A Bayesian Geo-statistical Approach for Plantation Forest Productivity Assessment after the Fast-Track Land Reform in Zimbabwe

by

TSIKAI. S CHINEMBIRI

220112413

Submitted in partial fulfilment of the academic requirements of the Degree of Doctor of Philosophy in Environmental Science

School of Agricultural, Earth and Environmental Sciences

University of KwaZulu-Natal

Pietermaritzburg

South Africa

December 2023

ABSTRACT

The principal objective of the current study was to investigate how the new generation multispectral remote sensing, along with variants of the Bayesian hierarchical geostatistical methodology, could handle prediction uncertainty of carbon (C) stock. The assessment of C stock prediction uncertainty was conducted in a managed and disturbed plantation forest ecosystem located in Manicaland province of Zimbabwe. To achieve this, the study made use of ancillary data from the multispectral (Landsat-8 and Sentinel-2) remote sensing platforms, which informed the application of different inferential and methodological variants within the Bayesian hierarchical geostatistical framework. Allometric equations suited for the target plantation tree species in the sampled region were used to derive C stock from Above ground Biomass (AGB) sampled on 500 m^2 circular supports. These Bayesian geostatistical models utilized a combination of Landsat-8 and Sentinel-2 derived vegetation indices, along with climatic and topographic variables. The study found that the Normalized Difference Vegetation Index ($NDVI$), Distance to settlements ($DIST$), and Soil Adjusted Vegetation Index ($SAVI$) played crucial roles in influencing the spatial distribution of C stock in the studied region. Enhanced Vegetation Index (EVI) is an insignificant predictor for both Landsat-8 and Sentinel-2 driven C stock predictions. Among the tested Bayesian approaches, the spatially varying coefficient (SVC) model, the multi-source data-driven Bayesian geostatistical approach, and the frequentist geostatistical framework were examined. Regardless of the various specifications for independent variables in the predictive C stock modelling within the Bayesian framework, $NDVI$ and $DIST$ emerged as significant predictors of the modelled response variable. The non-stationary and Sentinel-2 driven Bayesian hierarchical model, with $NDVI$ and DIST covariables, proved to be the most effective prediction model in the studied plantation forest ecosystem in Zimbabwe. This best-performing C stock predictive model was subsequently used to predict C stock under both current (1970-2000) and future (SSP5-8.5) 2075 climate scenarios. The results of the Bayesian constructed hierarchical model indicate a significant shrinkage of forest C stock density and distribution under the future SSP5-8 (2075) business-as-usual climate projection. Basically, the findings of this study highlight the critical role of new generation multispectral remote sensing and Bayesian geostatistical approaches in assessing and predicting carbon stock uncertainty in forest ecosystems. These insights have significant implications for informed land management strategies, aligning with the goals and recommendations of the Intergovernmental Panel on Climate Change (IPCC) to effectively address climate challenges and enhance sustainable land management practices.

PREFACE

This research was conducted in the School of Agricultural, Earth and Environmental Sciences (SAEES), University of KwaZulu-Natal, Pietermaritzburg, from January 2021 to September 2023, under the supervision of Prof. Onisimo Mutanga and Prof. Timothy Dube.

I wish to declare that the research reported in this thesis has never been submitted in any form to any other university and represents my original work, except where due acknowledgments are made.

Tsikai Solomon Chinembiri

Signed:___ [redacted] Date:_21/08/ 2023_____

As the candidate's supervisor, I certify the above statement and have approved this thesis for submission.

Prof. Onisimo Mutanga

Signed: _ [redacted] _____ Date: __21/08/2023_____

Prof. Timothy Dube

Signed: _____ Date: _____

DECLARATION 1: PLAGIARISM

I, **Tsikai Solomon Chinembiri**, declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other institution.
3. This thesis does not contain other people's data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted:

a. Their words have been rewritten and the general information attributed to them has been referenced.

b. Where their exact words have been used, their writing has been placed in italics inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the references section.

Singed _____ [redacted] _____

DECLARATION 2: PUBLICATIONS

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

1. Tsikai Solomon Chinembiri*, Onisimo Mutanga, Timothy Dube, "Hierarchical Bayesian geostatistics for C stock prediction in disturbed plantation forest ecosystems", ***Ecological Informatics***, 73 (101934), 2023, DOI: 10.1016/j.ecoinf.2022.101934, 2023.

