions include highlighting knowledge shortfalls while developing a novel framework for continuous carbon stock assessments to support climate-smart urban greening and mitigation. The approach underscores immense yet underexploited opportunities for geospatial capabilities to inform management. In summary, outcomes equipped policymakers with an efficient monitoring tool to incentivize maximized carbon mitigation via urban reforestation. The answers to the research questions can be synthesized as follows:

Q1: How accurately can remote sensing data map and monitor forest aboveground biomass and carbon accumulation across an urban reforestation landscape when integrated with machine learning algorithms?

This research question was addressed in Chapter 3 using Sentinel-1, Sentinel-2 and Planetscope images to estimate aboveground biomass, and in Chapter 4 using Planetscope to estimate aboveground carbon stocks. Findings showed the most effective models for capturing both aboveground biomass (AGB) and carbon stocks mainly leveraged vegetation indices linked to greenness. Specifically, optimal predictors included the green normalized difference vegetation index (GNDVI), red edge NDVI, and red edge simple ratio index. This

aligns with the sensitivity of these bands and spectral transformations to photosynthetic material.

The study integrating multi-source satellite data (Sentinel-1, Sentinel-2, Planetscope) with machine learning achieved accurate AGB estimates for the heterogeneous urban forest. This showcases the value of the data fusion approach in data-scarce regions to enable wall-to-wall mapping and monitoring. Broader applications could extend to assessing additional biophysical parameters to inform management.

Additionally, analysis focused specifically on Planetscope data demonstrated reliable modeling of aboveground carbon stocks when combining vegetation indices with terrain variables. This again underscores the potential of remote sensing analytics to quantify ecosystem structure metrics associated with climate regulation functions.

Overall, both investigations emphasize the importance of continuous evaluation of biomass and carbon stocks enabled through scalable remote sensing and machine learning frameworks. This delivers an efficient information pipeline for both reporting and guiding adaptive management of reforestation initiatives over time. Findings carry valuable implications for planning carbon-conscious and ecologically resilient urban landscapes.

In summary, key outcomes accentuate the immense yet underexploited potential of recent advances in earth observation technology combined with sophisticated computational tools to revolutionize monitoring of critical ecosystem processes, even across challenging heterogeneous environments like cities.

Q2: Which category of machine learning algorithms (e.g. ensemble methods and/or neural networks) demonstrates optimal performance for spatially explicit quantification of urban reforestation carbon accumulation using remote sensing inputs?

Specifically, the extreme gradient boosting (XGB) algorithm was identified as the top performing modeling approach for mapping both aboveground biomass and carbon stocks in the reforested urban landscape. The paper evaluating multiple machine learning regression algorithms found XGB achieved the highest accuracy estimates, outperforming other techniques including support vector machines and artificial neural networks. Additionally, the study focused solely on Planetscope data integration with machine learning and concluded that XGB again demonstrated the most robust and reliable carbon stock modeling capacity. Comparative analysis against an ANN model showed higher predictive force. The ensemble method combines multiple decision trees to produce robust estimates. The algorithm's strength likely stems from effectively accounting for nonlinearity in complex remote sensing

feature spaces associated with heterogeneous urban forest structure and condition. Therefore, based on reported evaluations, extreme gradient boosting emerges as the optimal machine learning approach for leveraging remote sensing inputs to quantify and map reforestation carbon sequestration accrual across the urban study region. The technique demonstrates particular promise for continuous wall-to-wall monitoring to guide climate-smart management.

Q3: What specific combinations of spectral bands and, vegetation indices calculated from multispectral data, offer the highest sensitivity for capturing variability in aboveground carbon stocks across diverse urban reforested landscapes?

Based on the research findings, the specific combinations of spectral bands and vegetation indices that offer the highest sensitivity for capturing variability in aboveground carbon stocks across diverse urban reforested landscapes are:

1. Green Normalized Difference Vegetation Index (GNDVI)

The GNDVI leverages the green and NIR bands and showed strong sensitivity to photosynthetic green material. As an indicator of canopy density and health, it effectively captured variability in carbon stocks stored in vegetation biomass.

2. Red Edge Normalized Difference Vegetation Index (NDVI_RE)

The red edge NDVI takes advantage of spectral bands straddling the red edge zone (transition from red to NIR) that track subtle changes in vegetation condition. This sensitivity enabled modeling of differences in carbon storage.

3. Red Edge Simple Ratio Index (SR_RE)

Similarly, the red edge simple ratio index compares a red edge band vs NIR reflectance. This transformation using narrow bands precisely tuned to vegetative properties provided capacity to discriminate variable carbon accumulation levels.

Furthermore, band 3 (green), band 4 (red), band 5-7(red-edge bands) and band 8 (near infrared) were all sensitive to biomass and carbon stock estimation.

In summary, spectral transformations targeting the green peak and red edge zones of vegetation spectral signatures, including the GNDVI, NDVI_RE and SR_RE allowed accurate indirect measurement of photosynthetic material strongly linked to aboveground carbon content across the reforestation project. These indices offer optimal sensitivity for ecosystem structure metrics relevant to climate regulation functions.

Therefore, these three vegetation indices calculated using key multispectral bands (blue, green, red edge, NIR) prove most effective for quantifying and differentiating stored forest carbon across diverse urban reforested landscapes using remote sensing.

Q4: How do environmental factors including terrain, disturbance history, climate, and land use change trajectories influence the spatial distribution of carbon storage across complex urban reforestation landscapes?

The study found that environmental factors like terrain, disturbance history, climate, and land use changes significantly influenced the spatial distribution and accumulation of carbon stocks across the complex urban reforestation landscape in Buffelsdraai, South Africa. The fine-resolution remote sensing data and digital elevation models captured nuanced spatio-temporal variations in vegetation and topography that impacted carbon sequestration. Past disturbances from fire outbreaks and drought led to a decrease in forest cover and carbon stocks between 2014-2016. The use of high-resolution climate data from MERRAclim helped model the effects of climate variables on vegetation growth and carbon capture over time. Furthermore, the trajectory of land use change from agriculture and urban expansion to reforestation directly influenced carbon accumulation patterns. Advanced geospatial and machine learning techniques proved useful in reliably modeling and mapping carbon stock changes across the heterogeneous reforestation landscape under varying environmental conditions over decades. The study highlights the need for integrated urban planning and forest management approaches that incorporate climate-resilient reforestation to enhance carbon sequestration and sustainable urban development.

Q5: How can time-series based neural network models be leveraged to reliably forecast trajectories of aboveground carbon sequestration by urban forests under varying climate and land use change scenarios to inform adaptive management?

The study demonstrates that time-series based neural network models can reliably forecast trajectories of carbon sequestration by urban forests under varying climate and land use scenarios. The models leverage multi-temporal remote sensing data and high-resolution climate datasets to capture nuanced spatial and temporal dynamics in vegetation growth and carbon accumulation. By incorporating robust climate and land use change projections, the neural network models can forecast long-term carbon sequestration potential and spatial variability in urban reforestation areas. This enables adapting management strategies to optimize carbon capture, such as prioritizing planting in areas forecasted to have higher sequestration. The reliable forecasts empower policymakers to integrate climate-resilient

urban reforestation into development plans to mitigate emissions and meet climate goals. Overall, the advanced modeling approach provides data-driven forecasts to guide adaptive management of urban forests for enhanced carbon sequestration and sustainable climate action.

7.2 Prospects of earth observation and machine learning in research for monitoring ecological reforestation: challenges, opportunities and contributions of this thesis.

Recent advances in remote sensing technologies and machine learning have significantly impacted the planning, monitoring, and implementation of reforestation interventions globally (Goetz and Dubayah, 2011, Taylor et al., 2020). The advent of high-precision airborne and satellite optical sensors with enhanced spatial, spectral, and temporal resolutions has enabled timely upscaling of reforestation monitoring worldwide. Current satellite missions are providing unprecedented public access to diverse high-quality datasets and derived products on open platforms (Ustin and Jacquemoud, 2020, Matiza et al., 2023a). These new resources allow detailed characterization of vegetation structure, function, and dynamics at regional to global scales. While freely accessible satellite data enables many crucial assessments, high-resolution imagery under 10m is essential for precision monitoring and mapping of forests at finer scales. However, acquiring proprietary very high-resolution data can be prohibitively expensive for many applications. This thesis utilized freely available 10m resolution radar and optical satellite imagery to assess forest carbon and biomass dynamics at moderate resolution. Additionally, high-resolution (3-6m) optical data enabled fine-scale spatial and temporal analysis. However, global forest monitoring would greatly benefit from spaceborne lidar datasets like GEDI that capture 3D vegetation structure (Silva et al., 2021, Coops et al., 2021). Usage of lidar data has been limited thus far due to complex preprocessing requirements. Future integration of multilayered lidar data with optical/radar imagery and field measurements could significantly improve large-scale monitoring of forest structure, biomass, and carbon sequestration.

The integration of remote sensing with advanced data analytical techniques such as machine learning can strengthen monitoring and modelling of urban reforestation landscapes (Jafarzadeh et al., 2021). In this study, machine learning algorithms combined with satellite data effectively processed and analysed key forest parameters including biomass, generating results comparable to field measurements. Additionally, machine learning forecasted potential carbon stock trajectories under varying climate and land use change scenarios. Total site-level

carbon estimates were reliably modelled from limited field samples, demonstrating the power of fusing remote sensing with machine learning. Overall, this emerging approach enables efficient large-scale mapping and projection of forest structure and carbon dynamics. By leveraging remote sensing's spatial coverage and spectral sensitivity with machine learning's computing power, these technologies provide synergistic capabilities to continuously monitor, understand, and predict reforestation success across heterogenous urban environments.

Emerging collaborative frameworks among interdisciplinary scientists are enabling access to pre-processed datasets, open-source tools, and integrated data fusion techniques (Rex et al., 2020). By harnessing these resources, novel techniques can track forest dynamics at finer resolutions. For instance, this study leveraged continuous spatio-temporal data to detect subtle carbon stock changes across an urban reforestation landscape. Additionally, open data sharing facilitates combining diverse sensors, field measurements, and airborne systems to capitalize on their complementary strengths (Campesato, 2020). Pre-processed data from emerging earth observation platforms and ecosystem decision support systems expands accessibility and uptake of advanced monitoring techniques. Furthermore, open-source software like R, QGIS, and Python empowers widespread application of cutting-edge algorithms. Through embracing these technologies and collaborative frameworks, an interdisciplinary community of practice is forming to transform reforestation research. Their pioneering applications in remote sensing, informatics, and ecological modelling will be critical for scaling intelligent monitoring systems that guide adaptive management for successful global reforestation initiatives.

Despite advancements in remote sensing and monitoring tools, ground-truth data remains essential for validation and calibration (Chave et al., 2019). Earth observation data holds tremendous potential for enhancing carbon and biomass models if effectively leveraged for calibration and validation (Cockburn et al., 2016). Emerging methods that streamline validation workflows could significantly improve large-scale assessments of reforestation success over time using satellite data. For instance, proliferating ground-based monitoring via static cameras provides validation data to verify satellite observations (Van Rooyen et al., 2013). Complementary airborne and UAV data at higher resolutions and from customizable sensors can add spectral, spatial, and temporal detail at key phases. Especially for initial planning and monitoring of reforestation sites, drones offer immense utility if costs can be reduced to enable widespread adoption. Overall, strategically integrating ground, airborne, and satellite systems and developing efficient validation techniques is critical to unlock the

full capabilities of earth observation for continuously evaluating and guiding adaptive management of reforestation projects across scales.

A major barrier to effective monitoring is the predominance of short-term programs, whereas at least ten years are required to observe conservation impacts and ecological change in reforestation efforts (Cockburn et al., 2016, Matiza et al., 2024a). Long-term monitoring is critical to evaluate if urban reforestation landscapes stay on target trajectories over time (Mngadi et al., 2022b). Robust systems are needed that can detect deviations and trigger interventions. However, most initiatives lack monitoring beyond a few years post-implementation. This study demonstrates the value of long-term remote sensing to assess variability in biomass and carbon stocks beyond the initial reforestation timeline. Still, improved understanding and communication is needed around applying remote sensing and machine learning in ecological restoration. Additional research should focus on quantifying uncertainty in these datasets for reforestation specifically. Furthermore, advanced monitoring systems must translate complex spatio-temporal insights into actionable information for practitioners, land managers, and policy makers. Overall, implementing continuous monitoring frameworks with adaptive management protocols is essential to guide reforestation trajectories for lasting success.

7.3 Conclusion

This research presented a remote sensing and machine learning framework for quantifying and mapping aboveground carbon sequestration by reforestation initiatives in urban landscapes over time. Comparing biomass and carbon capture overtime revealed key environmental drivers and management strategies for maximizing reforestation's climate impact. Overall, findings showed significant carbon accumulation across the study region between 2012-2022 resulting from tree plantings, with rates varying based on ecological histories. This highlights the potential of urban reforestation in offsetting carbon emissions through nature-based solutions. Variability based on local conditions points to opportunities for targeted monitoring and interventions amplifying carbon capture efficiency.

Additionally, the simulation of future carbon trajectories between 2022 and 2040 using forecasted changes in climate and land development patterns underscored risks of losses from extreme events and land use pressures, respectively. This emphasizes the need for proactive policymaking centered on adaptive expansion and protecting high carbon sequestering green spaces. The methodological framework presented enables localized tracking of spatio-temporal dynamics in aboveground carbon stocks across reforestation sites through

integration of multi-source optical and radar satellite imagery with sophisticated machine learning. The system has potential to strengthen localized carbon accounting, transparency in monitoring, reporting and verification, and evidence-based management practices fundamental to urban climate change mitigation policy.

Conclusions based on the results from each objective were:

Objective 1: A systematic review shows that remote sensing and machine learning can effectively estimate carbon storage in natural forests, guiding the selection of optimal sensors and algorithms.

Objective 2: Integrating satellite imagery with machine learning accurately quantifies above-ground biomass and carbon in heterogeneous reforested urban landscapes, enabling monitoring where field data is limited.

Objective 3: PlanetScope data can reliably model and map carbon stocks in reforested urban areas when combined with spectral indices and terrain variables, highlighting the utility of remote sensing and machine learning.

Objective 4: Analysis of time-series RapidEye data demonstrates that collaborative urbn reforestation increased carbon sequestration over time. Machine learning can create a timeous robust framework for monitoring the progress of reforested landscapes.

Objective 5: Advanced geospatial and machine learning techniques enable modeling of past, current and projected carbon stock changes under varying scenarios specific to reforested urban landscapes to support planning.

Overall, the research highlights the significant capabilities of remote sensing and machine learning to quantify, map, and track carbon accumulation by urban reforestation initiatives over time. The findings provide robust evidence to guide the optimization of urban reforestation for maximizing climate change mitigation through targeted monitoring, strategic expansion, and adaptive management.

7.4 Implications for Policy and Practice:

1. Develop targeted urban reforestation strategies: Based on the findings that carbon accumulation rates vary depending on local ecological histories and conditions, policymakers and urban planners should develop targeted reforestation strategies tailored to specific urban areas. This could involve identifying and prioritizing areas with high carbon sequestration potential, considering factors such as soil quality, microclimate, and existing vegetation.

2. Implement adaptive management and monitoring practices: To address the risks of carbon loss due to extreme weather events and land-use pressures, policymakers should adopt adaptive management practices. This could involve regularly monitoring carbon stocks, implementing measures to protect high-sequestering green spaces, and developing contingency plans for replanting or restoration efforts in the event of disturbances.
3. Promote cross-sectoral collaboration and public-private partnerships: Enhancing urban reforestation efforts and carbon sequestration monitoring requires collaboration among various stakeholders, including policymakers, urban planners, researchers, environmental organizations, and private entities. Public-private partnerships can leverage resources, expertise, and funding to support large-scale reforestation initiatives and implement robust monitoring systems.
4. Integrate remote sensing and machine learning into urban carbon accounting: Policymakers should consider integrating the proposed remote sensing and machine learning framework into their urban carbon accounting and reporting processes. This approach can provide localized, spatially explicit data on carbon stocks, enabling more accurate monitoring, reporting, and verification of carbon sequestration efforts.
5. Develop incentives and regulations for carbon sequestration: To encourage and incentivize urban reforestation efforts, policymakers could explore mechanisms such as carbon credits, tax incentives, or regulatory requirements for incorporating carbon sequestration considerations into urban development projects.
6. Raise public awareness and promote community engagement: Successful urban reforestation initiatives require public support and participation. Policymakers should promote public awareness campaigns and community engagement programs to educate residents about the benefits of urban reforestation and encourage active involvement in tree planting and maintenance efforts.

By implementing these recommendations, policymakers, urban planners, and other stakeholders can leverage the research findings to develop more effective strategies for enhancing urban reforestation efforts, maximizing carbon sequestration potential, and contributing to climate change mitigation efforts in urban areas.

7.5 Prospects for the implementation of earth observation and machine learning based monitoring and evaluation of reforestation intervations .

While this research demonstrates the immense potential of remote sensing and machine learning for quantifying and tracking carbon sequestration by urban reforestation efforts, further investigations are needed to strengthen monitoring and evidence-based management capabilities. Specific directions for additional research include:

- Further testing of deep learning approaches for carbon stock modeling in urban forests using multi-temporal, multi-sensor data to improve accuracy and robustness.

Exploring integration of hyper-spatial drone imagery with satellite data to capture structural complexity for enhanced carbon mapping in heterogeneous urban reforestation.

- Leveraging thermal imagery to assess vegetation health and link to urban forest productivity and carbon sequestration potentials.
- Utilizing forthcoming satellite missions with imaging spectroscopy to map plant biodiversity and associated carbon impacts across reforested city landscapes.
- Expanding the spatio-temporal scope of analyses to cities across climate zones for comparative urban reforestation carbon assessments.
- Incorporating projected land use, land cover and climate data into carbon projections to quantify uncertainty and guide adaptive management.
- Linking carbon mapping to economic valuations and cost-benefit analysis to inform reforestation policy and investment prioritization.
- Developing monitoring systems and decision support tools for near real-time tracking and adaptive management of urban forest projects.
- Refining biomass estimation equations and quantifying carbon fractions specific to diverse reforested urban landscapes through collaborative field, lab and modeling research.

7.6 Future Research Directions:

While this research provides valuable insights into quantifying and mapping aboveground carbon sequestration through urban reforestation initiatives, several gaps in knowledge and areas for further investigation have been identified:

1. Long-term monitoring and modeling: This study focused on a specific time frame (2012-2022) and simulated future scenarios until 2040. However, continuous long-term

monitoring and modeling of urban reforestation sites are necessary to better understand the long-term dynamics of carbon sequestration and the impacts of climate change, urbanization, and other factors over extended periods.