2. Tsikai Solomon Chinembiri*, Onisimo Mutanga, Timothy Dube, "Landsat-8 and Sentinel-2 Based Prediction of Forest Plantation C Stock Using Spatially Varying Coefficient Bayesian Hierarchical Models", ***Remote Sensing***, 14 (22), 2022, DOI: 10.3390/rs14225676.

3. Tsikai Solomon Chinembiri*, Onisimo Mutanga, Timothy Dube, "Carbon Stock Prediction in Managed Forest Ecosystems Using Bayesian and Frequentist Geostatistical Techniques and New Generation Remote Sensing Metrics", ***Remote Sensing***, 15 (6), DOI: 10.3390/rs15061649

4. Tsikai Solomon Chinembiri*, Onisimo Mutanga, Timothy Dube, "A multi-source data approach to Carbon stock prediction using Bayesian hierarchical geostatistical models in managed forest ecosystems, GIScience and Remote Sensing, Ms. No. TGRS-2023-0217 232584508), 2023, **Under Review**

5. Tsikai Solomon Chinembiri*, Onisimo Mutanga, Timothy Dube, "Multispectral remote sensing-based hierarchical modelling of forest C stock under projected climate warming in Zimbabwe, 2023, **Under review**

DEDICATION

Dedication to this thesis is extended to my family Amai Ambassador, Ambassador, Rutendo and Deborah Chinembiri and my late mother and father, Ronica Chinembiri and Tsikayi Solomon Chinembiri, respectively.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the University of KwaZulu-Natal (College of Agricultural, Earth and Environmental Sciences) for allowing me the opportunity to undertake my PhD study under the most difficult circumstances at a time when Covid-19 made it impossible for students to step foot into the campus. I also wish to thank my employer, the Zimbabwe Land Commission (ZLC) for supporting my application to study abroad whilst I enjoyed the full benefits of my job. The successful completion of this work could not have been possible without the tremendous support from academics, organisations and friends named henceforth:

I am very grateful to my supervisors, Prof. O. Mutanga and Prof. T. Dube for the generous support and advise throughout the duration of this study. Prof. Mutanga's simplistic and friendly approach to dealing with students makes you a unique and outstanding academic as this gave me the opportunity for self-discovery. Keep up the good work Prof. and may the Almighty Lord continue to bless your much-respected work in the region and abroad. Thank you, Prof. Dube for opening this golden avenue for me and I greatly appreciate imparting the spirit of hard work in me, our manuscripts never faced major challenges and may the Almighty Lord bless you abundantly. I would like to thank and acknowledge the National Research Foundation (NRF) of South Africa Chair in Land Use Planning and Management for their financial assistance. I also wish to give special mention to my MSc. second supervisor, Dr. D.G. Rossiter, for introducing and mentoring my appetite for geostatistical modelling at ITC (University of Twente) in the Netherlands.

I wish to acknowledge and thank the immense support I received from UKZN (College of Agricultural, Earth and Environmental Sciences) staff (Ms Andile Etah Mshengu, and Ms J. Manickum) for the great support provided during times of registration even when circumstances were unfavourable due to Covid-19. You made my life smooth ladies and may the mighty lord continue to bless you for the good work you are doing at the UKZN. I express my thanks to my generous friends Dr. P. Kowe, Dr. T. Mushore and Dr. Chemura who provided the much-needed advice and support in times of need, may the almighty bless you abundantly. Many thanks to Dr. Matongera for aiding in manuscript formatting during times of need.

My family deserves a special mention, Amai Ambassador, Ambassador, Rutendo and Deborah. A special mention also goes to my late mother, Mrs. Ronica Chinembiri who predicted this achievement some 23 years ago, even when the outlook appeared gloomy, you were a pillar of strength and courage, thank you Amai and may your soul continue to rest in eternal peace. I also wish to thank my late father, Tsikayi Solomon Chinembiri for laying out the foundation and encouraging me to acquire the necessary 21st century survival skills.

Table of Contents

LIST OF FIGURES

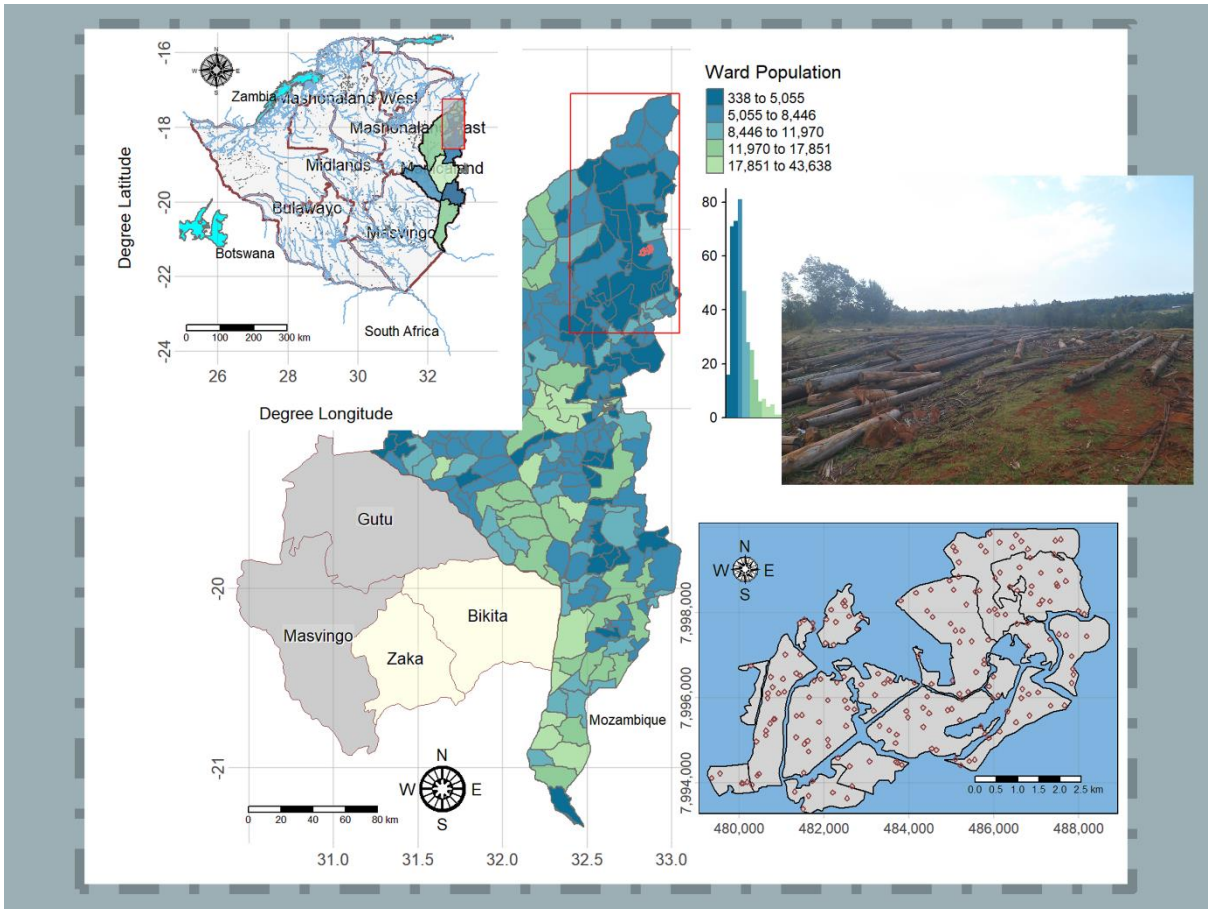
LIST OF TABLES

LIST OF ACRONYMS

AGB	Above Ground Biomass
C	C stock
CIW	Credible Interval Width
CI	Credible Interval
BEF	Biomass Expansion Factor
DBH	Diameter at Breast Height
DIST	Distance to Settlements
EVI	Enhanced Vegetation Index
GIS	Geographic Information System
FTLRP	Fast track Land Resettlement Programme
LULC	Land Use and Land Cover
LIDAR	Light Detection and Ranging
MSI	Multi-Spectral Instrument
MSS	Multi-Spectral Scanner
MAT	Mean Annual Temperature
MAP	Mean Annual Precipitation
MSSD	Mean Squared Shortest Distance
MCMC	Markov Chain Monte Carlo
NDVI	Normalized Difference Vegetation Index
NIR	Near infrared
OLI	Operational Land Imager
RMSPE	Root Mean Squared Prediction Error
RMSE	Root Mean Square Error
RRMSE	Relative Root Mean Square Error
RS	Remote Sensing
SAVI	Soil Adjusted Vegetation Index
SVC	Spatially Varying Coefficient
SVI	Spatiallly Varying Intercept
SRTM	Shuttle Radar Topography Mission
TOA	Top of the Atmosphere
UNFCCC	United Nations Framework Convention on Climate Change
USGS-EROS	United States Global Survey Earth Resources System