2. Integrating belowground carbon dynamics: The current study focused on aboveground carbon stocks in urban forests. Future research should aim to incorporate belowground carbon dynamics, including root biomass, soil organic carbon, and their interactions with aboveground components, to provide a more comprehensive understanding of the total carbon sequestration potential of urban reforestation initiatives.

3. Exploring the role of species composition and biodiversity: This study examined the overall carbon sequestration patterns of urban reforestation sites. However, future research could investigate the specific contributions of different tree species and the potential benefits of promoting biodiversity in urban forests for enhanced carbon capture and storage.

4. Socio-economic and environmental justice considerations: While this research focused on the ecological aspects of urban reforestation, future studies could explore the socio-economic and environmental justice implications of urban reforestation initiatives, including the equitable distribution of benefits, community engagement, and the potential for addressing environmental inequalities.

5. Refining remote sensing and machine learning techniques: As remote sensing technology and machine learning algorithms continue to evolve, future research should explore ways to further refine and improve the methodological framework presented in this study, potentially incorporating new data sources, advanced algorithms, and higher-resolution imagery for more accurate and detailed carbon stock mapping.

6. Comparative studies across diverse urban landscapes: This research focused on a specific urban landscape. Conducting comparative studies across diverse urban environments, considering factors such as climate, geography, urbanization patterns, and socio-economic conditions, could provide valuable insights into the transferability and scalability of the findings and methodologies.

7. Integrating urban reforestation with other climate mitigation strategies: Future research could investigate the synergies and trade-offs between urban reforestation and other climate mitigation strategies, such as renewable energy deployment, sustainable transportation, and energy-efficient buildings, to develop integrated and comprehensive approaches to urban climate change mitigation.

By addressing these future research directions, scientists, policymakers, and urban planners can further strengthen the knowledge base and practical applications of urban reforestation as a nature-based solution for carbon sequestration and climate change mitigation in urban areas.

8 REFERENCES

ABDEL-HAMID, A., DUBOVYK, O., ABOU EL-MAGD, I. & MENZ, G. 2018. Mapping Mangroves Extents on the Red Sea Coastline in Egypt using Polarimetric SAR and High Resolution Optical Remote Sensing Data. *SUSTAINABILITY*, 10.

ABIODUN, O. I., KIRU, M. U., JANTAN, A. B., OMOLARA, A. E., DADA, K. V., UMAR, A., LINUS, O. U., ARSHAD, H., KAZAURE, A. A. & GANA, U. M. 2019. Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition. *IEEE Access*, 7**,** 158820-158846.

ABUDU, D., PARVIN, N. S. & ANDOGAH, G. 2020. Reviewing the pertinence of Sentinel-1 SAR for urban land use land cover classification.

AGJEE, N. E. H., ISMAIL, R. & MUTANGA, O. 2016. Identifying relevant hyperspectral bands using Boruta: a temporal analysis of water hyacinth biocontrol. *Journal of Applied Remote Sensing*, 10**,** 042002.

AHIRWAL, J., NATH, A., BRAHMA, B., DEB, S., SAHOO, U. K. & NATH, A. J. 2021. Patterns and driving factors of biomass carbon and soil organic carbon stock in the Indian Himalayan region. *Science of the Total Environment*, 770**,** 145292.

AHMAD, A., GILANI, H. & AHMAD, S. R. 2021. Forest aboveground biomass estimation and mapping through high-resolution optical satellite imagery—A literature review. *Forests*, 12**,** 914.

AHMAD, A., LIU, Q. J., NIZAMI, S. M., MANNAN, A. & SAEED, S. 2018. Carbon emission from deforestation, forest degradation and wood harvest in the temperate region of Hindukush Himalaya, Pakistan between 1994 and 2016. *Land Use Policy*, 78**,** 781-790.

AHMADPOUR KASGARI, Z. 2018. Elaborative text modification vs. input flooding: a case study on non-congruent collocations. *Asian-Pacific Journal of Second and Foreign Language Education*, 3**,** 1-15.

AHMED, N., ATZBERGER, C. & ZEWDIE, W. 2022. The potential of modeling Prosopis Juliflora invasion using Sentinel-2 satellite data and environmental variables in the dryland ecosystem of Ethiopia. *Ecological Informatics*, 68**,** 101545.

AHMED, O. S., FRANKLIN, S. E., WULDER, M. A. & WHITE, J. C. 2015. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101**,** 89-101.

AKHTAR, A. M., QAZI, W. A., AHMAD, S. R., GILANI, H., MAHMOOD, S. A. & RASOOL, A. 2020. Integration of high-resolution optical and SAR satellite remote sensing datasets for aboveground biomass estimation in subtropical pine forest, Pakistan. *Environmental Monitoring and Assessment*, 192**,** 584.

ALI, A. & MATTSSON, E. 2017. Individual tree size inequality enhances aboveground biomass in homegarden agroforestry systems in the dry zone of Sri Lanka. *Science of the Total Environment*, 575**,** 6-11.

ALI, A., ULLAH, S., BUSHRA, S., AHMAD, N., ALI, A. & KHAN, M. A. 2018. Quantifying forest carbon stocks by integrating satellite images and forest inventory data. *Austrian Journal of Forest Science*, 135**,** 93-117.

ALI, I., GREIFENEDER, F., STAMENKOVIC, J., NEUMANN, M. & NOTARNICOLA, C. 2015. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing*, 7**,** 16398-16421.

ALIG, R. J. 2003. *Land use changes involving forestry in the United States, 1952 to 1997, with projections to 2050*, Pacific Northwest Research Station.

ALMEIDA, C. T. D., GALVÃO, L. S., ARAGÃO, L. E. D. O. C. E., OMETTO, J. P. H. B., JACON, A. D., PEREIRA, F. R. D. S., SATO, L. Y., LOPES, A. P., GRAÇA, P. M. L. D. A., SILVA, C. V. D. J., FERREIRA-FERREIRA, J. & LONGO, M. 2019. Combining LiDAR and hyperspectral data for aboveground biomass modeling in the

Brazilian Amazon using different regression algorithms. *Remote Sensing of Environment,* 232**,** 111323.
AMATULLI, G., DOMISCH, S., TUANMU, M.-N., PARMENTIER, B., RANIPETA, A., MALCZYK, J. & JETZ, W. 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific data,* 5**,** 1-15.
ANAYA, J. A., CHUVIECO, E. & PALACIOS-ORUETA, A. 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management,* 257**,** 1237-1246.
ANDERSON, C., NAUGHTON, D., BRUNN, A. & THIELE, M. Radiometric correction of RapidEye imagery using the on-orbit side-slither method. Remote Sensing, 2011.
ANDREATTA, D., GIANELLE, D., SCOTTON, M. & DALPONTE, M. 2022. Estimating grassland vegetation cover with remote sensing: A comparison between Landsat-8, Sentinel-2 and PlanetScope imagery. *Ecological Indicators,* 141**,** 109102.
ANGELSEN, A., BROWN, S. & LOISEL, C. 2009. Reducing emissions from deforestation and forest degradation (REDD): an options assessment report.
ANJALI, K., KHUMAN, Y. & SOKHI, J. 2020. A Review of the interrelations of terrestrial carbon sequestration and urban forests: Terrestrial carbon sequestration and urban forests. *AIMS Environmental Science,* 7.
ANSELIN, L. 2020. Local spatial autocorrelation. *Other Local Spatial Autocorrelation Statistics*.
APOSTOL, B., PETRILA, M., LORENŢ, A., CICEU, A., GANCZ, V. & BADEA, O. 2020. Species discrimination and individual tree detection for predicting main dendrometric characteristics in mixed temperate forests by use of airborne laser scanning and ultra-high-resolution imagery. *Science of The Total Environment,* 698**,** 134074.
APPS, M. J. & PRICE, D. T. Forest ecosystems, forest management and the global carbon cycle. 1996.
ARYAL, A., BRUNTON, D., PANDIT, R., SHRESTHA, T. K., LORD, J., KOIRALA, R. K., THAPA, Y. B., ADHIKARI, B., JI, W. & RAUBENHEIMER, D. 2012. Biological diversity and management regimes of the northern Barandabhar Forest Corridor: an essential habitat for ecological connectivity in Nepal. *Tropical Conservation Science,* 5**,** 38-49.
ASRAT, Z., TADDESE, H., ØRKA, H. O., GOBAKKEN, T., BURUD, I. & NÆSSET, E. 2018. Estimation of Forest Area and Canopy Cover Based on Visual Interpretation of Satellite Images in Ethiopia. *Land,* 7**,** 92.
AUSTIN, K., BAKER, J., SOHNGEN, B., WADE, C., DAIGNEAULT, A., OHREL, S., RAGNAUTH, S. & BEAN, A. 2020. The economic costs of planting, preserving, and managing the world's forests to mitigate climate change. *Nature communications,* 11**,** 5946.
AVITABILE, V., BACCINI, A., FRIEDL, M. A. & SCHMULLIUS, C. 2012. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sensing of Environment,* 117**,** 366-380.
AWAD, M., KHANNA, R., AWAD, M. & KHANNA, R. 2015. Support vector regression. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers***,** 67-80.
BACCINI, A., FRIEDL, M., WOODCOCK, C. & WARBINGTON, R. 2004. Forest biomass estimation over regional scales using multisource data. *Geophysical research letters,* 31.
BAINES, O., WILKES, P. & DISNEY, M. 2020. Quantifying urban forest structure with open-access remote sensing data sets. *Urban Forestry & Urban Greening,* 50**,** 126653.
BAKER, T. R., PHILLIPS, O. L., MALHI, Y., ALMEIDA, S., ARROYO, L., DI FIORE, A., ERWIN, T., KILLEEN, T. J., LAURANCE, S. G. & LAURANCE, W. F. 2004. Variation in wood density determines spatial patterns inAmazonian forest biomass. *Global Change Biology,* 10**,** 545-562.
BALDOCCHI, D. D., RYU, Y., DECHANT, B., EICHELMANN, E., HEMES, K. S., MA, S., SANCHEZ, C. R., SHORTT, R., SZUTU, D. J., VALACH, A. C., VERFAILLIE, J. G., BADGLEY, G., ZENG, Y. & BERRY, J. A. 2020. Outgoing Near-Infrared Radiation From Vegetation Scales With Canopy Photosynthesis Across a Spectrum of

Function, Structure, Physiological Capacity, and Weather. *Journal of Geophysical Research: Biogeosciences*, 125.
BALOLOY, A. B., BLANCO, A.C., CANDIDO, C.G., ARGAMOSA, R.J.L., DUMALAG, J.B.L.C., DIMAPILIS, L.L.C. AND PARINGIT, E.C. 2018. Estimation of mangrove forest aboveground biomass using multispectral bands, vegetation indices and biophysical variables derived from optical satellite imageries: rapideye, planetscope and sentinel-2. . *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences,*, 4.
BANNARI, A., MORIN, D., BONN, F. & HUETE, A. 1995. A review of vegetation indices. *Remote sensing reviews,* 13**,** 95-120.
BARBOSA, J., BROADBENT, E. & BITENCOURT, M. 2014. Remote sensing of aboveground biomass in tropical secondary forests: A review. *International Journal of Forestry Research,* 2014.
BASYUNI, M., WIRASATRIYA, A., IRYANTHONY, S. B., AMELIA, R., SLAMET, B., SULISTIYONO, N., PRIBADI, R., SUMARGA, E., EDDY, S. & AL MUSTANIROH, S. S. 2023. Aboveground biomass and carbon stock estimation using UAV photogrammetry in Indonesian mangroves and other competing land uses. *Ecological Informatics,* 77**,** 102227.
BAZZATO, E., ROSATI, L., CANU, S., FIORI, M., FARRIS, E. & MARIGNANI, M. 2021. High spatial resolution bioclimatic variables to support ecological modelling in a Mediterranean biodiversity hotspot. *Ecological Modelling,* 441**,** 109354.
BELGIU, M. & DRĂGUŢ, L. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing,* 114**,** 24-31.
BHATTARAI, T., SKUTSCH, M., MIDMORE, D. & SHRESTHA, H. L. 2015. Carbon measurement: An overview of forest carbon estimation methods and the role of geographical information system and remote sensing techniques for REDD+ implementation. *Journal of Forest and Livelihood,* 13**,** 69-86.
BIFARIN, O. O. 2022. Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification. *PLOS ONE,* 18.
BIRDSEY, R. & PAN, Y. 2015. Trends in management of the world's forests and impacts on carbon stocks. *Forest Ecology and Management,* 355**,** 83-90.
BJÖRK, S., ANFINSEN, S. N., NÆSSET, E., GOBAKKEN, T. & ZAHABU, E. Generation of lidar-predicted forest biomass maps from radar backscatter with conditional generative adversarial networks. IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, 2020. IEEE, 4327-4330.
BJÖRK, S., ANFINSEN, S. N., NÆSSET, E., GOBAKKEN, T. & ZAHABU, E. 2022. On the Potential of Sequential and Nonsequential Regression Models for Sentinel-1-Based Biomass Prediction in Tanzanian Miombo Forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 15**,** 4612-4639.
BORT ESCABIAS, C. 2017. *Tree Boosting Data Competitions with XGBoost.* Universitat Politècnica de Catalunya.
BREIMAN, L. 2001. Random forests. *Machine learning,* 45**,** 5-32.
BROWN, S. 1997. *Estimating biomass and biomass change of tropical forests: a primer*, Food & Agriculture Org.
BUYANTUYEV, A. & WU, J. 2010. Urban heat islands and landscape heterogeneity: linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landscape ecology,* 25**,** 17-33.
C VEGA, G., PERTIERRA, L. R. & OLALLA-TÁRRAGA, M. Á. 2017. MERRAclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling. *Scientific data,* 4**,** 1-12.
CAI, B. & ZHANG, L. 2014. Urban CO2 emissions in China: spatial boundary and performance comparison. *Energy Policy,* 66**,** 557-567.
CAMPESATO, O. 2020. *Artificial Intelligence, Machine Learning, and Deep Learning*, Stylus Publishing, LLC.

CAO, L., COOPS, N. C., INNES, J. L., SHEPPARD, S. R. J., FU, L., RUAN, H. & SHE, G. 2016. Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data. *Remote Sensing of Environment,* 178**,** 158-171.

CARTER, T., PARRY, M., HARASAWA, H. & NISHIOKA, S. IPCC technical guidelines for assessing climate change impacts and adaptations. Part of the IPCC Special Report to the First Session of the Conference of the Parties to the UN Framework Convention on Climate Change, Intergovernmental Panel on Climate Change. Department of Geography, University College London, UK and Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan, 1994.

CASTILLO, J. A. A., APAN, A. A., MARASENI, T. N. & SALMO III, S. G. 2017. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS Journal of Photogrammetry and Remote Sensing,* 134**,** 70-85.

CHAN, T., GOMEZ, C. A., KOTHIKAR, A. & BAIZ, P. 2022. Joint Study of Above Ground Biomass and Soil Organic Carbon for Total Carbon Estimation using Satellite Imagery in Scotland. *arXiv preprint arXiv:2205.04870*.

CHAVE, J., DAVIES, S. J., PHILLIPS, O. L., LEWIS, S. L., SIST, P., SCHEPASCHENKO, D., ARMSTON, J., BAKER, T. R., COOMES, D. & DISNEY, M. 2019. Ground data are essential for biomass remote sensing missions. *Surveys in Geophysics,* 40**,** 863-880.

CHAVE, J., RÉJOU-MÉCHAIN, M., BÚRQUEZ, A., CHIDUMAYO, E., COLGAN, M. S., DELITTI, W. B., DUQUE, A., EID, T., FEARNSIDE, P. M. & GOODMAN, R. C. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global change biology,* 20**,** 3177-3190.

CHAVES, P. P., ZUQUIM, G., RUOKOLAINEN, K., KALLIOLA, R., GÓMEZ RIVERO, E. & TUOMISTO, H. 2020. Mapping floristic patterns of trees in Peruvian Amazonia using remote sensing and machine learning. *Remote Sensing,* 12**,** 1523.

CHEN, C., MA, Y., REN, G. & WANG, J. 2022a. Aboveground biomass of salt-marsh vegetation in coastal wetlands: Sample expansion of in situ hyperspectral and Sentinel-2 data using a generative adversarial network. *Remote Sensing of Environment,* 270**,** 112885.

CHEN, C. & XU, Z. 2010. Forest ecosystem responses to environmental changes: the key regulatory role of biogeochemical cycling. *Journal of Soils and Sediments,* 10**,** 210-214.

CHEN, H., COVERT, I. C., LUNDBERG, S. M. & LEE, S.-I. 2023. Algorithms to estimate Shapley value feature attributions. *Nature Machine Intelligence***,** 1-12.

CHEN, M., QIU, X., ZENG, W. & PENG, D. 2022b. Combining Sample Plot Stratification and Machine Learning Algorithms to Improve Forest Aboveground Carbon Density Estimation in Northeast China Using Airborne LiDAR Data. *Remote Sensing,* 14**,** 1477.

CHEN, S., HE, C., HUANG, Z., XU, X., JIANG, T., HE, Z., LIU, J., SU, B., FENG, H. & YU, Q. 2022c. Using support vector machine to deal with the missing of solar radiation data in daily reference evapotranspiration estimation in China. *Agricultural and Forest Meteorology,* 316**,** 108864.

CHEN, T. & GUESTRIN, C. 2016a. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* San Francisco, California, USA: Association for Computing Machinery.

CHEN, T. & GUESTRIN, C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016b. 785-794.

CHEN, Y. & XIAO, W. 2019. Estimation of forest NPP and carbon sequestration in the Three Gorges Reservoir Area, using the biome-BGC model. *Forests,* 10**,** 149.

CHEN, Y. Z., CHEN, L. Y., CHENG, Y., JU, W. M., CHEN, H. Y. H. & RUAN, H. H. 2020. Afforestation promotes the enhancement of forest LAI and NPP in China. *Forest Ecology and Management,* 462**,** 8.

CHENG, F., OU, G., WANG, M. & LIU, C. 2024. Remote Sensing Estimation of Forest Carbon Stock Based on Machine Learning Algorithms. *Forests,* 15**,** 681.
CHENG, Y. & XIAO, Y. 2022. Factors of carbon emissions from Chinese urban and rural residents: a time-varying study. *Applied Economics Letters,* 29**,** 1696-1701.
CHINEMBIRI, T. S., MUTANGA, O. & DUBE, T. 2023. Hierarchical Bayesian geostatistics for C stock prediction in disturbed plantation forest in Zimbabwe. *Ecological Informatics,* 73**,** 101934.
CHURKINA, G. 2016. The role of urbanization in the global carbon cycle. *Frontiers in Ecology and Evolution,* 3**,** 144.
CLEUGH, H. & GRIMMOND, S. 2011. Urban climates and global climate change. *The Future of the World's Climate (Second Edition)***,** 47-76.
CLEVERS, J. & GITELSON, A. Using the red-edge bands on Sentinel-2 for retrieving canopy chlorophyll and nitrogen content. 2012.
CLEVERS, J. G. & GITELSON, A. A. 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation,* 23**,** 344-351.
CLIFF, A. & ORD, K. 1972. Testing for spatial autocorrelation among regression residuals. *Geographical analysis,* 4**,** 267-284.
COCKBURN, J., ROUGET, M., SLOTOW, R., ROBERTS, D., BOON, R., DOUWES, E., O'DONOGHUE, S., DOWNS, C. T., MUKHERJEE, S. & MUSAKWA, W. 2016. How to build science-action partnerships for local land-use planning and management: lessons from Durban, South Africa. *Ecology and Society,* 21.
COOK-PATTON, S. C., LEAVITT, S. M., GIBBS, D., HARRIS, N. L., LISTER, K., ANDERSON-TEIXEIRA, K. J., BRIGGS, R. D., CHAZDON, R. L., CROWTHER, T. W. & ELLIS, P. W. 2020. Mapping carbon accumulation potential from global natural forest regrowth. *Nature,* 585**,** 545-550.
COOPS, N. C., TOMPALSKI, P., GOODBODY, T. R. H., QUEINNEC, M., LUTHER, J. E., BOLTON, D. K., WHITE, J. C., WULDER, M. A., VAN LIER, O. R. & 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**,** 112477.
COPPIN, P. R. & BAUER, M. E. 1994. Processing of multitemporal Landsat TM imagery to optimize extraction of forest cover change features. *IEEE Transactions on Geoscience and remote Sensing,* 32**,** 918-927.
COSENZA, D. N., PACKALEN, P., MALTAMO, M., VARVIA, P., RÄTY, J., SOARES, P., TOMÉ, M., STRUNK, J. L. & KORHONEN, L. 2022. Effects of numbers of observations and predictors for various model types on the performance of forest inventory with airborne laser scanning. *Canadian Journal of Forest Research,* 52**,** 385-395.
COURSOLLE, C., MARGOLIS, H., GIASSON, M.-A., BERNIER, P.-Y., AMIRO, B., ARAIN, M., BARR, A., BLACK, T., GOULDEN, M. & MCCAUGHEY, J. 2012. Influence of stand age on the magnitude and seasonality of carbon fluxes in Canadian forests. *Agricultural and Forest Meteorology,* 165**,** 136-148.
COWAN, D. R. & COOPER, G. R. J. 2004. The Shuttle Radar Topography Mission–a new source of near-global digital elevation data. *ASEG Extended Abstracts,* 2004**,** 1 - 4.
CROMPTON, J. L. 2005. The impact of parks on property values: empirical evidence from the past two decades in the United States. *Managing Leisure,* 10**,** 203-218.
CUI, L., CHEN, P., WANG, L., LI, J. & LING, H. 2021. Application of extreme gradient boosting based on grey relation analysis for prediction of compressive strength of concrete. *Advances in Civil Engineering,* 2021**,** 1-14.
CUNHA-LIGNON, M., MAHIQUES, M. M., SCHAEFFER-NOVELLI, Y., RODRIGUES, M., KLEIN, D. A., GOYA, S. C., MENGHINI, R. P., TOLENTINO, C. C., CINTRÓN-MOLERO, G. & DAHDOUH-GUEBAS, F. 2009. Analysis of mangrove forest succession, using sediment cores: a case study in the Cananéia-Iguape coastal system, São Paulo Brazil. *Brazilian Journal of Oceanography,* 57**,** 161-174.
CUTLER, M., BOYD, D., FOODY, G. & VETRIVEL, A. 2012. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An

assessment of predictions between regions. *ISPRS Journal of Photogrammetry and Remote Sensing,* 70**,** 66-77.
DAHY, B., ISSA, S., KSIKSI, T. & SALEOUS, N. Geospatial Technology Methods for Carbon Stock Assessment: A Comprehensive Review. IOP Conference Series: Earth and Environmental Science, 2020. IOP Publishing, 012036.
DAINELLI, R., TOSCANO, P., DI GENNARO, S. F. & MATESE, A. 2021. Recent advances in unmanned aerial vehicle forest remote sensing—A systematic review. part I: A general framework. *Forests,* 12**,** 327.
DANG, A. T. N., NANDY, S., SRINET, R., LUONG, N. V., GHOSH, S. & KUMAR, A. S. 2019. Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam. *Ecological Informatics,* 50**,** 24-32.
DANTAS, D., TERRA, M. D. C. N. S., SCHORR, L. P. B. & CALEGARIO, N. 2021. Machine learning for carbon stock prediction in a tropical forest in Southeastern Brazil. *Revista Bosque,* 42**,** 131-140.
DAS, S. K., PANT, M. & BEBORTTA, S. 2020. Geospatial data analytics: A machine learning perspective. *Available at SSRN 3599656*.
DATTA, D., DEY, M., KUMAR GHOSH, P., NEOGY, S. & KUMAR ROY, A. 2023. Coupling multi-sensory earth observation datasets, in-situ measurements, and machine learning algorithms for total blue C stock estimation of an estuarine mangrove forest. *Forest Ecology and Management,* 546**,** 121345.
DAVID, R. M., ROSSER, N. J. & DONOGHUE, D. N. 2022. Improving above ground biomass estimates of Southern Africa dryland forests by combining Sentinel-1 SAR and Sentinel-2 multispectral imagery. *Remote Sensing of Environment,* 282**,** 113232.
DE LUCIA, G., LAPEGNA, M. & ROMANO, D. 2022. Towards explainable AI for hyperspectral image classification in edge computing environments. *Computers and Electrical Engineering,* 103**,** 108381.
DE OCAMPO, A. L. P. Normalized Difference Vegetation Index (NDVI) Estimation based on Filter Augmented Imaging. 2023 International Electrical Engineering Congress (iEECON), 2023. IEEE, 84-88.
DEMUZERE, M., ORRU, K., HEIDRICH, O., OLAZABAL, E., GENELETTI, D., ORRU, H., BHAVE, A. G., MITTAL, N., FELIÚ, E. & FAEHNLE, M. 2014. Mitigating and adapting to climate change: Multi-functional and multi-scale assessment of green urban infrastructure. *Journal of environmental management,* 146**,** 107-115.
DEN BESTEN, J. W., ARTS, B. & VERKOOIJEN, P. 2014. The evolution of REDD+: an analysis of discursive-institutional dynamics. *Environmental Science & Policy,* 35**,** 40-48.
DHUMAL, R. K., VIBHUTE, A. D., NAGNE, A. D., KALE, K. V. & MEHROTRA, S. C. Performance analysis of spectral features based on narrowband vegetation indices for cotton and maize crops by EO-1 Hyperion dataset. Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM, 10-11 December 2016, 2017. Springer, 581-590.
DI SACCO, A., HARDWICK, K. A., BLAKESLEY, D., BRANCALION, P. H., BREMAN, E., CECILIO REBOLA, L., CHOMBA, S., DIXON, K., ELLIOTT, S. & RUYONGA, G. 2021. Ten golden rules for reforestation to optimize carbon sequestration, biodiversity recovery and livelihood benefits. *Global Change Biology,* 27**,** 1328-1348.
DIETRICH, R., HOUDE, J., JONES, L., PRIME, J., SCHWENKER, C. & STAPLES, E. 2014. Reforestation Success in an Urban-industrial Landscape.
DJOMO, A. N., PICARD, N., FAYOLLE, A., HENRY, M., NGOMANDA, A., PLOTON, P., MCLELLAN, J., SABOROWSKI, J., ADAMOU, I. & LEJEUNE, P. 2016. Tree allometry for estimation of carbon stocks in African tropical forests. *Forestry,* 89**,** 446-455.
DOBBS, C., HERNÁNDEZ-MORENO, Á., REYES-PAECKE, S. & MIRANDA, M. D. 2018. Exploring temporal dynamics of urban ecosystem services in Latin America: The case of Bogota (Colombia) and Santiago (Chile). *Ecological Indicators,* 85**,** 1068-1080.

DOBBS, C., KENDAL, D. & NITSCHKE, C. R. 2014. Multiple ecosystem services and disservices of the urban forest establishing their connections with landscape structure and sociodemographics. *Ecological Indicators,* 43**,** 44-55.

DOMINGUES, G. F., SOARES, V. P., LEITE, H. G., FERRAZ, A. S., RIBEIRO, C. A. A. S., LORENZON, A. S., MARCATTI, G. E., TEIXEIRA, T. R., DE CASTRO, N. L. M., MOTA, P. H. S., DE SOUZA, G. S. A., DE MENEZES, S. J. M. D. C., DOS SANTOS, A. R. & DO AMARAL, C. H. 2020. Artificial neural networks on integrated multispectral and SAR data for high-performance prediction of eucalyptus biomass. *Computers and Electronics in Agriculture,* 168**,** 105089.

DONNENFELD, Z., HEDDEN, S. & CROOKES, C. 2018. A delicate balance: Water scarcity in South Africa.

DORNING, M. A., KOCH, J., SHOEMAKER, D. A. & MEENTEMEYER, R. K. 2015. Simulating urbanization scenarios reveals tradeoffs between conservation planning strategies. *Landscape and Urban Planning,* 136**,** 28-39.

DOU, X. & YANG, Y. 2018. Estimating forest carbon fluxes using four different data-driven techniques based on long-term eddy covariance measurements: Model comparison and evaluation. *The Science of the total environment,* 627**,** 78-94.

DOUWES, E., ROUGET, M., DIEDERICHS, N., O'DONOGHUE, S., ROY, K. & ROBERTS, D. Buffelsdraai landfill site community reforestation project. XIV World Forestry Congress, 2015a.

DOUWES, E., ROY, K., DIEDERICHS-MANDER, N., MAVUNDLA, K. & ROBERTS, D. 2015b. The Buffelsdraai Landfill Site Community Reforestation Project: Leading the way in community ecosystem-based adaptation to climate change. eThekwini Municipality. *South Africa*.

DRUSCH, M., DEL BELLO, U., CARLIER, S., COLIN, O., FERNANDEZ, V., GASCON, F., HOERSCH, B., ISOLA, C., LABERINTI, P. & MARTIMORT, P. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment,* 120**,** 25-36.

DUBE, T., GARA, T. W., MUTANGA, O., SIBANDA, M., SHOKO, C., MURWIRA, A., MASOCHA, M., NDAIMANI, H. & HATENDI, C. M. 2018. Estimating forest standing biomass in savanna woodlands as an indicator of forest productivity using the new generation WorldView-2 sensor. *Geocarto International,* 33**,** 178-188.

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 Journal of Photogrammetry and Remote Sensing,* 101**,** 36-46.

DUBE, T. & MUTANGA, O. 2016. The impact of integrating WorldView-2 sensor and environmental variables in estimating plantation forest species aboveground biomass and carbon stocks in uMgeni Catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing,* 119**,** 415-425.

DUBE, T., MUTANGA, O., ADAM, E. & ISMAIL, R. 2014a. Intra-and-inter species biomass prediction in a plantation forest: testing the utility of high spatial resolution spaceborne multispectral rapideye sensor and advanced machine learning algorithms. *Sensors,* 14**,** 15348-15370.

DUBE, T., MUTANGA, O., ELHADI, A. & ISMAIL, R. 2014b. Intra-and-inter species biomass prediction in a plantation forest: testing the utility of high spatial resolution spaceborne multispectral rapideye sensor and advanced machine learning algorithms. *Sensors,* 14**,** 15348-15370.

DUBE, T., MUTANGA, O. & ISMAIL, R. 2016. Quantifying aboveground biomass in African environments: A review of the trade-offs between sensor estimation accuracy and costs. *Tropical Ecology,* 57**,** 393-405.

DUCHESNE, L., HOULE, D., OUIMET, R., LAMBERT, M.-C. & LOGAN, T. 2016. Aboveground carbon in Quebec forests: stock quantification at the provincial scale and assessment of temperature, precipitation and edaphic properties effects on the potential stand-level stocking. *PeerJ,* 4.

ECKERT, S. 2012. Improved forest biomass and carbon estimations using texture measures from WorldView-2 satellite data. *Remote sensing,* 4**,** 810-829.

EHDI, AKHSHANDEHROO, OHD, OHARI, USOF, OOZBEH, RABI, ASUL & AHANDARFARD. STRATEGIES TO IMPROVE SUSTAINABILITY IN URBAN LANDSCAPE. 2016.
ENDSLEY, K. A. 2018. 9.09 - Remote Sensing of Socio-Ecological Dynamics in Urban Neighborhoods. *In:* LIANG, S. (ed.) *Comprehensive Remote Sensing.* Oxford: Elsevier.
ESCOBEDO, F. J., KROEGER, T. & WAGNER, J. E. 2011. Urban forests and pollution mitigation: Analyzing ecosystem services and disservices. *Environmental Pollution,* 159**,** 2078-2087.
ESKANDARI, S. & SARAB, S. A. M. 2022. Mapping land cover and forest density in Zagros forests of Khuzestan province in Iran: A study based on Sentinel-2, Google Earth and field data. *Ecological Informatics,* 70**,** 101727.
ESTORNELL, J., MARTÍ-GAVILÁ, J. M., SEBASTIÁ, M. T. & MENGUAL, J. 2013. Principal component analysis applied to remote sensing. *Modelling in Science Education and Learning,* 6**,** 83-89.
EVANGELIDES, C. & NOBAJAS, A. 2020. Red-Edge Normalised Difference Vegetation Index (NDVI705) from Sentinel-2 imagery to assess post-fire regeneration. *Remote Sensing Applications: Society and Environment,* 17**,** 100283.
FANG, J., CHEN, A., PENG, C., ZHAO, S. & CI, L. 2001. Changes in forest biomass carbon storage in China between 1949 and 1998. *Science,* 292**,** 2320-2322.
FAO. Southern Africa's forests and people : Investing in a sustainable future successes, challanges and ways forward WFC 2015 XIV World Forestry Congress 2015 Durban, South Africa United Nations Food and Agriculture Organisation
FARARODA, R., REDDY, R. S., RAJASHEKAR, G., CHAND, T. K., JHA, C. & DADHWAL, V. 2021a. Improving forest above ground biomass estimates over Indian forests using multi source data sets with machine learning algorithm. *Ecological Informatics,* 65**,** 101392.
FARARODA, R., REDDY, R. S., RAJASHEKAR, G., CHAND, T. K., JHA, C. S. & DADHWAL, V. 2021b. Improving forest above ground biomass estimates over Indian forests using multi source data sets with machine learning algorithm. *Ecological Informatics,* 65**,** 101392.
FASSNACHT, F. E., HARTIG, F., LATIFI, H., BERGER, C., HERNÁNDEZ, J., CORVALÁN, P. & KOCH, B. 2014. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sensing of Environment,* 154**,** 102-114.
FASSNACHT, K. S., GOWER, S. T., MACKENZIE, M. D., NORDHEIM, E. V. & LILLESAND, T. M. 1997. Estimating the leaf area index of North Central Wisconsin forests using the landsat thematic mapper. *Remote Sensing of Environment,* 61**,** 229-245.
FENG, Y., ZENG, Z., SEARCHINGER, T. D., ZIEGLER, A. D., WU, J., WANG, D., HE, X., ELSEN, P. R., CIAIS, P. & XU, R. 2022. Doubling of annual forest carbon loss over the tropics during the early twenty-first century. *Nature Sustainability,* 5**,** 444-451.
FICK, S. E. & HIJMANS, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology,* 37**,** 4302-4315.
FINK, H. S. 2016. Human-Nature for Climate Action: Nature-Based Solutions for Urban Sustainability. *Sustainability,* 8**,** 254.
FORKUOR, G., HOUNKPATIN, O. K., WELP, G. & THIEL, M. 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PloS one,* 12**,** e0170478.
FRAZIER, A. E. & HEMINGWAY, B. L. 2021. A technical review of planet smallsat data: Practical considerations for processing and using planetscope imagery. *Remote Sensing,* 13**,** 3930.
FU, Y., HE, H. S., HAWBAKER, T. J., HENNE, P. D., ZHU, Z. & LARSEN, D. R. 2019. Evaluating k-Nearest Neighbor (k NN) Imputation Models for Species-Level Aboveground Forest Biomass Mapping in Northeast China. *Remote sensing,* 11**,** 2005.

GAO, F., XIN, X., SONG, J., LI, X., ZHANG, L., ZHANG, Y. & LIU, J. 2023. Simulation of LUCC Dynamics and Estimation of Carbon Stock under Different SSP-RCP Scenarios in Heilongjiang Province. *Land,* 12**,** 1665.

GAO, Y., FU, J. S., DRAKE, J. B., LAMARQUE, J. F. & LIU, Y. 2013. The impact of emission and climate change on ozone in the United States under representative concentration pathways (RCPs). *Atmospheric chemistry and physics,* 13 18**,** 9607-9621.

GAO, Y., LU, D., LI, G., WANG, G., CHEN, Q., LIU, L. & LI, D. 2018. Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region. *Remote Sensing,* 10**,** 627.

GARA, T. W., MURWIRA, A. & NDAIMANI, H. 2016. Predicting forest carbon stocks from high resolution satellite data in dry forests of Zimbabwe: exploring the effect of the red-edge band in forest carbon stocks estimation. *Geocarto international,* 31**,** 176-192.

GARCÍA-GUTIÉRREZ, J., GONZÁLEZ-FERREIRO, E., MATEOS-GARCÍA, D. & RIQUELME-SANTOS, J. C. A preliminary study of the suitability of deep learning to improve LiDAR-derived biomass estimation. International Conference on Hybrid Artificial Intelligence Systems, 2016. Springer, 588-596.

GARCÍA, M., SAATCHI, S., USTIN, S. & BALZTER, H. 2018. Modelling forest canopy height by integrating airborne LiDAR samples with satellite Radar and multispectral imagery. *International Journal of Applied Earth Observation and Geoinformation,* 66**,** 159-173.

GAŠPAROVIĆ, M. & DOBRINIĆ, D. 2020. Comparative assessment of machine learning methods for urban vegetation mapping using multitemporal sentinel-1 imagery. *Remote Sensing,* 12**,** 1952.

GHARAIBEH, A. A., SHAAMALA, A., OBEIDAT, R. & AL-KOFAHI, S. D. 2020. Improving land-use change modeling by integrating ANN with Cellular Automata-Markov Chain model. *Heliyon,* 6.