CHAPTER ONE

General Introduction



1.1. Introduction

Due to compelling evidence of the earth's increasing global average temperature, extensive research has been conducted to comprehend carbon sequestration dynamics of forest ecosystems (Wysowski, 2010). The disappearance of certain forest types and species through disruptions of ecological functions is clear testimony of the adverse effects of rising global temperatures (Wysowski, 2010). In addition, probable impacts of climate change are confirmed by the continued sea level rise, massive decrease in agricultural production and continued aridity (Wysowski, 2010). The biggest culprit behind the global warming phenomenon is the inert carbon dioxide (CO_2) gas, making solutions and efforts aimed at curtailing its levels in the atmosphere the most pragmatic approach (UNFCCC, 1998) for combating climate change impacts. Making headway in this endeavour, particluarly on the African continent is challenging as lives of the majority of the population is heavily dependent on the exploitation of primary resources that are responsible for atmospheric CO_2 emissions. For instance, the UNFCCC (UNFCCC, 2020), attributes the sub-Saharan African region as one of the overexploited forest resource regions of the world. This development poses significant threat to the global climate mitigation policies under the United Nations Framework Convention on Climate Change.

Planted and indigenous forest ecosystems in the developing world are increasingly coming under pressure from the rising demand of forest and land-based products and services (European Commission, 2017). This has led to the conversion or degradation of forest plantations and natural woodlands into unsustainable forms of land use (Ter-Mikaelian & Korzukhin, 1997). Such cases are particularly widespread in the developing world where appetite for forest products is high, mainly due to poverty and population growth. The severe degradation of forests compromises their capacity to function as regulators of the natural environment resulting in increased flood, erosion hazards, reduction in soil fertility and a loss of animal and plant life (Overman, Witte, & Saldarriaga, 1994). Accordingly, sustainable supply of goods and services from plantation ecosystems is threatened. The growing population and subsequent expansion of anthropogenic activities like infrastructure and settlements is threatening global forest cover. This trend is evident and predominant in most developing regions of the world, particularly the tropical rainforest belt of the African continent where the potential for carbon sequestration is greatest (Overman, Witte, & Saldarriaga, 1994).

1.2. The land reform programme and plantation forests in Zimbabwe

Zimbabwe's land reform officially started in 1980 after the signing of the Lancaster House Agreement (Falkowski et al., 2000). This was done in order to create an equitable distribution of land between subsistence farmers and white commercial farmers who owned large commercial farms in the country (FAO, 2018). The programme was mainly implemented on a willing buyer willing seller basis. However, the government fast tracked the reform programme at the beginning of the year 2000. By the year 2011, approximately 237,859 Zimbabwean households had been resettled and provided with access to land under the Fast Track Land Reform programme (Forestry-Commission, 2021). In essence, about 10,816,886 hectares were acquired in 2000 compared to 3,498,445 purchased from voluntary sellers between 1980 and 1998 (Forestry-Commission, 2021).

Given the significant contribution of the Sub-Saharan African region to the universal climate change mitigation as one of the world's biggest carbon reservoir, none of its member countries have undertaken a land redistribution programme to the scale and magnitude done in Zimbabwe in the past 50 years (Initiative For Climate Action Transparency (ICAT), 2022). Ironically, there have not been any study undertaken to date, to assess how the scale and extent of Zimbabwe's land reform programme has influenced the country's carbon budget. Zimbabwe has a long-established commitment to protection of biodiversity and its sustainable use as 16% of the country's land was under reserved forest and national parks before the 2000 Agrarian Land Reform Programme (FAO, 2018). Subsequently, the country signed and ratified the Convention on Biological Diversity (CBD) as a recognition of the role of natural resources to the national economy (Falkowski et al., 2000). However, there has been a steady decline of land under forest in Zimbabwe between 2000 and 2016 due to resettlement, uncontrolled fires and illegal timber harvesting from 168 581 hectares to 85 026 hectares in 2016 (European Commission, 2017). As pronounced by the Parliamentary Portfolio Committee on Environment and Tourism (2018), continued reduction of land under forest plantations and woodlots would ultimately lead to severe productivity losses and a delusional policy on the government's climate change policy agenda (Álvarez-Dávila *et al.*, 2017).