GHORBANIAN, A., AHMADI, S. A., AMANI, M., MOHAMMADZADEH, A. & JAMALI, S. 2022. Application of Artificial Neural Networks for Mangrove Mapping Using Multi-Temporal and Multi-Source Remote Sensing Imagery. *Water,* 14**,** 244.

GHOSH, S. M. & BEHERA, M. D. 2018. Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. *Applied Geography,* 96**,** 29-40.

GITELSON, A. A., KAUFMAN, Y. J. & MERZLYAK, M. N. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment,* 58**,** 289-298.

GLEASON, C. J. & IM, J. 2012. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sensing of Environment,* 125**,** 80-91.

GLENDAY, J. 2007. Carbon storage and sequestration analysis for the eThekwini environmental services management plan open space system. *Durban (South Africa): eThekwini Municipality Environmental Management Department*.

GOETZ, S. & DUBAYAH, R. 2011. Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Management,* 2**,** 231-244.

GOETZ, S. J., BACCINI, A., LAPORTE, N. T., JOHNS, T., WALKER, W., KELLNDORFER, J., HOUGHTON, R. A. & SUN, M. 2009. Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. *Carbon balance and management,* 4**,** 1-7.

GRANIERI, D., CHIODINI, G., AVINO, R. & CALIRO, S. 2014. Carbon dioxide emission and heat release estimation for Pantelleria Island (Sicily, Italy). *Journal of volcanology and geothermal research,* 275**,** 22-33.

GRENIER, M., LANTZ, N. J., SOULARD, F. & WANG, J. M.-P. 2020. The use of combined Landsat and Radarsat data for urban ecosystem accounting in Canada. *Statistical journal of the IAOS,* 36**,** 823-839.

GRIMM, N. B., FAETH, S. H., GOLUBIEWSKI, N. E., REDMAN, C. L., WU, J., BAI, X. & BRIGGS, J. M. 2008. Global change and the ecology of cities. *science,* 319**,** 756-760.

GRINAND, C., LE MAIRE, G., VIEILLEDENT, G., RAZAKAMANARIVO, H., RAZAFIMBELO, T. & BERNOUX, M. 2017. Estimating temporal changes in soil carbon stocks at ecoregional scale in Madagascar using remote-sensing. *International journal of applied earth observation and geoinformation,* 54**,** 1-14.

GUERRA-HERNÁNDEZ, J., NARINE, L. L., PASCUAL, A., GONZALEZ-FERREIRO, E., BOTEQUIM, B., MALAMBO, L., NEUENSCHWANDER, A., POPESCU, S. C. & GODINHO, S. 2022. Aboveground biomass mapping by integrating ICESat-2, SENTINEL-1, SENTINEL-2, ALOS2/PALSAR2, and topographic information in Mediterranean forests. *GIScience & Remote Sensing,* 59**,** 1509-1533.

GÜNLÜ, A. & ERCANLı, İ. 2020. Artificial neural network models by ALOS PALSAR data for aboveground stand carbon predictions of pure beech stands: a case study from northern of Turkey. *Geocarto International,* 35**,** 17-28.

GYAMFI-AMPADU, E. & GEBRESLASIE, M. 2021. Two Decades Progress on the Application of Remote Sensing for Monitoring Tropical and Sub-Tropical Natural Forests: A Review. *Forests,* 12**,** 739.

GYAMFI-AMPADU, E., GEBRESLASIE, M. & MENDOZA-PONCE, A. 2021. Evaluating multi-sensors spectral and spatial resolutions for tree species diversity prediction. *Remote Sensing,* 13**,** 1033.

HAJIMA, T., WATANABE, M., YAMAMOTO, A., TATEBE, H., NOGUCHI, M. A., ABE, M., OHGAITO, R., ITO, A., YAMAZAKI, D., OKAJIMA, H., ITO, A., TAKATA, K., OGOCHI, K., WATANABE, S. & KAWAMIYA, M. 2020. Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*.

HALL, M. A. 1999. *Correlation-based feature selection for machine learning.* The University of Waikato.

HALME, E., PELLIKKA, P. & MOTTUS, M. 2019a. Utility of hyperspectral compared to multispectral remote sensing data in estimating forest biomass and structure variables in Finnish boreal forest. *INTERNATIONAL JOURNAL OF APPLIED EARTH OBSERVATION AND GEOINFORMATION,* 83.

HALME, E., PELLIKKA, P. & MÕTTUS, M. 2019b. Utility of hyperspectral compared to multispectral remote sensing data in estimating forest biomass and structure variables in Finnish boreal forest. *International Journal of Applied Earth Observation and Geoinformation,* 83**,** 101942.

HAME, T., RAUSTE, Y., ANTROPOV, O., AHOLA, H. A. & KILPI, J. 2013. Improved mapping of tropical forests with optical and SAR imagery, Part II: Above ground biomass estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 6**,** 92-101.

HAMIDA, A. B., BENOIT, A., LAMBERT, P. & AMAR, C. B. 2018. 3-D deep learning approach for remote sensing image classification. *IEEE Transactions on geoscience and remote sensing,* 56**,** 4420-4434.

HAN, B., MA, X., ZHAO, R., ZHANG, J., WEI, X., LIU, X., LIU, X., ZHANG, C., TAN, C., JIANG, Y. & CHEN, Y. 2012. Development and experimental test of support vector machines virtual screening method for searching Src inhibitors from large compound libraries. *Chemistry Central Journal,* 6**,** 139.

HAN, R., LIU, P., WANG, H., YANG, L., ZHANG, H. & MA, C. 2017. An Improved Urban Mapping Strategy Based on Collaborative Processing of Optical and SAR Remotely Sensed Data. *Mathematical Problems in Engineering,* 2017**,** 1-9.

HANIF, I. 2018. Impact of economic growth, nonrenewable and renewable energy consumption, and urbanization on carbon emissions in Sub-Saharan Africa. *Environmental Science and Pollution Research,* 25**,** 15057-15067.

HARALICK, R. M., SHANMUGAM, K. & DINSTEIN, I. H. 1973. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics***,** 610-621.

HE, C., LIU, Z., TIAN, J. & MA, Q. 2014. Urban expansion dynamics and natural habitat loss in China: A multiscale landscape perspective. *Global change biology,* 20**,** 2886-2902.

HEDBLOM, M., ANDERSSON, E. & BORGSTRÖM, S. 2017. Flexible land-use and undefined governance: From threats to potentials in peri-urban landscape planning. *Land Use Policy,* 63**,** 523-527.
HEENKENDA, M. K., JOYCE, K. E., MAIER, S. W. & BARTOLO, R. 2014. Mangrove species identification: Comparing WorldView-2 with aerial photographs. *Remote Sensing,* 6**,** 6064-6088.
HEISKANEN, J., PELLIKKA, P., BETEMARIAM, E. A. & PACKALEN, P. 2013. Field measurement guidelines for aboveground biomass and fuel wood stocks. *Building Biocarbon and Rural Development in West Africa (BIODEV); World Agroforestry Centre: Nairobi, Kenya*.
HERNÁNDEZ-CLEMENTE, R., NAVARRO-CERRILLO, R. M. & ZARCO-TEJADA, P. J. 2012. Carotenoid content estimation in a heterogeneous conifer forest using narrow-band indices and PROSPECT + DART simulations. *Remote Sensing of Environment,* 127**,** 298-315.
HERNÁNDEZ-GUZMÁN, R., RUIZ-LUNA, A. & GONZÁLEZ, C. 2019. Assessing and modeling the impact of land use and changes in land cover related to carbon storage in a western basin in Mexico. *Remote Sensing Applications: Society and Environment,* 13**,** 318-327.
HLATSHWAYO, S. T., MUTANGA, O., LOTTERING, R. T., KIALA, Z. & ISMAIL, R. 2019. Mapping forest aboveground biomass in the reforested Buffelsdraai landfill site using texture combinations computed from SPOT-6 pan-sharpened imagery. *International Journal of Applied Earth Observation and Geoinformation,* 74**,** 65-77.
HOJAS-GASCON, L. & HUGH, E. 2014. Developing methods for monitoring forests degradation in Tanzania using fine spatial resolution RapidEye data. *Joint Research Centre, European Commission, Ispra, Italy*.
HONG, C., BURNEY, J. A., PONGRATZ, J., NABEL, J. E., MUELLER, N. D., JACKSON, R. B. & DAVIS, S. J. 2021. Global and regional drivers of land-use emissions in 1961–2017. *Nature,* 589**,** 554-561.
HOSSEINI, Z., LATIFI, H., NAGHAVI, H., BAKHTIARI, S. B. & FASSNACHT, F. E. 2021. Influence of plot and sample sizes on aboveground biomass estimations in plantation forests using very high resolution stereo satellite imagery. *Forestry,* 94**,** 278-291.
HU, L., GRIFFITH, D. A. & CHUN, Y. 2020. Impacts of spatial autocorrelation in georeferenced beta and multinomial random variables. *Geographical analysis,* 52**,** 278-298.
HUANG, Q., ROBINSON, D. T. & PARKER, D. C. 2014. Quantifying spatial–temporal change in land-cover and carbon storage among exurban residential parcels. *Landscape Ecology,* 29**,** 275-291.
HUANG, W., DOLAN, K., SWATANTRAN, A., JOHNSON, K., TANG, H., O'NEIL-DUNNE, J., DUBAYAH, R. & HURTT, G. 2019. High-resolution mapping of aboveground biomass for forest carbon monitoring system in the Tri-State region of Maryland, Pennsylvania and Delaware, USA. *Environmental Research Letters,* 14**,** 095002.
HUANG, W., LI, W., XU, J., MA, X., LI, C. & LIU, C. 2022a. Hyperspectral Monitoring Driven by Machine Learning Methods for Grassland Above-Ground Biomass. *Remote Sensing,* 14**,** 2086.
HUANG, Z., TIAN, Y., ZHANG, Q., HUANG, Y., LIU, R., HUANG, H., ZHOU, G., WANG, J., TAO, J. & YANG, Y. 2022b. Estimating mangrove above-ground biomass at Maowei Sea, Beibu Gulf of China using machine learning algorithm with Sentinel-1 and Sentinel-2 data. *Geocarto International***,** 1-28.
HUETE, A. R. 1988. A soil-adjusted vegetation index (SAVI). *Remote sensing of environment,* 25**,** 295-309.
HUTYRA, L. R., DUREN, R., GURNEY, K. R., GRIMM, N., KORT, E. A., LARSON, E. & SHRESTHA, G. 2014. Urbanization and the carbon cycle: Current capabilities and research outlook from the natural sciences perspective. *Earth's Future,* 2**,** 473-495.
IEA 2008. World energy outlook. Washington dc, usa IEA.

IENCO, D., INTERDONATO, R., GAETANO, R. & MINH, D. H. T. 2019. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS Journal of Photogrammetry and Remote Sensing,* 158**,** 11-22.

IMRAN, A., KHAN, K., ALI, N., AHMAD, N., ALI, A. & SHAH, K. 2020. Narrow band based and broadband derived vegetation indices using Sentinel-2 Imagery to estimate vegetation biomass. *Global Journal of Environmental Science and Management,* 6**,** 97-108.

IPCC-GPG 2004. I nte rgovernmental Panel on Climate Change. Good Practice Guidance for Land Use, Land-Use Change and Fore st ry. Currently (January 2004) subject to Final Copyedit. IPCC National Greenhouse ….

IPCC 2006. 2006 IPCC guidelines for national greenhouse gas inventories. *Institute for Global Environmental Strategies, Hayama, Kanagawa, Japan*.

IPCC 2019. Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems *In:* P.R. SHUKLA, J. S., E. CALVO BUENDIA, V. MASSON-DELMOTTE, H.-O. PÖRTNER, D. C. ROBERTS, P. ZHAI, R. SLADE, S. CONNORS, R. VAN DIEMEN, M. FERRAT, E. HAUGHEY, S. LUZ, S. NEOGI, M. PATHAK, J. PETZOLD, J. PORTUGAL PEREIRA, P. VYAS, E. HUNTLEY, K. KISSICK, M. BELKACEMI, J. MALLEY, (ed.). IPCC.

JACOBSEN, K. Analysis of SRTM elevation models. EARSeL 3D-Remote Sensing Workshop, Porto, 2005.

JACON, A. D., GALVÃO, L. S., DALAGNOL, R. & DOS SANTOS, J. R. 2021. Aboveground biomass estimates over Brazilian savannas using hyperspectral metrics and machine learning models: experiences with Hyperion/EO-1. *GIScience & Remote Sensing***,** 1-18.

JAFARZADEH, H., MAHDIANPARI, M., GILL, E., MOHAMMADIMANESH, F. & HOMAYOUNI, S. 2021. Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: a comparative evaluation. *Remote Sensing,* 13**,** 4405.

JHAMTANI, A., MEHTA, R. & SINGH, S. 2021. Size of wallet estimation: application of K-nearest neighbour and quantile regression. *IIMB Management Review,* 33**,** 184-190.

JIANG, F., KUTIA, M., MA, K., CHEN, S., LONG, J. & SUN, H. 2021. Estimating the aboveground biomass of coniferous forest in Northeast China using spectral variables, land surface temperature and soil moisture. *Science of The Total Environment,* 785**,** 147335.

JIANG, Z., HUETE, A. R., DIDAN, K. & MIURA, T. 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote sensing of Environment,* 112**,** 3833-3845.

JIANG, Z., HUETE, A. R., KIM, Y. & DIDAN, K. 2-band enhanced vegetation index without a blue band and its application to AVHRR data. Remote Sensing and Modeling of Ecosystems for Sustainability IV, 2007. SPIE, 45-53.

JINDAL, R., SWALLOW, B. & KERR, J. Forestry-based carbon sequestration projects in Africa: Potential benefits and challenges. Natural Resources Forum, 2008. Wiley Online Library, 116-130.

JUNG-ROTHENHÄUSLER, F., WEICHELT, H. & PACH, M. Rapideye-A novel approach to space borne geo-information solutions. ISPRS Hannover Workshop, 2007. 4-7.

JUTZ, S. & MILAGRO-PÉREZ, M. 2020. Copernicus: the European Earth Observation programme. *Revista de Teledetección***,** V-XI.

KANEVSKI, M. F., POZDNUKHOV, A. & TIMONIN, V. Machine Learning Algorithms for GeoSpatial Data. Applications and Software Tools. 2008.

KANOWSKI, J. & CATTERALL, C. P. 2010. Carbon stocks in above-ground biomass of monoculture plantations, mixed species plantations and environmental restoration plantings in north-east Australia. Wiley Online Library.

KEMPENEERS, P., SEDANO, F., SEEBACH, L., STROBL, P. & SAN-MIGUEL-AYANZ, J. 2011. Data fusion of different spatial resolution remote sensing images applied to

forest-type mapping. *IEEE Transactions on Geoscience and Remote Sensing,* 49**,** 4977-4986.
KHAN, K., IQBAL, J., ALI, A. & KHAN, S. 2020a. Assessment of sentinel-2-derived vegetation indices for the estimation of above-ground biomass/carbon stock, temporal deforestation and carbon emissions estimation in the moist temperate forests of pakistan. *Appl. Ecol. Environ. Res,* 18**,** 783-815.
KHAN, K., IQBAL, J., ALI, A. & KHAN, S. N. 2020b. ASSESSMENT OF SENTINEL-2-DERIVED VEGETATION INDICES FOR THE ESTIMATION OF ABOVE-GROUND BIOMASS/CARBON STOCK, TEMPORAL DEFORESTATION AND CARBON EMISSIONS ESTIMATION IN THE MOIST TEMPERATE FORESTS OF PAKISTAN. *Applied Ecology and Environmental Research,* 18**,** 783-815.
KIM, B., LEE, D.-E., HU, G., NATARAJAN, Y., PREETHAA, S. & RATHINAKUMAR, A. P. 2022. Ensemble machine learning-based approach for predicting of frp–concrete interfacial bonding. *Mathematics,* 10**,** 231.
KOEDSIN, W. & VAIPHASA, C. 2013. Discrimination of tropical mangroves at the species level with EO-1 Hyperion data. *Remote Sensing,* 5**,** 3562-3582.
KOENIG, W. D. 1999. Spatial autocorrelation of ecological phenomena. *Trends in Ecology & Evolution,* 14**,** 22-26.
KOLDASBAYEVA, D., TREGUBOVA, P., GASANOV, M., ZAYTSEV, A., PETROVSKAIA, A. & BURNAEV, E. V. 2023. Challenges in data-based geospatial modeling for environmental research and practice. *ArXiv,* abs/2311.11057.
KOWARIK, I., HILLER, A., PLANCHUELO, G., SEITZ, B., VON DER LIPPE, M. & BUCHHOLZ, S. 2019. Emerging Urban Forests: Opportunities for Promoting the Wild Side of the Urban Green Infrastructure. *Sustainability*.
KOWE, P., MUTANGA, O. & DUBE, T. 2021. Advancements in the remote sensing of landscape pattern of urban green spaces and vegetation fragmentation. *International Journal of Remote Sensing,* 42**,** 3797-3832.
KRISCHKE, M., NIEMEYER, W. & SCHERER, S. 2000. RapidEye satellite based geo-information system. *Acta Astronautica,* 46**,** 307-312.
KROSS, A., MCNAIRN, H., LAPEN, D., SUNOHARA, M. & CHAMPAGNE, C. 2015. Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *International Journal of Applied Earth Observation and Geoinformation,* 34**,** 235-248.
KRUK, M. 2023. Prediction of environmental factors responsible for chlorophyll a-induced hypereutrophy using explainable machine learning. *Ecological Informatics,* 75**,** 102005.
KUMAR, M., KUMAR, R., BISHNOI, P., SIHAG, V., BISHNOI, R., RANI, S., SINDHU, P., BUDHWAR, S., KUMAR, P., SHARMA, S., SHARMA, P., SHARMA, R., PANDEY, V., DAHIYA, M., ARYA, V. S., SINGH, T. P. & KUMAR, V. 2021. A geo-spatial approach to assess Trees outside Forest (ToF) in Haryana State, India. *Land Degradation & Development,* 32**,** 3588-3597.
KUMAR, M. K. & SHIVA NAGENDRA, S. M. 2016. Quantification of anthropogenic CO2 emissions in a tropical urban environment. *Atmospheric Environment,* 125**,** 272-282.
KURNIATI, F. T., MANONGGA, D. H., SEDIYONO, E., PRASETYO, S. Y. J. & HUIZEN, R. R. 2024. GLCM-Based Feature Combination for Extraction Model Optimization in Object Detection Using Machine Learning. *arXiv preprint arXiv:2404.04578*.
KURSA, M. B. & RUDNICKI, W. R. 2010. Feature selection with the Boruta package. *Journal of statistical software,* 36**,** 1-13.
KUYPER, J., SCHROEDER, H. & LINNÉR, B.-O. 2018. The Evolution of the UNFCCC. *Annual Review of Environment and Resources,* 43**,** 343-368.
LABRECQUE, S., FOURNIER, R. A., LUTHER, J. E. & PIERCEY, D. 2006. A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland. *Forest Ecology and Management,* 226**,** 129-144.
LANDGREBE, D. 2002. Hyperspectral image data analysis. *IEEE Signal processing magazine,* 19**,** 17-28.
LANG, M., MCCARTY, G., OESTERLING, R. & YEO, I.-Y. 2013. Topographic metrics for improved mapping of forested wetlands. *Wetlands,* 33**,** 141-155.

LANGUILLE, F., DECHOZ, C., GAUDEL, A., GRESLOU, D., DE LUSSY, F., TRÉMAS, T. & POULAIN, V. Sentinel-2 geometric image quality commissioning: first results. SPIE Remote Sensing, 2015.
LATIFAH, S., PURWOKO, A., HARTINI, K. S. & FACHRUDIN, K. A. Allometric models to estimate the aboveground biomass of forest: A literature review. IOP Conference Series: Materials Science and Engineering, 2021. IOP Publishing, 012047.
LAURANCE, W. F. 2010. Habitat destruction: death by a thousand cuts. *Conservation biology for all,* 1**,** 73-88.
LAURIN, G. V., BALLING, J., CORONA, P., MATTIOLI, W., PAPALE, D., PULETTI, N., RIZZO, M., TRUCKENBRODT, J. & URBAN, M. 2018a. Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data. *Journal of Applied Remote Sensing,* 12**,** 016008-016008.
LAURIN, G. V., BALLING, J., CORONA, P., MATTIOLI, W., PAPALE, D., PULETTI, N., RIZZO, M., TRUCKENBRODT, J. & URBAN, M. 2018b. Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data. *Journal of Applied Remote Sensing,* 12**,** 016008.
LEE, D.-H., KIL, S.-H., JO, H.-K. & CHOI, B. 2019. Spatial distributions of carbon storage and uptake of urban forests in Seoul, South Korea. *Sensors and Materials,* 31**,** 3811-3826.
LEWIS, T., VERSTRATEN, L., HOGG, B., WEHR, B. J., SWIFT, S., TINDALE, N., MENZIES, N. W., DALAL, R. C., BRYANT, P. & FRANCIS, B. 2019. Reforestation of agricultural land in the tropics: The relative contribution of soil, living biomass and debris pools to carbon sequestration. *Science of the Total Environment,* 649**,** 1502-1513.
LI, LI, M., LI, C. & LIU, Z. 2020a. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Scientific reports,* 10**,** 1-12.
LI, B., CHEN, D., WU, S., ZHOU, S., WANG, T. & CHEN, H. 2016a. Spatio-temporal assessment of urbanization impacts on ecosystem services: Case study of Nanjing City, China. *Ecological Indicators,* 71**,** 416-427.
LI, C., LI, M. Y., LI, Y. C. & QIAN, P. 2020b. Estimating aboveground forest carbon density using Landsat 8 and field-based data: a comparison of modelling approaches. *International Journal of Remote Sensing,* 41**,** 4269-4292.
LI, D., CHEN, J. M., YU, W., ZHENG, H., YAO, X., ZHU, Y., CAO, W. & CHENG, T. 2024a. A chlorophyll-constrained semi-empirical model for estimating leaf area index using a red-edge vegetation index. *Computers and Electronics in Agriculture,* 220**,** 108891.
LI, H., ZHANG, G., ZHONG, Q., XING, L. & DU, H. 2023. Prediction of Urban Forest Aboveground Carbon Using Machine Learning Based on Landsat 8 and Sentinel-2: A Case Study of Shanghai, China. *Remote Sensing,* 15**,** 284.
LI, L., FU, W. & LUO, M. 2022. Spatial and Temporal Variation and Prediction of Ecosystem Carbon Stocks in Yunnan Province Based on Land Use Change. *International Journal of Environmental Research and Public Health,* 19**,** 16059.
LI, N., HUO, L. & ZHANG, X. 2024b. Using only the red-edge bands is sufficient to detect tree stress: A case study on the early detection of PWD using hyperspectral drone images. *Computers and Electronics in Agriculture,* 217**,** 108665.
LI, P., ZHU, J., HU, H., GUO, Z., PAN, Y., BIRDSEY, R. & FANG, J. 2016b. The relative contributions of forest growth and areal expansion to forest biomass carbon. *Biogeosciences,* 13**,** 375-388.
LI, W., NIU, Z., WANG, C., HUANG, W., CHEN, H., GAO, S., LI, D. & MUHAMMAD, S. 2015. Combined Use of Airborne LiDAR and Satellite GF-1 Data to Estimate Leaf Area Index, Height, and Aboveground Biomass of Maize During Peak Growing Season. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 8**,** 4489-4501.
LI, X., CHEN, W. Y., SANESI, G. & LAFORTEZZA, R. 2019a. Remote Sensing in Urban Forestry: Recent Applications and Future Directions. *Remote Sensing,* 11**,** 1144.

LI, X., ZHANG, M., LONG, J. & LIN, H. 2021. A Novel Method for Estimating Spatial Distribution of Forest Above-Ground Biomass Based on Multispectral Fusion Data and Ensemble Learning Algorithm. *Remote Sensing,* 13**,** 3910.
LI, Y., LI, C., LI, M. & LIU, Z. 2019b. Influence of variable selection and forest type on forest aboveground biomass estimation using machine learning algorithms. *Forests,* 10**,** 1073.
LI, Y., LI, M., LI, C. & LIU, Z. 2020c. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Scientific reports,* 10**,** 1-12.
LIBEN-NOWELL, D., SHARP, A., WEXLER, T. & WOODS, K. M. Computing Shapley Value in Supermodular Coalitional Games. International Computing and Combinatorics Conference, 2012.
LIN, D., LAI, J., MULLER-LANDAU, H. C., MI, X. & MA, K. 2012. Topographic variation in aboveground biomass in a subtropical evergreen broad-leaved forest in China. *PloS one,* 7**,** e48244.
LIN, Q. Enhanced vegetation index using moderate resolution imaging spectroradiometers. 2012 5th International Congress on Image and Signal Processing, 2012. IEEE, 1043-1046.
LINDSELL, J. A. & KLOP, E. 2013. Spatial and temporal variation of carbon stocks in a lowland tropical forest in West Africa. *Forest Ecology and Management,* 289**,** 10-17.
LIU, N. & NAN, H. 2018. Carbon stocks of three secondary coniferous forests along an altitudinal gradient on Loess Plateau in inland China. *PLoS One,* 13**,** e0196927.
LIU, W., ZHANG, X. & MA, J. Application of remote-sensing technology to forest carbon storage estimate. 2011 International Conference on Remote Sensing, Environment and Transportation Engineering, 2011. IEEE, 3858-3860.
LIU, X., TIAN, Z. & CHEN, C. 2021. Total organic carbon content prediction in lacustrine shale using extreme gradient boosting machine learning based on bayesian optimization. *Geofluids,* 2021**,** 1-18.
LIU, Y., GONG, W., XING, Y., HU, X. & GONG, J. 2019. Estimation of the forest stand mean height and aboveground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery. *ISPRS Journal of Photogrammetry and Remote Sensing,* 151**,** 277-289.
LOPATIN, J., KATTENBORN, T., GALLEGUILLOS, M., PÉREZ-QUEZADA, J. F. & SCHMIDTLEIN, S. 2019. Using aboveground vegetation attributes as proxies for mapping peatland belowground carbon stocks. *Remote Sensing of Environment*.
LÓPEZ-SERRANO, P. M., CÁRDENAS DOMÍNGUEZ, J. L., CORRAL-RIVAS, J. J., JIMÉNEZ, E., LÓPEZ-SÁNCHEZ, C. A. & VEGA-NIEVA, D. J. 2020. Modeling of aboveground biomass with Landsat 8 OLI and machine learning in temperate forests. *Forests,* 11**,** 11.
LÓPEZ-SERRANO, P. M., CORRAL-RIVAS, J. J., DÍAZ-VARELA, R. A., ÁLVAREZ-GONZÁLEZ, J. G. & LÓPEZ-SÁNCHEZ, C. A. 2016a. Evaluation of radiometric and atmospheric correction algorithms for aboveground forest biomass estimation using Landsat 5 TM data. *Remote sensing,* 8**,** 369.
LÓPEZ-SERRANO, P. M., LÓPEZ-SÁNCHEZ, C. A., ALVAREZ-GONZALEZ, J. G. & GARCIA-GUTIERREZ, J. 2016b. A comparison of machine learning techniques applied to Landsat-5 TM spectral data for biomass estimation. *Canadian Journal of Remote Sensing,* 42**,** 690-705.
LOURENÇO, P. 2021. Biomass Estimation Using Satellite-Based Data. *Forest Biomass-From Trees to Energy*. IntechOpen.
LOVELL, H., SALES DE AGUIAR, T., BEBBINGTON, J. & LARRINAGA, C. 2010. Accounting for carbon. *ACCA Research Report,* 122.
LU, D., CHEN, Q., WANG, G., LIU, L., LI, G. & MORAN, E. 2016. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth,* 9**,** 63-105.
LU, D. & JIANG, X. 2024. A brief overview and perspective of using airborne Lidar data for forest biomass estimation. *International Journal of Image and Data Fusion***,** 1-24.

LU, X. M., ZHENG, G., MILLER, C. & ALVARADO, E. 2017. Combining Multi-Source Remotely Sensed Data and a Process-Based Model for Forest Aboveground Biomass Updating. *Sensors,* 17**,** 21.
LUCKE, T. & BEECHAM, S. 2019. An infiltration approach to reducing pavement damage by street trees. *Science of the Total Environment,* 671**,** 94-100.
LULUCF, G. 2004. I nte rgovernmental Panel on Climate Change. Good Practice Guidance for Land Use, Land-Use Change and Fore st ry. Currently (January 2004) subject to Final Copyedit. IPCC National Greenhouse ….
LUO, K., WEI, Y., DU, J., LIU, L., LUO, X., SHI, Y., PEI, X., LEI, N., SONG, C., LI, J. & TANG, X. 2022. Machine learning-based estimates of aboveground biomass of subalpine forests using Landsat 8 OLI and Sentinel-2B images in the Jiuzhaigou National Nature Reserve, Eastern Tibet Plateau. *Journal of Forestry Research,* 33**,** 1329-1340.
LUO, M., WANG, Y., XIE, Y., ZHOU, L., QIAO, J., QIU, S. & SUN, Y. 2021. Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests,* 12**,** 216.
LUO, W., KIM, H. S., ZHAO, X., RYU, D., JUNG, I., CHO, H., HARRIS, N., GHOSH, S., ZHANG, C. & LIANG, J. 2020. New forest biomass carbon stock estimates in Northeast Asia based on multisource data. *Global Change Biology,* 26**,** 7045-7066.
MACDICKEN, K. G. 1997. A guide to monitoring carbon storage in forestry and agroforestry projects.
MACKEY, B. G., MOOMAW, W. R., LINDENMAYER, D. B. & KEITH, H. 2022. Net carbon accounting and reporting are a barrier to understanding the mitigation value of forest protection in developed countries. *Environmental Research Letters,* 17.
MAIN-KNORN, M., MOISEN, G. G., HEALEY, S. P., KEETON, W. S., FREEMAN, E. A. & HOSTERT, P. 2011. Evaluating the remote sensing and inventory-based estimation of biomass in the Western Carpathians. *Remote Sensing,* 3**,** 1427-1446.
MARSHALL, M. & THENKABAIL, P. 2015. Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing,* 108**,** 205-218.
MATIZA, C., MUTANGA, O., ODINDI, J. & MNGADI, M. 2024a. The utility of Planetscope spectral data in quantifying above-ground carbon stock in an urban reforested landscape. *Ecological Informatics,* 80**,** 102472.
MATIZA, C., MUTANGA, O., PEERBHAY, K., ODINDI, J. & LOTTERING, R. 2023a. A systematic review of remote sensing and machine learning approaches for accurate carbon storage estimation in natural forests. *Southern Forests: a Journal of Forest Science,* 85**,** 123-141.
MATIZA, C., MUTANGA, O., PEERBHAY, K., ODINDI, J. & LOTTERING, R. 2023b. A systematic review of remote sensing and machine learning approaches for accurate carbon storage estimation in natural forests. *Southern Forests,* 85**,** 123-141.
MATIZA, C., MUTANGA, O., PEERBHAY, K., ODINDI, J. & LOTTERING, R. 2024b. Assessing above-ground biomass in reforested urban landscapes using machine learning and remotely sensed data. *Journal of Spatial Science***,** 1-28.
MCBRIDE, L. A., HOPE, A. P., CANTY, T. P., BENNETT, B. F., TRIBETT, W. R. & SALAWITCH, R. J. 2021. Comparison of CMIP6 historical climate simulations and future projected warming to an empirical model of global climate. *Earth System Dynamics,* 12**,** 545-579.
MCCARTHY, M. P., BEST, M. J. & BETTS, R. A. 2010. Climate change in cities due to global warming and urban effects. *Geophysical Research Letters,* 37.
MELILLOS, G. & HADJIMITSIS, D. G. Using simple ratio (SR) vegetation index to detect deep man-made infrastructures in Cyprus. Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXV, 2020. SPIE, 105-113.
MEMON, N., PATEL, S. B. & PATEL, D. P. Comparative analysis of artificial neural network and XGBoost algorithm for PolSAR image classification. International Conference on Pattern Recognition and Machine Intelligence, 2019. Springer, 452-460.

MERMOZ, S., LE TOAN, T., VILLARD, L., RÉJOU-MÉCHAIN, M. & SEIFERT-GRANZIN, J. 2014. Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote sensing of environment,* 155**,** 109-119.

MILLER, C. S. 2014. *520,000 years of environmental change in West Africa*. The Open University.

MITCHELL, A. L., ROSENQVIST, A. & MORA, B. 2017. Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. *Carbon balance and management,* 12**,** 1-22.

MITCHELL, M. G., JOHANSEN, K., MARON, M., MCALPINE, C. A., WU, D. & RHODES, J. R. 2018a. Identification of fine scale and landscape scale drivers of urban aboveground carbon stocks using high-resolution modeling and mapping. *Science of the total Environment,* 622**,** 57-70.

MITCHELL, M. G. E., JOHANSEN, K. L., MARON, M., MCALPINE, C. A., WU, D. & RHODES, J. R. 2018b. Identification of fine scale and landscape scale drivers of urban aboveground carbon stocks using high-resolution modeling and mapping. *The Science of the total environment,* 622-623**,** 57-70.

MNGADI, M., ODINDI, J. & MUTANGA, O. 2021a. The utility of Sentinel-2 spectral data in quantifying above-ground carbon stock in an urban reforested landscape. *Remote Sensing,* 13**,** 4281.

MNGADI, M., ODINDI, J. & MUTANGA, O. 2022a. Quantifying Carbon Stock Variability of Species Within a Reforested Urban Landscape Using Texture Measures Derived from Remotely Sensed Imagery. *Advances in Remote Sensing for Forest Monitoring***,** 150-166.

MNGADI, M., ODINDI, J., MUTANGA, O. & SIBANDA, M. 2022b. Estimating aboveground net primary productivity of reforested trees in an urban landscape using biophysical variables and remotely sensed data. *Science of The Total Environment,* 802**,** 149958.

MNGADI, M., ODINDI, J., PEERBHAY, K. & MUTANGA, O. 2021b. Examining the effectiveness of Sentinel-1 and 2 imagery for commercial forest species mapping. *Geocarto International,* 36**,** 1-12.

MOHD ZAKI, N. A. & ABD LATIF, Z. 2017. Carbon sinks and tropical forest biomass estimation: a review on role of remote sensing in aboveground-biomass modelling. *Geocarto International,* 32**,** 701 - 716.

MOHER, D., SHAMSEER, L., CLARKE, M., GHERSI, D., LIBERATI, A., PETTICREW, M., SHEKELLE, P. & STEWART, L. A. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews,* 4**,** 1-9.

MONTESANO, P., COOK, B., SUN, G., SIMARD, M., NELSON, R., RANSON, K., ZHANG, Z. & LUTHCKE, S. 2013. Achieving accuracy requirements for forest biomass mapping: A spaceborne data fusion method for estimating forest biomass and LiDAR sampling error. *Remote Sensing of Environment,* 130**,** 153-170.

MORIN, D., PLANELLS, M., GUYON, D., VILLARD, L., MERMOZ, S., BOUVET, A., THEVENON, H., DEJOUX, J. F., TOAN, T. L. & DEDIEU, G. 2019. Estimation and Mapping of Forest Structure Parameters from Open Access Satellite Images: Development of a Generic Method with a Study Case on Coniferous Plantation. *Remote Sensing,* 11**,** 25.

MORLEY, S. K., BRITO, T. & WELLING, D. T. 2018. Measures of Model Performance Based On the Log Accuracy Ratio. *Space Weather,* 16**,** 69 - 88.

MOSIN, V., AGUILAR, R., PLATONOV, A., VASILIEV, A., KEDROV, A. & IVANOV, A. Remote sensing and machine learning for tree detection and classification in forestry applications. Image and Signal Processing for Remote Sensing XXV, 2019. SPIE, 130-141.

MOUNTRAKIS, G., IM, J. & OGOLE, C. 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66**,** 247-259.