Efforts towards resuscitation of the timber industry are particularly significant as evidenced by the government's introduction of the tobacco levy in 2015 meant to regulate and manage access to forest resources (López-Serrano *et al.*, 2016). The reforestation fund is derived from a 0.75% tax on total sales charged on the country's 90 000 tobacco farmers depending on wood for tobacco curing (FAO, 2018). Unfortunately, this fund aimed at enhancing the country's carbon

footprint comes with no scientific backing to guide the implementation plans on areas with the most potential for C sequestration. Implementation of the forest resuscitation programmes at local scale is meant to encourage farmers to grow trees in order to outweigh forest losses due to Land Use Land Cover Change (LULC) activities (FAO, 2016).

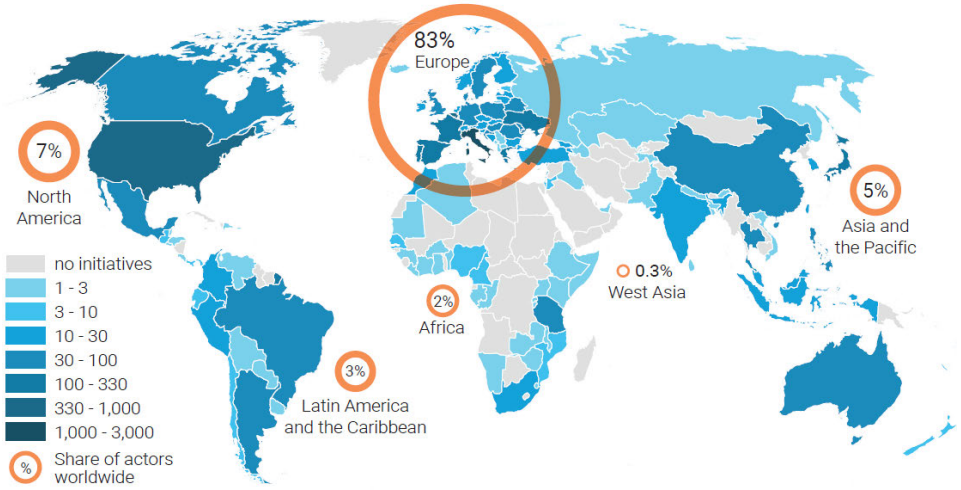


Figure 1-1: NSA participants and their geographical distribution. Source: Yale, New Climate Institute, 2018

This consequently points to the need for an objective implementation process that would drive existing carbon stock levels to their maximum potential. This would also be most appropriate for the REDD+ activities as the interventions would enhance Above Ground Biomass (ABG) within existing forested areas of local communities (FAO, 2016).

The timber industry forms one of the most important pillars of national development and contributes around 6 % of the country's Gross Domestic Product (GDP) in addition to vast employment opportunities in the sector (FAO, 2018). The post Agrarian Land Reform period has resulted in increased interface between communities and forest areas formerly designated for timber plantations (Forestry-Commission, 2021). In fact, the amount of additional biomass that can be accumulated in these areas depend much on the forest condition and management practices (Whitlow, 1998). The need for building models that depict the current state of affairs in these ecosystems can therefore not be overemphasized. Despite the growing contribution of Non-State and Sub-national Actors (NSA) in global climate governance after the 2015 Paris climate agreement, Zimbabwe remains one of the many Sub-Saharan African (SSA) countries with limited participation on climate initiatives (**Figure 1-1**). However, progressive participation from NSA and other arms of government will remain suppressed as long as there is no objective monitoring framework on the revival programme of the disturbed forest areas (Tran, Reef and Zhu, 2022). Engagement of private investment and other non-state actors in

the timber industry can enhance international cooperative initiatives that will lead to significant emission reduction.

1.3. C stock estimation methods and limitations

Various methods documented in literature for estimating AGB use either direct or indirect techniques from forest inventory data. Allometric equations, conversion factors like biomass expansion factors and wood density, form the major direct and indirect AGB estimation from forest inventory data (Matose, 2008). Despite the advantages associated with the use of conventional methods as detailed in Fuller *et al.* (2010) and Wang *et al.* (2020, the methods have a drawback of cost and time ineffectiveness in addition to posing a threat to the environment. Complicated topography and landscapes make some environments inaccessible, rendering conventional methods less practical for extensive AGB estimation programmes (Mutanga, Dube and Ahmed, 2016). Remote sensing is an emerging and promising method immune from the aforementioned shortfalls as it provides cost effective techniques for AGB assessment through stratification of forest types and canopy densities. Continuous and repeated application of remote sensing to the same area of interest generates time series data that can be used for change detection analysis and incorporation into a Geographic Information System (GIS) (Hoeting, 2009). Nevertheless, remote sensing methods also face limitations in situations of unfavourable weather conditions like cloud cover, requiring extensive processing and cleaning before the data can become fit for remote sensing applications (Fuller *et al.*, 2010).

The rapid growth pertaining to the application of remote sensing derived data in recent years has become the basis for informing global policy initiatives and direction linked to carbon dioxide (CO_2) emissions into the atmosphere, chiefly from deforestation and land cover changes (Zunguze, 2012). Thus, satellite based earth observation has become a powerful tool for the derivation of forest structural attributes and AGB since it provides feasible and practical methods of spatially distributed carbon from local to global scales (Gelfand *et al.*, 2003; Zvobgo and Tsoka, 2021). Three broad categories of satellite based earth observation methods namely, light detection and ranging (LiDAR), passive optical remote sensing and radio detection and ranging (radar) are used for measuring forest biomass (Zunguze, 2012; Mutanga, Dube and Ahmed, 2016). Passive optical remote sensing is chiefly used for detecting vegetation structural attributes, texture and shadow (Mutanga, Dube and Ahmed, 2016; Chrysafis *et al.*, 2017). Leaf area index, tree density and crown size have a strong correlation with AGB. On the other hand, radar and LiDAR remote sensing are used for measuring dielectric, geometrical attributes and characterising forest vertical structure and height, respectively (Korhonen *et al.*,

2017). When combined with GIS technologies, remote sensing data can provide cost effective and practical methods for research on extensive and inaccessible forest areas.

Prediction and estimation of AGB together with other structural forest parameters using new generation remote sensing multispectral data is a relatively new research ground within the climate change and carbon monitoring arena. The possibility of utilising new generation remote sensing for developing AGB estimates coupled with other earth observation methodologies or as standalone methods is therefore an active field of research (Gelfand *et al.*, 2003; Green, Finley and Strawderman, 2020). In the same vein, application of spatial regression models with new generation remote sensing data without accommodating spatial dependence gives imprecise C stock predictions (Popescu *et al.*, 2011; Tonolli *et al.*, 2011a). Furthermore, non-Bayesian spatial modelling of forest attributes can lead to the underestimation of prediction uncertainty (Hoeting, 2009; Babcock *et al.*, 2015) as classical spatial regression techniques do not make correct assumptions regarding the covariance structure, $\boldsymbol{\Sigma}$. A well-documented limitation of error maps generated from classical approaches regards their limitations for taking on board uncertainty emanating from variogram-derived spatial covariance parameters (Cressie, 1993). On the other hand, the employment of the Bayesian hierarchical framework to inference offers superior advantages including its ability to access the entire posterior predictive distribution (Goulard and Voltz, 1992).

Governments in developing countries including Zimbabwe are trying to come up with initiatives ranging from shifting to greener energy sources to afforestation and reforestation programmes as efforts for reducing their carbon footprint (Wang et al., 2020). Such initiatives are usually undertaken by communities at local levels without scientific backing to guide the implementation plans on areas with the most potential for the reforestation programmes. Given the government of Zimbabwe's quest for private investment in the timber industry, quantification of C stock concentration and distribution with an associated measure of uncertainty is critical for decision making. The Bayesian modelling and estimation framework is a novel approach with the potential to assist players in the timber industry with the needed facts and scientific backup for decision making.