MOURATIDIS, A., BRIOLE, P. & KATSAMBALOS, K. 2010. SRTM 3 ″DEM (versions 1, 2, 3, 4) validation by means of extensive kinematic GPS measurements: a case study from North Greece. *International Journal of Remote Sensing,* 31**,** 6205-6222.
MOVSESSIAN, A., CAVA, D. G. & TCHERNIAK, D. 2021. Interpretable machine learning in damage detection using Shapley Additive Explanations. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*.
MUGABOWINDEKWE, M., BRANDT, M., CHAVE, J., REINER, F., SKOLE, D. L., KARIRYAA, A., IGEL, C., HIERNAUX, P., CIAIS, P. & MERTZ, O. 2022. Nation-wide mapping of tree-level aboveground carbon stocks in Rwanda. *Nature Climate Change***,** 1-7.
MUGWEDI, L., ROUGET, M., EGOH, B. N., SERSHEN, RAMDHANI, S., SLOTOW, R. & RENTERIA, J. L. 2017. An Assessment of a Community-Based, Forest Restoration Programme in Durban (eThekwini), South Africa. *Forests,* 8**,** 255.
MUHE, S. & ARGAW, M. 2021. Modeling Forest Carbon Estimation Using Sentinel-2 Derived Indices in Yayu Afro-Montane Forest, South West Ethiopia.
MULATU, K. A., DECUYPER, M., BREDE, B., KOOISTRA, L., REICHE, J., MORA, B. & HEROLD, M. 2019. Linking Terrestrial LiDAR Scanner and Conventional Forest Structure Measurements with Multi-Modal Satellite Data. *Forests,* 10**,** 291.
MUNYATI, C. 2022. Detecting the distribution of grass aboveground biomass on a rangeland using Sentinel-2 MSI vegetation indices. *Advances in Space Research,* 69**,** 1130-1145.
MUTANGA, O., ADAM, E. & CHO, M. A. 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation,* 18**,** 399-406.
MUUKKONEN, P. 2006. Forest inventory-based large-scale forest biomass and carbon budget assessment: new enhanced methods and use of remote sensing for verification. *Dissertationes Forestales*.
MYEONG, S., NOWAK, D. J. & DUGGIN, M. J. 2006. A temporal analysis of urban forest carbon storage using remote sensing. *Remote Sensing of Environment,* 101**,** 277-282.
NAKAKAAWA, C. A., VEDELD, P. O. & AUNE, J. B. 2011. Spatial and temporal land use and carbon stock changes in Uganda: implications for a future REDD strategy. *Mitigation and Adaptation Strategies for Global Change,* 16**,** 25-62.
NANDY, S. & KUSHWAHA, S. P. S. 2021. Forest Biomass Assessment Integrating Field Inventory and Optical Remote Sensing Data: A Systematic Review. *INTERNATIONAL JOURNAL OF PLANT AND ENVIRONMENT*.
NANDY, S., SINGH, R., GHOSH, S., WATHAM, T., KUSHWAHA, S. P. S., KUMAR, A. S. & DADHWAL, V. K. 2017. Neural network-based modelling for forest biomass assessment. *Carbon Management,* 8**,** 305-317.
NAVARRO-CERRILLO, R. M., DUQUE-LAZO, J., RODRIGUEZ-VALLEJO, C., VARO-MARTINEZ, M. A. & PALACIOS-RODRIGUEZ, G. 2018. Airborne Laser Scanning Cartography of On-Site Carbon Stocks as a Basis for the Silviculture of Pinus Halepensis Plantations. *Remote Sensing,* 10**,** 22.
NAVARRO, A., YOUNG, M., ALLAN, B., CARNELL, P., MACREADIE, P. & IERODIACONOU, D. 2020. The application of Unmanned Aerial Vehicles (UAVs) to estimate above-ground biomass of mangrove ecosystems. *Remote Sensing of Environment,* 242**,** 111747.
NAVARRO, J. A., ALGEET, N., FERNÁNDEZ-LANDA, A., ESTEBAN, J., RODRÍGUEZ-NORIEGA, P. & GUILLÉN-CLIMENT, M. L. 2019. Integration of UAV, Sentinel-1, and Sentinel-2 data for mangrove plantation aboveground biomass monitoring in Senegal. *Remote Sensing,* 11**,** 77.
NAVE, L. E., WALTERS, B. F., HOFMEISTER, K., PERRY, C. H., MISHRA, U., DOMKE, G. M. & SWANSTON, C. 2019. The role of reforestation in carbon sequestration. *New Forests,* 50**,** 115-137.
NEVZATI, F., VELDI, M., STORIE, J. & KÜLVIK, M. 2024. Leveraging Ecosystem Services and Well-Being in Urban Landscape Planning for Nature Conservation: A Case Study of Peri-Urban Dynamics. *Conservation,* 4**,** 1-22.

NGO, T. H., NGUYEN, T. M. C., DUONG, T. G. H. & LY, T. H. 2021. Forest - Related Culture and Contribution to Sustainable Development in the Northern Mountain Region in Vietnam. *Forest and Society*.
NGUYEN, T. T., NGUYEN, V. P., NGUYEN, V. Q. & HOANG, T. P. N. 2022. Applied Machine Learning Algorithms and Landsat 8 for Estimating Aboveground Carbon Stock in Evergreen Broadleaf Forest in Binh Phuoc Province. *VNU Journal of Science: Earth and Environmental Sciences*, 38.
NI-MEISTER, W., LEE, S., STRAHLER, A. H., WOODCOCK, C. E., SCHAAF, C., YAO, T., RANSON, K. J., SUN, G. & BLAIR, J. B. 2010. Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research: Biogeosciences*, 115.
NOACK, A. 2007. Energy models for graph clustering. *J. Graph Algorithms Appl.*, 11**,** 453-480.
NOWAK, D. J., CRANE, D. E. & STEVENS, J. C. 2006. Air pollution removal by urban trees and shrubs in the United States. *Urban forestry & urban greening*, 4**,** 115-123.
NOWAK, D. J., GREENFIELD, E. J., HOEHN, R. E. & LAPOINT, E. 2013. Carbon storage and sequestration by trees in urban and community areas of the United States. *Environmental pollution*, 178**,** 229-236.
NOWAK, D. J., HIRABAYASHI, S., BODINE, A. & GREENFIELD, E. 2014. Tree and forest effects on air quality and human health in the United States. *Environmental Pollution*, 193**,** 119-129.
NRDC. 2018. *Fossil Fuels: The Dirty Facts* [Online]. online: NRDC. Available: https://www.nrdc.org/stories/fossil-fuels-dirty-facts#sec-whatis [Accessed 29 december 2023].
NUTHAMMACHOT, N., PHAIRUANG, W., WICAKSONO, P. & SAYEKTININGSIH, T. 2018. Estimating aboveground biomass on private forest using Sentinel-2 imagery. *Journal of Sensors*, 2018.
ODEBIRI, O., MUTANGA, O. & ODINDI, J. 2022. Deep learning-based national scale soil organic carbon mapping with Sentinel-3 data. *Geoderma*, 411**,** 115695.
ODEBIRI, O., MUTANGA, O., ODINDI, J., NAICKER, R., SLOTOW, R. & MNGADI, M. 2023. Evaluation of projected soil organic carbon stocks under future climate and land cover changes in South Africa using a deep learning approach. *Journal of Environmental Management*, 330**,** 117127.
ODEBIRI, O., MUTANGA, O., ODINDI, J., PEERBHAY, K. & DOVEY, S. 2020a. Predicting soil organic carbon stocks under commercial forest plantations in KwaZulu-Natal province, South Africa using remotely sensed data. *GIScience & remote sensing*, 57**,** 450-463.
ODEBIRI, O., MUTANGA, O., ODINDI, J., PEERBHAY, K., DOVEY, S. & ISMAIL, R. 2020b. Estimating soil organic carbon stocks under commercial forestry using topo-climate variables in KwaZulu-Natal, South Africa. *South African Journal of Science*, 116**,** 1-8.
ODINDI, J. & MHANGARA, P. 2012. Green spaces trends in the city of Port Elizabeth from 1990 to 2000 using remote sensing.
ODINDI, J., MHANGARA, P. & KAKEMBO, V. 2012. Remote sensing land-cover change in Port Elizabeth during South Africa's democratic transition. *South African Journal of Science*, 108**,** 1-7.
OEHMCKE, S., LI, L., TREPEKLI, K., REVENGA, J. C., NORD-LARSEN, T., GIESEKE, F. & IGEL, C. 2024. Deep point cloud regression for above-ground forest biomass estimation from airborne LiDAR. *Remote Sensing of Environment*, 302**,** 113968.
OGUNBODE, T. O. & ASIFAT, J. T. 2021. Sustainability and challenges of climate change mitigation through urban reforestation-a review. *Journal of Forest and Environmental Science*, 37**,** 1-13.
OJOYI, M., MUTANGA, O., ODINDI, J. & ABDEL-RAHMAN, E. M. 2016. Application of topo-edaphic factors and remotely sensed vegetation indices to enhance biomass estimation in a heterogeneous landscape in the Eastern Arc Mountains of Tanzania. *Geocarto International*, 31**,** 1-21.

OTUNGA, C., ODINDI, J., MUTANGA, O. & ADJORLOLO, C. 2019. Evaluating the potential of the red edge channel for C3 (Festuca spp.) grass discrimination using Sentinel-2 and Rapid Eye satellite image data. *Geocarto International,* 34**,** 1123-1143.

PACHECO-PASCAGAZA, A. M., GOU, Y., LOUIS, V., ROBERTS, J. F., RODRÍGUEZ-VEIGA, P., DA CONCEIÇÃO BISPO, P., ESPÍRITO-SANTO, F. D., ROBB, C., UPTON, C. & GALINDO, G. 2022. Near Real-Time Change Detection System Using Sentinel-2 and Machine Learning: A Test for Mexican and Colombian Forests. *Remote Sensing,* 14**,** 707.

PADALIA, H., PRAKASH, A. & WATHAM, T. 2023. Modelling aboveground biomass of a multistage managed forest through synergistic use of Landsat-OLI, ALOS-2 L-band SAR and GEDI metrics. *Ecological Informatics,* 77**,** 102234.

PANDIT, S., TSUYUKI, S. & DUBE, T. 2018. Landscape-scale aboveground biomass estimation in buffer zone community forests of central Nepal: Coupling in situ measurements with Landsat 8 satellite data. *Remote Sensing,* 10**,** 1848.

PANDIT, S., TSUYUKI, S. & DUBE, T. 2020. Exploring the inclusion of Sentinel-2 MSI texture metrics in above-ground biomass estimation in the community forest of Nepal. *Geocarto International,* 35**,** 1832-1849.

PANNAKKONG, W., THIWA-ANONT, K., SINGTHONG, K., PARTHANADEE, P. & BUDDHAKULSOMSIRI, J. 2022. Hyperparameter Tuning of Machine Learning Algorithms Using Response Surface Methodology: A Case Study of ANN, SVM, and DBN. *Mathematical Problems in Engineering*.

PASHER, J., MCGOVERN, M., KHOURY, M. & DUFFE, J. 2014. Assessing carbon storage and sequestration by Canada's urban forests using high resolution earth observation data. *Urban forestry & urban greening,* 13**,** 484-494.

PECHANEC, V., PURKYT, J., BENC, A., NWAOGU, C., ŠTĚRBOVÁ, L. & CUDLÍN, P. 2018. Modelling of the carbon sequestration and its prediction under climate change. *Ecological Informatics,* 47**,** 50-54.

PEERBHAY, K., MUTANGA, O., LOTTERING, R. & ISMAIL, R. 2016. Mapping Solanum mauritianum plant invasions using WorldView-2 imagery and unsupervised random forests. *Remote Sensing of Environment,* 182**,** 39-48.

PEERBHAY, K. Y., MUTANGA, O. & ISMAIL, R. 2013. Investigating the capability of few strategically placed Worldview-2 multispectral bands to discriminate forest species in KwaZulu-Natal, South Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 7**,** 307-316.

PENG, X., ZHAO, A., CHEN, Y., CHEN, Q., LIU, H., WANG, J. & LI, H. 2020. Comparison of Modeling Algorithms for Forest Canopy Structures Based on UAV-LiDAR: A Case Study in Tropical China. *Forests,* 11**,** 1324.

PENMAN, J., GREEN, C., OLOFSSON, P., RAISON, J., WOODCOCK, C., BALZTER, H., BALTUCK, M. & FOODY, G. M. 2016. Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative.

PETTORELLI, N. 2013. *The normalized difference vegetation index,* Oxford University Press.

PETTORELLI, N., VIK, J. O., MYSTERUD, A., GAILLARD, J.-M., TUCKER, C. J. & STENSETH, N. C. 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in ecology & evolution,* 20**,** 503-510.

PHAM, T. D., LE, N. N., HA, N. T., NGUYEN, L. V., XIA, J., YOKOYA, N., TO, T. T., TRINH, H. X., KIEU, L. Q. & TAKEUCHI, W. 2020a. Estimating mangrove above-ground biomass using extreme gradient boosting decision trees algorithm with fused sentinel-2 and ALOS-2 PALSAR-2 data in can Gio biosphere reserve, Vietnam. *Remote Sensing,* 12**,** 777.

PHAM, T. D., YOKOYA, N., BUI, D. T., YOSHINO, K. & FRIESS, D. A. 2019. Remote sensing approaches for monitoring mangrove species, structure, and biomass: Opportunities and challenges. *Remote Sensing,* 11**,** 230.

PHAM, T. D., YOKOYA, N., NGUYEN, T. T. T., LE, N. N., HA, N. T., XIA, J., TAKEUCHI, W. & PHAM, T. D. 2021. Improvement of mangrove soil carbon stocks

estimation in North Vietnam using Sentinel-2 data and machine learning approach. *GIScience & Remote Sensing,* 58**,** 68-87.
PHAM, T. D., YOKOYA, N., XIA, J., HA, N. T., LE, N. N., NGUYEN, T. T. T., DAO, T. H., VU, T. T. P., PHAM, T. D. & TAKEUCHI, W. 2020b. Comparison of Machine Learning Methods for Estimating Mangrove Above-Ground Biomass Using Multiple Source Remote Sensing Data in the Red River Delta Biosphere Reserve, Vietnam. *Remote Sensing,* 12**,** 1334.
PHAM, T. D. & YOSHINO, K. 2017. Aboveground biomass estimation of mangrove species using ALOS-2 PALSAR imagery in Hai Phong City, Vietnam. *Journal of Applied Remote Sensing,* 11**,** 026010.
PHAM, T. D., YOSHINO, K., LE, N. N. & BUI, D. T. 2018. Estimating aboveground biomass of a mangrove plantation on the Northern coast of Vietnam using machine learning techniques with an integration of ALOS-2 PALSAR-2 and Sentinel-2A data. *International Journal of Remote Sensing,* 39**,** 7761-7788.
PIBUMRUNG, P., GAJASENI, N. & POPAN, A. 2008. Profiles of carbon stocks in forest, reforestation and agricultural land, Northern Thailand. *Journal of Forestry Research,* 19**,** 11-18.
PINTY, B. & VERSTRAETE, M. M. 1992. GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetatio,* 101**,** 15-20.
PLANET LABS INC. 2020. Planet Products Image Specifications.
POGGIO, L., SIMONETTI, E. & GIMONA, A. 2018. Enhancing the WorldClim data set for national and regional applications. *The Science of the total environment,* 625**,** 1628-1643.
POLEY, G., LUCY, MCDERMID, J. & GREGORY 2020. A systematic review of the factors influencing the estimation of vegetation aboveground biomass using unmanned aerial systems. *Remote Sensing,* 12**,** 1052.
PONCE-HERNANDEZ, R., KOOHAFKAN, P. & ANTOINE, J. 2004. *Assessing carbon stocks and modelling win-win scenarios of carbon sequestration through land-use changes,* Food & Agriculture Org.
PONTIUS, J., SCHABERG, P. & HANAVAN, R. 2020. Remote sensing for early, detailed, and accurate detection of forest disturbance and decline for protection of biodiversity. *Remote sensing of plant biodiversity***,** 121-154.
POPESCU, S. C. & HAUGLIN, M. 2014. Estimation of Biomass Components by Airborne Laser Scanning. *In:* MALTAMO, M., NAESSET, E. & VAUHKONEN, J. (eds.) *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies.* Dordrecht: Springer.
POUYAT, R. V., RUSSELL-ANELLI, J., YESILONIS, I. D. & GROFFMAN, P. M. 2002. Soil carbon in urban forest ecosystems. *The potential of US forest soils to sequester carbon and mitigate the greenhouse effect.* CRC Press.
PRAKASH, A. J., BEHERA, M., GHOSH, S., DAS, A. & MISHRA, D. 2022. A new synergistic approach for Sentinel-1 and PALSAR-2 in a machine learning framework to predict aboveground biomass of a dense mangrove forest. *Ecological Informatics,* 72**,** 101900.
PREGITZER, C. C., HANNA, C., CHARLOP-POWERS, S. & BRADFORD, M. A. 2022. Estimating carbon storage in urban forests of New York City. *Urban Ecosystems,* 25**,** 617-631.
QI, J., CHEHBOUNI, A., HUETE, A. R., KERR, Y. H. & SOROOSHIAN, S. 1994. A modified soil adjusted vegetation index. *Remote sensing of environment,* 48**,** 119-126.
QI, S., SONG, B., LIU, C., GONG, P., LUO, J., ZHANG, M. & XIONG, T. 2022. Bamboo Forest Mapping in China Using the Dense Landsat 8 Image Archive and Google Earth Engine. *Remote Sensing,* 14**,** 762.
QI, W., SAARELA, S., ARMSTON, J., STÅHL, G. & DUBAYAH, R. 2019. Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data. *Remote Sensing of Environment,* 232**,** 111283.
RACITI, S. M., HUTYRA, L. R. & NEWELL, J. D. 2014. Mapping carbon storage in urban trees with multi-source remote sensing data: Relationships between biomass, land use,

and demographics in Boston neighborhoods. *Science of the Total Environment,* 500**,** 72-83.
RAICH, J. W., RUSSELL, A. E., KITAYAMA, K., PARTON, W. J. & VITOUSEK, P. M. 2006. Temperature influences carbon accumulation in moist tropical forests. *Ecology,* 87**,** 76-87.
RAJASUGUNASEKAR, D., PATEL, A. K., DEVI, K. B., SINGH, A., SELVAM, P. & CHANDRA, A. 2023. An Integrative Review for the Role of Forests in Combating Climate Change and Promoting Sustainable Development. *International Journal of Environment and Climate Change,* 13**,** 4331-4341.
RAMACHANDRA, T., AITHAL, B. H. & SREEJITH, K. 2015. GHG footprint of major cities in India. *Renewable and Sustainable energy reviews,* 44**,** 473-495.
RAMDANI, F. & FURQON, M. T. 2022. The simplicity of XGBoost algorithm versus the complexity of Random Forest, Support Vector Machine, and Neural Networks algorithms in urban forest classification. *F1000Research,* 11**,** 1069.
RAMJEAWON, M., DEMLIE, M., TOUCHER, M. L. & JANSE VAN RENSBURG, S. 2020. Analysis of three decades of land cover changes in the Maputaland Coastal Plain, South Africa. *Koedoe: African Protected Area Conservation and Science,* 62**,** 1-12.
RAPIDEYE, A. 2012. Satellite imagery product specifications. 2012.
REIERSEN, G., DAO, D., LÜTJENS, B., KLEMMER, K., AMARA, K., STEINEGGER, A., ZHANG, C. & ZHU, X. 2022. ReforesTree: A Dataset for Estimating Tropical Forest Carbon Stock with Deep Learning and Aerial Imagery. *arXiv preprint arXiv:2201.11192*.
RÉJOU-MÉCHAIN, M., BARBIER, N., COUTERON, P., PLOTON, P., VINCENT, G., HEROLD, M., MERMOZ, S., SAATCHI, S., CHAVE, J. & DE BOISSIEU, F. 2019. Upscaling forest biomass from field to satellite measurements: sources of errors and ways to reduce them. *Surveys in Geophysics,* 40**,** 881-911.
REN, Y., WEI, X., WEI, X., PAN, J., XIE, P., SONG, X., PENG, D. & ZHAO, J. 2011. Relationship between vegetation carbon storage and urbanization: A case study of Xiamen, China. *Forest Ecology and Management,* 261**,** 1214-1223.
REN, Y., YAN, J., WEI, X., WANG, Y., YANG, Y., HUA, L., XIONG, Y., NIU, X. & SONG, X. 2012. Effects of rapid urban sprawl on urban forest carbon stocks: Integrating remotely sensed, GIS and forest inventory data. *Journal of environmental management,* 113**,** 447-455.
REN, Z., ZHENG, H., HE, X., ZHANG, D., SHEN, G. & ZHAI, C. 2019. Changes in spatio-temporal patterns of urban forest and its above-ground carbon storage: Implication for urban CO2 emissions mitigation under China's rapid urban expansion and greening. *Environment international,* 129**,** 438-450.
REX, F. E., SILVA, C. A., DALLA CORTE, A. P., KLAUBERG, C., MOHAN, M., CARDIL, A., DA SILVA, V. S., DE ALMEIDA, D. R. A., GARCIA, M., BROADBENT, E. N., VALBUENA, R., STODDART, J., MERRICK, T. & HUDAK, A. T. 2020. Comparison of Statistical Modelling Approaches for Estimating Tropical Forest Aboveground Biomass Stock and Reporting Their Changes in Low-Intensity Logging Areas Using Multi-Temporal LiDAR Data. *Remote Sensing,* 12**,** 20.
RIJAL, S. S., PHAM, T. D., NOER'AULIA, S., PUTERA, M. I. & SAINTILAN, N. 2023. Mapping Mangrove Above-Ground Carbon Using Multi-Source Remote Sensing Data and Machine Learning Approach in Loh Buaya, Komodo National Park, Indonesia. *Forests,* 14**,** 94.
RODRÍGUEZ-VEIGA, P., WHEELER, J., LOUIS, V., TANSEY, K. & BALZTER, H. 2017. Quantifying Forest Biomass Carbon Stocks From Space. *Current Forestry Reports,* 3**,** 1-18.
RONOUD, G., FATEHI, P., DARVISHSEFAT, A. A., TOMPPO, E., PRAKS, J. & SCHAEPMAN, M. E. 2021. Multi-Sensor Aboveground Biomass Estimation in the Broadleaved Hyrcanian Forest of Iran. *Canadian Journal of Remote Sensing,* 47**,** 818-834.
RUDNICKI, W. R., WRZESIEŃ, M. & PAJA, W. 2015. All relevant feature selection methods and applications. *Feature Selection for Data and Pattern Recognition***,** 11-28.

SAFARI, A., SOHRABI, H., POWELL, S. & SHATAEE, S. 2017. A comparative assessment of multi-temporal Landsat 8 and machine learning algorithms for estimating aboveground carbon stock in coppice oak forests. *International Journal of Remote Sensing,* 38**,** 6407-6432.
SALINAS-MELGOZA, M. A., SKUTSCH, M. & LOVETT, J. C. 2018. Predicting aboveground forest biomass with topographic variables in human-impacted tropical dry forest landscapes. *Ecosphere,* 9**,** e02063.
SALUNKHE, O., KHARE, P. K., KUMARI, R. & KHAN, M. L. 2018. A systematic review on the aboveground biomass and carbon stocks of Indian forest ecosystems. *Ecological Processes,* 7**,** 12.
SANLC 2020. South African National Landcover 2020 Accuracy assessment report. Pretoria, South Africa Department of Environment, forestry and fisheries
SANTI, E., PALOSCIA, S., PETTINATO, S., CUOZZO, G., PADOVANO, A., NOTARNICOLA, C. & ALBINET, C. 2020. Machine-learning applications for the retrieval of forest biomass from airborne P-Band SAR data. *Remote Sensing,* 12**,** 804.
SARKER, L. R. & NICHOL, J. E. 2011. Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sensing of Environment,* 115**,** 968-977.
SCHOLZ, T., HOF, A. & SCHMITT, T. 2018. Cooling effects and regulating ecosystem services provided by urban trees—novel analysis approaches using urban tree cadastre data. *Sustainability,* 10**,** 712.
SETO, K. C. & CHRISTENSEN, P. 2013. Remote sensing science to inform urban climate change mitigation strategies. *Urban Climate,* 3**,** 1-6.
SETO, K. C., GÜNERALP, B. & HUTYRA, L. R. 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences,* 109**,** 16083-16088.
SHAFRI, H. Z. & HAMDAN, N. 2009. Hyperspectral imagery for mapping disease infection in oil palm plantationusing vegetation indices and red edge techniques. *American Journal of Applied Sciences,* 6**,** 1031.
SHAHANI, N. M., ZHENG, X., LIU, C., HASSAN, F. U. & LI, P. 2021. Developing an XGBoost regression model for predicting young's modulus of intact sedimentary rocks for the stability of surface and subsurface structures. *Frontiers in Earth Science,* 9**,** 761990.
SHAIK, A. B. & SRINIVASAN, S. 2018. A Brief Survey on Random Forest Ensembles in Classification Model. *International Conference on Innovative Computing and Communications*.
SHANG, X. & CHISHOLM, L. A. 2013. Classification of Australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 7**,** 2481-2489.
SHAO, Z., ZHANG, L. & WANG, L. 2017. Stacked Sparse Autoencoder Modeling Using the Synergy of Airborne LiDAR and Satellite Optical and SAR Data to Map Forest Above-Ground Biomass. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 10**,** 5569-5582.
SHARMA, K., SAIKIA, A., GOSWAMI, S. & BORTHAKUR, M. 2020. Aboveground biomass estimation and carbon stock assessment along a topographical gradient in the forests of Manipur, Northeast India. *Arabian Journal of Geosciences,* 13**,** 1-16.
SHAVER, T., KHOSLA, R. & WESTFALL, D. Utilizing green normalized difference vegetation indices (GNDVI) for production level management zone delineation in irrigated corn. The 18th World Congress of Soil Science, 2006.
SHEN, A., WU, C., JIANG, B., DENG, J., YUAN, W., WANG, K., HE, S., ZHU, E., LIN, Y. & WU, C. 2018a. Spatiotemporal variations of aboveground biomass under different terrain conditions. *Forests,* 9**,** 778.
SHEN, W., LI, M., HUANG, C., TAO, X. & WEI, A. 2018b. Annual forest aboveground biomass changes mapped using ICESat/GLAS measurements, historical inventory data, and time-series optical and radar imagery for Guangdong province, China. *Agricultural and Forest Meteorology,* 259**,** 23-38.

SHEN, X. J., JIANG, M., LU, X. G., LIU, X. T., LIU, B., ZHANG, J. Q., WANG, X. W., TONG, S. Z., LEI, G. C., WANG, S. Z., TONG, C. A., FAN, H. Q., TIAN, K., WANG, X. L., HU, Y. M., XIE, Y. H., MA, M. Y., ZHANG, S. W., CAO, C. X. & WANG, Z. C. 2021. Aboveground biomass and its spatial distribution pattern of herbaceous marsh vegetation in China. *Science China-Earth Sciences,* 64**,** 1115-1125.

SHIN, Y. J., MIDGLEY, G. F., ARCHER, E. R., ARNETH, A., BARNES, D. K., CHAN, L., HASHIMOTO, S., HOEGH-GULDBERG, O., INSAROV, G. & LEADLEY, P. 2022. Actions to halt biodiversity loss generally benefit the climate. *Global change biology,* 28**,** 2846-2874.

SHUKLA, P. R., SKEG, J., BUENDIA, E. C., MASSON-DELMOTTE, V., PÖRTNER, H.-O., ROBERTS, D., ZHAI, P., SLADE, R., CONNORS, S. & VAN DIEMEN, S. 2019. Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

SIBANDA, M., MUTANGA, O. & ROUGET, M. 2017. Testing the capabilities of the new WorldView-3 space-borne sensor's red-edge spectral band in discriminating and mapping complex grassland management treatments. *International Journal of Remote Sensing,* 38**,** 1-22.

SIBANDA, M., MUTANGA, O., ROUGET, M. & ODINDI, J. 2015. Exploring the potential of in situ hyperspectral data and multivariate techniques in discriminating different fertilizer treatments in grasslands. *Journal of Applied Remote Sensing,* 9**,** 096033-096033.

SILLEOS, N. G., ALEXANDRIDIS, T. K., GITAS, I. Z. & PERAKIS, K. 2006. Vegetation indices: advances made in biomass estimation and vegetation monitoring in the last 30 years. *Geocarto International,* 21**,** 21-28.

SILVA, C. A., DUNCANSON, L., HANCOCK, S., NEUENSCHWANDER, A., THOMAS, N., HOFTON, M., FATOYINBO, L., SIMARD, M., MARSHAK, C. Z., ARMSTON, J., LUTCHKE, S. & DUBAYAH, R. 2021. Fusing simulated GEDI, ICESat-2 and NISAR data for regional aboveground biomass mapping. *Remote Sensing of Environment,* 253**,** 112234.

SINGH, C., KARAN, S. K., SARDAR, P. & SAMADDER, S. R. 2022. Remote sensing-based biomass estimation of dry deciduous tropical forest using machine learning and ensemble analysis. *Journal of Environmental Management,* 308**,** 114639.

SINGH, N., NANDY, S. & VAN LEEUWEN, L. 2024. Tree aboveground carbon mapping in an Indian tropical moist deciduous forest using object-based image analysis and very high resolution satellite imagery. *Journal of the Indian Society of Remote Sensing,* 52**,** 723-734.

SITHOLE, K., ODINDI, J. & MUTANGA, O. 2018. Assessing the utility of topographic variables in predicting structural complexity of tree stands in a reforested urban landscape. *Urban Forestry & Urban Greening,* 31**,** 120-129.

SMITH, A., MAJOR, D., MCNEIL, R., WILLMS, W., BRISCO, B. & BROWN, R. 1995. Complementarity of radar and visible-infrared sensors in assessing rangeland condition. *Remote sensing of environment,* 52**,** 173-180.

SMITH, L. A., EISSENSTAT, D. M. & KAYE, M. W. 2017. Variability in aboveground carbon driven by slope aspect and curvature in an eastern deciduous forest, USA. *Canadian Journal of Forest Research,* 47**,** 149-158.

SOLOMON, N., BIRHANE, E., TADESSE, T., TREYDTE, A. C. & MELES, K. 2017. Carbon stocks and sequestration potential of dry forests under community management in Tigray, Ethiopia. *Ecological processes,* 6**,** 1-11.

SØRENSEN, R., ZINKO, U. & SEIBERT, J. 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrol. Earth Syst. Sci.,* 10**,** 101-112.

SORIANO-LUNA, M. D. L. Á., ÁNGELES-PÉREZ, G., GUEVARA, M., BIRDSEY, R., PAN, Y., VAQUERA-HUERTA, H., VALDEZ-LAZALDE, J. R., JOHNSON, K. D. & VARGAS, R. 2018. Determinants of above-ground biomass and its spatial variability in a temperate forest managed for timber production. *Forests,* 9**,** 490.

SOUSA-SILVA, R., DUFLOS, M., ORDÓÑEZ BARONA, C. & PAQUETTE, A. 2023. Keys to better planning and integrating urban tree planting initiatives. *Landscape and Urban Planning,* 231**,** 104649.
SOUZA, G. S. A. D., SOARES, V. P., LEITE, H. G., GLERIANI, J. M., DO AMARAL, C. H., FERRAZ, A. S., SILVEIRA, M. V. D. F., SANTOS, F. C. D., VELLOSO, S. G. S., DOMINGUES, G. F. & SILVA, S. 2019. Multi-sensor prediction of Eucalyptus stand volume: A support vector approach. *ISPRS Journal of Photogrammetry and Remote Sensing,* 156**,** 135-146.
STOLL, E., KONSTANSKI, H., ANDERSON, C., DOUGLASS, K. & OXFORT, M. The RapidEye constellation and its data products. 2012 IEEE Aerospace Conference, 2012. IEEE, 1-9.
STOW, D., NIPHADKAR, M. & KAISER, J. 2005. MODIS-derived visible atmospherically resistant index for monitoring chaparral moisture content. *International journal of remote sensing,* 26**,** 3867-3873.
SU, H., SHEN, W., WANG, J., ALI, A. & LI, M. 2020. Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests. *Forest Ecosystems,* 7**,** 1-20.
SU, Y., GUO, Q., XUE, B., HU, T., ALVAREZ, O., TAO, S. & FANG, J. 2016. Spatial distribution of forest aboveground biomass in China: Estimation through combination of spaceborne lidar, optical imagery, and forest inventory data. *Remote Sensing of Environment,* 173**,** 187-199.
SUARDANA, A. M. A. P., ANGGRAINI, N., NANDIKA, M. R., AZIZ, K., AS-SYAKUR, A. R., ULFA, A., WIJAYA, A. D., PRASETIO, W., WINARSO, G. & DEWANTI, R. 2023. Estimation and Mapping Above-Ground Mangrove Carbon Stock Using Sentinel-2 Data Derived Vegetation Indices in Benoa Bay of Bali Province, Indonesia. *Forest and Society,* 7**,** 116-134.
SUDHA, S. & AJI, S. 2019. A review on recent advances in remote sensing image retrieval techniques. *Journal of the Indian Society of Remote Sensing,* 47**,** 2129-2139.
SUN, R. & CHEN, L. 2017. Effects of green space dynamics on urban heat islands: Mitigation and diversification. *Ecosystem Services,* 23**,** 38-46.
SUN, S. & HUANG, R. An adaptive k-nearest neighbor algorithm. 2010 seventh international conference on fuzzy systems and knowledge discovery, 2010. IEEE, 91-94.
SUN, W. & LIU, X. 2020. Review on carbon storage estimation of forest ecosystem and applications in China. *Forest Ecosystems,* 7**,** 1-14.
SZATMÁRI, G., PIRKÓ, B., KOÓS, S., LABORCZI, A., BAKACSI, Z., SZABÓ, J. & PÁSZTOR, L. 2019. Spatio-temporal assessment of topsoil organic carbon stock change in Hungary. *Soil and Tillage Research,* 195**,** 104410.
TADDEO, S., DRONOVA, I. & DEPSKY, N. 2019. Spectral vegetation indices of wetland greenness: Responses to vegetation structure, composition, and spatial distribution. *Remote sensing of Environment,* 234**,** 111467.
TAMIMINIA, H., SALEHI, B., MAHDIANPARI, M., BEIER, C. M. & JOHNSON, L. 2022. EVALUATING PIXEL-BASED AND OBJECT-BASED APPROACHES FOR FOREST ABOVE-GROUND BIOMASS ESTIMATION USING A COMBINATION OF OPTICAL, SAR, AND AN EXTREME GRADIENT BOOSTING MODEL. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 3**,** 485-492.
TAMIMINIA, H., SALEHI, B., MAHDIANPARI, M., QUACKENBUSH, L., ADELI, S. & BRISCO, B. 2020. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing,* 164**,** 152-170.
TAMM, T. & REMM, K. 2009. Estimating the parameters of forest inventory using machine learning and the reduction of remote sensing features. *International Journal of Applied Earth Observation and Geoinformation,* 11**,** 290-297.
TANASE, M. A., MARIN, G., BELENGUER-PLOMER, M. A., BORLAF, I., POPESCU, F. & BADEA, O. Deep Neural Networks for Forest Growing Stock Volume Retrieval: A

Comparative Analysis for L-band SAR data. IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, 2020. IEEE, 4975-4978.
TANASE, M. A., PANCIERA, R., LOWELL, K., TIAN, S., HACKER, J. M. & WALKER, J. P. 2014. Airborne multi-temporal L-band polarimetric SAR data for biomass estimation in semi-arid forests. *Remote Sensing of Environment,* 145**,** 93-104.
TANG, X., WANG, Y.-P., ZHOU, G., ZHANG, D., LIU, S., LIU, S., ZHANG, Q., LIU, J. & YAN, J. 2011. Different patterns of ecosystem carbon accumulation between a young and an old-growth subtropical forest in Southern China. *Plant Ecology,* 212**,** 1385-1395.
TANG, X. L., PEREZ-CRUZADO, C., VOR, T., FEHRMANN, L., ALVAREZ-GONZALEZ, J. G. & KLEINN, C. 2016. Development of stand density management diagrams for Chinese fir plantations. *Forestry,* 89**,** 36-45.
TANVEER, M., RAJANI, T., RASTOGI, R., SHAO, Y.-H. & GANAIE, M. 2022. Comprehensive review on twin support vector machines. *Annals of Operations Research***,** 1-46.
TAUNK, K., DE, S., VERMA, S. & SWETAPADMA, A. A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 15-17 May 2019 2019. 1255-1260.
TAYLOR, R., DAVIS, C., BRANDT, J., PARKER, M., STÄUBLE, T. & SAID, Z. 2020. The rise of big data and supporting technologies in keeping watch on the world's forests. *International Forestry Review,* 22**,** 129-141.
TEMENOS, A., TEMENOS, N., KASELIMI, M., DOULAMIS, A. & DOULAMIS, N. 2023. Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP. *IEEE Geoscience and Remote Sensing Letters,* 20**,** 1-5.
TEO, H. C., ZENG, Y., SARIRA, T. V., FUNG, T. K., ZHENG, Q., SONG, X. P., CHONG, K. Y. & KOH, L. P. 2021. Global urban reforestation can be an important natural climate solution. *Environmental Research Letters,* 16**,** 034059.
TER-MIKAELIAN, M. T., GONSAMO, A., CHEN, J. M., MO, G. & CHEN, J. 2021. Historical and future carbon stocks in forests of northern Ontario, Canada. *Carbon Balance and Management,* 16**,** 1-18.
TIAN, Y., HUANG, H., ZHOU, G., ZHANG, Q., TAO, J., ZHANG, Y. & LIN, J. 2021. Aboveground mangrove biomass estimation in Beibu Gulf using machine learning and UAV remote sensing. *Science of The Total Environment,* 781**,** 146816.
TIMILSINA, N., ESCOBEDO, F. J., STAUDHAMMER, C. L. & BRANDEIS, T. J. 2014. Analyzing the causal factors of carbon stores in a subtropical urban forest. *Ecological Complexity,* 20**,** 23-32.
TOMPALSKI, P., WHITE, J. C., COOPS, N. C. & WULDER, M. A. 2019. Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data. *Remote Sensing of Environment,* 227**,** 110-124.
UNITED NATIONS. 2023. *Causes and Effects of Climate Change.* [Online]. online United Nations Available: https://www.un.org/en/climatechange/science/causes-effects-climate-change [Accessed 29 December 2023].
UNIYAL, S., PUROHIT, S., CHAURASIA, K., RAO, S. S. & AMMINEDU, E. 2022. Quantification of carbon sequestration by urban forest using Landsat 8 OLI and machine learning algorithms in Jodhpur, India. *Urban Forestry & Urban Greening,* 67**,** 127445.
UPRETI, A. 2022. Machine learning application in GIS and remote sensing: An overview.
URBAZAEV, M., THIEL, C., CREMER, F., DUBAYAH, R., MIGLIAVACCA, M., REICHSTEIN, M. & SCHMULLIUS, C. 2018. Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. *Carbon Balance and Management,* 13**,** 1-20.
USTIN, S. L. & JACQUEMOUD, S. 2020. How the optical properties of leaves modify the absorption and scattering of energy and enhance leaf functionality. *Remote sensing of plant biodiversity.* Springer, Cham.
VAHEDI, A., MATAJI, A., BABAYI-KAFAKI, S., ESHAGHI-RAD, J., HODJATI, S. & DJOMO, A. 2014. Allometric equations for predicting aboveground biomass of

beech-hornbeam standsin the Hyrcanian forests of Iran. *Journal of Forest Science,* 60**,** 236-247.
VALAVI, R., GUILLERA-ARROITA, G., LAHOZ-MONFORT, J. J. & ELITH, J. 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs,* 92**,** e01486.
VAN ECK, N. J. & WALTMAN, L. 2017. Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics,* 111**,** 1053-1070.
VAN NOIJE, T., BERGMAN, T., LE SAGER, P., O'DONNELL, D., MAKKONEN, R., GONÇALVES-AGEITOS, M., DÖSCHER, R., FLADRICH, U., VON HARDENBERG, J., KESKINEN, J.-P., KORHONEN, H., LAAKSO, A., MYRIOKEFALITAKIS, S., OLLINAHO, P., GARCÍA-PANDO, C. P., REERINK, T. J., SCHRÖDNER, R., WYSER, K. & YANG, S. 2021. EC-Earth3-AerChem: a global climate model with interactive aerosols and atmospheric chemistry participating in CMIP6. *Geoscientific Model Development*.
VAN ROOYEN, M., VAN ROOYEN, N. & STOFFBERG, G. 2013. Carbon sequestration potential of post-mining reforestation activities on the KwaZulu-Natal coast, South Africa. *Forestry,* 86**,** 211-223.
VELASCO, E., ROTH, M., NORFORD, L. & MOLINA, L. T. 2016. Does urban vegetation enhance carbon sequestration? *Landscape and urban planning,* 148**,** 99-107.
VENTER, Z. S., BARTON, D. N., CHAKRABORTY, T., SIMENSEN, T. & SINGH, G. 2022. Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover. *Remote Sensing,* 14**,** 4101.
VERLY, O. M., LEITE, R. V., DA SILVA TAVARES-JUNIOR, I., DA ROCHA, S. J. S. S., LEITE, H. G., GLERIANI, J. M., RUFINO, M. P. M. X., DE FATIMA SILVA, V., TORRES, C. M. M. E. & PLATA-RUEDA, A. 2023. Atlantic Forest woody carbon stock estimation for different successional stages using Sentinel-2 data. *Ecological Indicators,* 146**,** 109870.
VERONESI, F. & SCHILLACI, C. 2019. Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. *Ecological Indicators,* 101**,** 1032-1044.
VICHARNAKORN, P., SHRESTHA, R. P., NAGAI, M., SALAM, A. P. & KIRATIPRAYOON, S. 2014. Carbon stock assessment using remote sensing and forest inventory data in Savannakhet, Lao PDR. *Remote Sensing,* 6**,** 5452-5479.
VINCENT, M. A. & SAATCHI, S. S. 1999. Comparison of remote sensing techniques for measuring carbon sequestration. *JPL Publ.*
WALTARI, E., SCHROEDER, R., MCDONALD, K., ANDERSON, R. P. & CARNAVAL, A. 2014. Bioclimatic variables derived from remote sensing: Assessment and application for species distribution modelling. *Methods in Ecology and Evolution,* 5**,** 1033-1042.
WAN, S., HUI, D., WALLACE, L. & LUO, Y. 2005. Direct and indirect effects of experimental warming on ecosystem carbon processes in a tallgrass prairie. *Global Biogeochemical Cycles,* 19.
WANG, B., GRAY, J. M., WATERS, C. M., ANWAR, M. R., ORGILL, S. E., COWIE, A. L., FENG, P. & LI LIU, D. 2022a. Modelling and mapping soil organic carbon stocks under future climate change in south-eastern Australia. *Geoderma,* 405**,** 115442.
WANG, D., WAN, B., LIU, J., SU, Y., GUO, Q., QIU, P. & WU, X. 2020. Estimating aboveground biomass of the mangrove forests on northeast Hainan Island in China using an upscaling method from field plots, UAV-LiDAR data and Sentinel-2 imagery. *International Journal of Applied Earth Observation and Geoinformation,* 85**,** 101986.
WANG, L.-H. & XING, Y.-Q. 2008. Remote sensing estimation of natural forest biomass based on an artificial neural network. *Ying yong sheng tai xue bao= The journal of applied ecology,* 19**,** 261-266.
WANG, L., JIA, M., YIN, D. & TIAN, J. 2019a. A review of remote sensing for mangrove forests: 1956–2018. *Remote Sensing of Environment,* 231**,** 111223.

WANG, L., LIU, J., XU, S., DONG, J. & YANG, Y. 2017. Forest above ground biomass estimation from remotely sensed imagery in the mount tai area using the RBF ANN algorithm. *Intelligent Automation & Soft Computing*, 1-8.
WANG, L., ZHU, R., YIN, Z., CHEN, Z., FANG, C., LU, R., ZHOU, J. & FENG, Y. 2022b. Impacts of land-use change on the spatio-temporal patterns of terrestrial ecosystem carbon storage in the Gansu Province, Northwest China. *Remote Sensing,* 14**,** 3164.
WANG, Q., PUTRI, N. A., GAN, Y. & SONG, G. 2022c. Combining both spectral and textural indices for alleviating saturation problem in forest LAI estimation using Sentinel-2 data. *Geocarto International***,** 1-21.
WANG, S., BAUM, A., ZARCO-TEJADA, P. J., DAM-HANSEN, C., THORSETH, A., BAUER-GOTTWEIN, P., BANDINI, F. & GARCIA, M. 2019b. Unmanned Aerial System multispectral mapping for low and variable solar irradiance conditions: Potential of tensor decomposition. *ISPRS journal of photogrammetry and remote sensing,* 155**,** 58-71.
WANG, Y., QIN, R., CHENG, H., LIANG, T., ZHANG, K., CHAI, N., GAO, J., FENG, Q., HOU, M., LIU, J., LIU, C., ZHANG, W., FANG, Y., HUANG, J. & ZHANG, F. 2022d. Can Machine Learning Algorithms Successfully Predict Grassland Aboveground Biomass? *Remote. Sens.,* 14**,** 3843.
WANG, Z., LI, X., MAO, Y., LI, L., WANG, X. & LIN, Q. 2022e. Dynamic simulation of land use change and assessment of carbon storage based on climate change scenarios at the city level: A case study of Bortala, China. *Ecological Indicators,* 134**,** 108499.
WARWICK-CHAMPION, E., DAVIES, K. P., BARBER, P., HARDY, N. & BRUCE, E. 2022. Characterising the Aboveground Carbon Content of Saltmarsh in Jervis Bay, NSW, Using ArborCam and PlanetScope. *Remote Sensing,* 14**,** 1782.
WASSENAER, P. J. E. V., SATEL, A., KENNEY, W. A., URŠIČ, M., JOHNSTON, M. & PERCIVAL, G. A framework for strategic urban forest management planning and monitoring. 2012.
WENG, Q. 2012. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment,* 117**,** 34-49.
WIJAYA, A. & GLOAGUEN, R. Fusion of ALOS Palsar and Landsat ETM data for land cover classification and biomass modeling using non-linear methods. 2009 IEEE International Geoscience and Remote Sensing Symposium, 2009. IEEE, III-581-III-584.
WONG, F. K. & FUNG, T. 2014. Combining eo-1 hyperion and envisat asar data for mangrove species classification in Mai Po Ramsar Site, Hong Kong. *International Journal of Remote Sensing,* 35**,** 7828-7856.
WOOD, S. N. 2017. *Generalized additive models: an introduction with R*, CRC press.
WU, C., NIU, Z. & GAO, S. 2012. The potential of the satellite derived green chlorophyll index for estimating midday light use efficiency in maize, coniferous forest and grassland. *Ecological Indicators,* 14**,** 66-73.
WU, C., NIU, Z., TANG, Q. & HUANG, W. 2008. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and forest meteorology,* 148**,** 1230-1241.
WU, C. Y., HEMBER, R. A., CHEN, J. M., KURZ, W. A., PRICE, D. T., BOISVENUE, C., GONSAMO, A. & JU, W. M. 2014. Accelerating Forest Growth Enhancement due to Climate and Atmospheric Changes in British Colombia, Canada over 1956-2001. *Scientific Reports,* 4**,** 5.
WU, W., ZHU, Y. & WANG, Y. 2023. Spatio-Temporal Pattern, Evolution and Influencing Factors of Forest Carbon Sinks in Zhejiang Province, China. *Forests,* 14**,** 445.
WULDER, M. & BOOTS, B. 1998. Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *International Journal of Remote Sensing,* 19**,** 2223-2231.
WULDER, M. A., WHITE, J. C., NELSON, R. F., NÆSSET, E., ØRKA, H. O., COOPS, N. C., HILKER, T., BATER, C. W. & GOBAKKEN, T. 2012. Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment,* 121**,** 196-209.

XI, L., SHU, Q., SUN, Y., HUANG, J. & SONG, H. 2023. Carbon storage estimation of mountain forests based on deep learning and multisource remote sensing data. *Journal of Applied Remote Sensing,* 17**,** 014510-014510.
XIAO, C.-W. & CEULEMANS, R. 2004. Allometric relationships for below-and aboveground biomass of young Scots pines. *Forest ecology and management,* 203**,** 177-186.
XIAO, J., CHEN, L., ZHANG, T., LI, L., YU, Z., WU, R., BAI, L., XIAO, J. & CHEN, L. 2022. Identification of Urban Green Space Types and Estimation of Above-Ground Biomass Using Sentinel-1 and Sentinel-2 Data. *Forests,* 13**,** 1077.
XIAO, J., CHEVALLIER, F. D. R., GOMEZ, C., GUANTER, L., HICKE, J. A., HUETE, A. R., ICHII, K., NI, W., PANG, Y., RAHMAN, A. F., SUN, G., YUAN, W., ZHANG, L. & ZHANG, X. 2019. Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment*.
XIONG, Y. & WANG, H. 2022. Spatial relationships between NDVI and topographic factors at multiple scales in a watershed of the Minjiang River, China. *Ecological Informatics,* 69**,** 101617.
XU, S.-C., HE, Z.-X., LONG, R.-Y., CHEN, H., HAN, H.-M. & ZHANG, W.-W. 2016. Comparative analysis of the regional contributions to carbon emissions in China. *Journal of Cleaner Production,* 127**,** 406-417.
YAN, L., MENG, S., YANG, F., DAI, X. & WANG, H. 2023. Changes in Forest Vegetation Carbon Storage and Its Driving Forces in Subtropical Red Soil Hilly Region over the Past 34 Years: A Case Study of Taihe County, China. *Forests,* 14**,** 602.
YAO, Z., LIU, J., ZHAO, X., LONG, D. & WANG, L. 2015. Spatial dynamics of aboveground carbon stock in urban green space: a case study of Xi'an, China. *Journal of Arid Land,* 7**,** 350-360.
YORO, K. O. & DARAMOLA, M. O. 2020. CO2 emission sources, greenhouse gases, and the global warming effect. *Advances in carbon capture.* Elsevier.
YUE, S., JI, G., CHEN, W., HUANG, J., GUO, Y. & CHENG, M. 2023. Spatial and Temporal Variability Characteristics of Future Carbon Stocks in Anhui Province under Different SSP Scenarios Based on PLUS and InVEST Models. *Land,* 12**,** 1668.
ZAVALETA, E. S., THOMAS, B. D., CHIARIELLO, N. R., ASNER, G. P., SHAW, M. R. & FIELD, C. B. 2003. Plants reverse warming effect on ecosystem water balance. *Proceedings of the National Academy of Sciences,* 100**,** 9892-9893.
ZHANG, F., TIAN, X., ZHANG, H. & JIANG, M. 2022a. Estimation of aboveground carbon density of forests using deep learning and multisource remote sensing. *Remote Sensing,* 14**,** 3022.
ZHANG, F., TIAN, X., ZHANG, H. & JIANG, M. 2022b. Estimation of Aboveground Carbon Density of Forests Using Deep Learning and Multisource Remote Sensing. *Remote. Sens.,* 14**,** 3022.
ZHANG, J., ZHANG, H., WANG, R., ZHANG, M., HUANG, Y., HU, J. & PENG, J. 2022c. Measuring the critical influence factors for predicting carbon dioxide emissions of expanding megacities by XGBoost. *Atmosphere,* 13**,** 599.
ZHANG, L. 2022. Effects of urban land use transitions on ecosystem services: Implication on Chinese urban planning. *Journal of Land Science*.
ZHANG, L., SHAO, Z., LIU, J. & CHENG, Q. 2019a. Deep learning based retrieval of forest aboveground biomass from combined LiDAR and landsat 8 data. *Remote Sensing,* 11**,** 1459.
ZHANG, L., ZHANG, X., SHAO, Z., JIANG, W. & GAO, H. 2023. Integrating Sentinel-1 and 2 with LiDAR data to estimate aboveground biomass of subtropical forests in northeast Guangdong, China. *International Journal of Digital Earth,* 16**,** 158-182.
ZHANG, M., DU, H., ZHOU, G., LI, X., MAO, F., DONG, L., ZHENG, J., LIU, H., HUANG, Z. & HE, S. 2019b. Estimating forest aboveground carbon storage in Hang-Jia-Hu using landsat TM/OLI data and random forest model. *Forests,* 10**,** 1004.
ZHANG, M. Z., LIN, H., ZENG, S. Q., LI, J. P., SHI, J. N. & WANG, G. X. 2013. Impacts of Plot Location Errors on Accuracy of Mapping and Scaling Up Aboveground Forest Carbon Using Sample Plot and Landsat TM Data. *Ieee Geoscience and Remote Sensing Letters,* 10**,** 1483-1487.

ZHANG, Q. L., HE, H. S., LIANG, Y., HAWBAKER, T. J., HENNE, P. D., LIU, J. X., HUANG, S. L., WU, Z. W. & HUANG, C. 2018. Integrating forest inventory data and MODIS data to map species-level biomass in Chinese boreal forests. *Canadian Journal of Forest Research,* 48**,** 461-479.

ZHANG, R., ZHOU, X., OUYANG, Z., AVITABILE, V., QI, J., CHEN, J. & GIANNICO, V. 2019c. Estimating aboveground biomass in subtropical forests of China by integrating multisource remote sensing and ground data. *Remote Sensing of Environment,* 232**,** 111341.

ZHANG, X., LI, X., JI, X., ZHANG, Z., ZHANG, H., ZHA, T. & JIANG, L. 2021. Elevation and total nitrogen are the critical factors that control the spatial distribution of soil organic carbon content in the shrubland on the Bashang Plateau, China. *Catena,* 204**,** 105415.

ZHANG, X., ZHAO, Y. M. & ASHTON, M. Chapter 7 Methods of Measuring Carbon in Forests. 2009.

ZHANG, X. Y. & NI-MEISTER, W. 2014. Remote Sensing of Forest Biomass. *In:* HANES, J. M. (ed.) *Biophysical Applications of Satellite Remote Sensing.* Berlin: Springer-Verlag Berlin.

ZHANG, Y., MA, J., LIANG, S., LI, X. & LI, M. 2020. An evaluation of eight machine learning regression algorithms for forest aboveground biomass estimation from multiple satellite data products. *Remote Sensing,* 12**,** 4015.

ZHANG, Y., MA, J., LIANG, S., LI, X. & LIU, J. 2022d. A stacking ensemble algorithm for improving the biases of forest aboveground biomass estimations from multiple remotely sensed datasets. *GIScience & Remote Sensing,* 59**,** 234-249.

ZHANG, Y. & SHAO, Z. 2021. Assessing of urban vegetation biomass in combination with LiDAR and high-resolution remote sensing images. *International Journal of Remote Sensing,* 42**,** 964-985.

ZHAO, C. & SANDER, H. A. 2015. Quantifying and mapping the supply of and demand for carbon storage and sequestration service from urban trees. *PLoS One,* 10**,** e0136392.

ZHAO, J., JI, G., YUE, Y., LAI, Z., CHEN, Y., YANG, D., YANG, X. & WANG, Z. 2019a. Spatio-temporal dynamics of urban residential CO2 emissions and their driving forces in China using the integrated two nighttime light datasets. *Applied Energy,* 235**,** 612-624.

ZHAO, J., MA, J., HOU, M. & LI, S. 2020. Spatial–temporal variations of carbon storage of the global forest ecosystem under future climate change. *Mitigation and Adaptation Strategies for Global Change,* 25**,** 603-624.

ZHAO, K., POPESCU, S., MENG, X., PANG, Y. & AGCA, M. 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sensing of Environment,* 115**,** 1978-1996.

ZHAO, Q., YU, S., ZHAO, F., TIAN, L. & ZHAO, Z. 2019b. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. *Forest Ecology and Management,* 434**,** 224-234.

ZHENG, D., RADEMACHER, J., CHEN, J., CROW, T., BRESEE, M., LE MOINE, J. & RYU, S.-R. 2004. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote sensing of environment,* 93**,** 402-411.

ZHOU, J.-J., ZAO, Z., ZHAO, Q., ZHAO, J. & WANG, H. 2013. Quantification of aboveground forest biomass using Quickbird imagery, topographic variables, and field data. *Journal of Applied Remote Sensing,* 7**,** 073484.

ZHOU, J., ZHOU, Z., ZHAO, Q., HAN, Z., WANG, P., XU, J. & DIAN, Y. 2020. Evaluation of Different Algorithms for Estimating the Growing Stock Volume of Pinus massoniana Plantations Using Spectral and Spatial Information from a SPOT6 Image. *Forests,* 11**,** 540.

ZOLKOS, S. G., GOETZ, S. J. & DUBAYAH, R. 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment,* 128**,** 289-298.