1.4. Aim of the study

The study aimed at assessing how multispectral remote sensing informed Bayesian hierarchical geostatistical models handle uncertainty in predicting C stock in a disturbed plantation forest ecosystem in Zimbabwe.

1.5. Objectives of the study

The following objectives guided the study:

1. To assess how related multispectral remote sensing ancillary data deal with C stock prediction uncertainty in disturbed plantation forest ecosystems
2. To compare the performance of Sentinel-2 and Landsat-8 derived spatially varying coefficient (SVC) Bayesian hierarchical models in predicting C stock in plantation forests
3. To compare the potential of Bayesian geostatistical framework with a classical geostatistical approach for predicting plantation forest C stock
4. To assess how a remote sensing based multi-source data approach influences C stock prediction in managed plantation forest ecosystems.
5. To evaluate and characterise C stock distribution under a projected warming future climate scenario.

1.6. Research hypotheses

The study sought to validate the following research hypotheses:

1. Improved spatial and spectral properties of the Sentinel-2 will significantly improve C stock prediction uncertainty compared to the Landsat-8 derived spectral predictors.
2. A Bayesian hierarchical and Sentinel-2 led spatially varying coefficient C stock model offers improved better prediction uncertainty than a Landsat-8 spatially varying coefficient Bayesian model
3. A Sentinel-2 driven multi-source prediction approach provides lower C stock prediction uncertainty than the Landsat-8 driven multisource C stock model constructed within a Bayesian hierarchical framework.
4. The Bayesian hierarchical geostatistical inferential framework handles C stock prediction uncertainty better than its frequentist based geostatistical modelling
5. There will be a significant reduction in C stock density in the sampled region under a projected and warming climate scenario than the current (1970-2000) climate scenario.

1.7. Scope of the study

The study aimed at assessing how multispectral remote sensing informed Bayesian hierarchical geostatistical models handle uncertainty in predicting C stock in a disturbed plantation forest environment in Zimbabwe. New generation remote sensing platforms of Sentine-2 and Landsat-8 are the main sources of ancillary data for assessing how a Bayesian hierarchical geostatistical framework account for C stock prediction uncertainty. Vegetation indices and

anthropogenic data derived from the aforementioned remote sensing platforms in addition to climatic data are the main building blocks for assessing the models for C stock prediction uncertainty and accuracy. Lot 75 A of Nyanga Downs in Nyanga district in the Eastern Highlands of Zimbabwe in Manicaland province characterised and dominated by *Pinus* and *Eucalyptus* species forest plantations was utilised as the study area.

1.8. Description of the study area

The studied region is located at lot 75 A of Nyanga Downs in Nyanga district of Manicaland province in the Eastern Highlands of Zimbabwe, (**Figure 1-2**). The study area is made up of Eucalyptus species dominated plantation forest which has some of its patches being used for agriculture, grazing and gold panning and is located between latitude 32°40′ E and 32°54′ E and 18°10′12″ S and 18°25′48″ S longitude as illustrated in **Figure 1-2**. The study area is a former plantation forest estate owned by big corporates but now has some of its parts sub-divided into farming areas. The study domain covers an area of approximately 2 767 ha.

Rainfall amounts are variable and range from 741 mm to 2,997 mm with a mean annual precipitation of 1200 mm. Annual mean temperatures span from a minimum of 10°C to 13°C to a maximum of 26 °C to 29°C. The weather is relatively hot with extensive wild fires occurring in the high-elevation grasslands from September to November when the grasses are dry. Vegetation in the vicinity is mainly montane and sub-montane grassland. Some parts of the study area have a typical rain forest of the moist tropics comprising of five layers: The upper canopy (more than 20 m height), the medium canopy (between 8 and 20 m height), the lower layer (between 2 and 8 m height) the sapling layer (1 to 2 m height) and the ground layer characterized by seedlings; litter which in some locations is as thick as 20 cm and several herbal species.

1.9. General structure of the dissertation

The dissertation is comprised of seven chapters in which chapter one forms the introduction and chapter seven forms the synthesis chapter. Five research papers answering each of the aforesaid research objectives outlined in section 1.3, also form part of the thesis. The literature review and methodology are engrained in each of the mentioned papers.

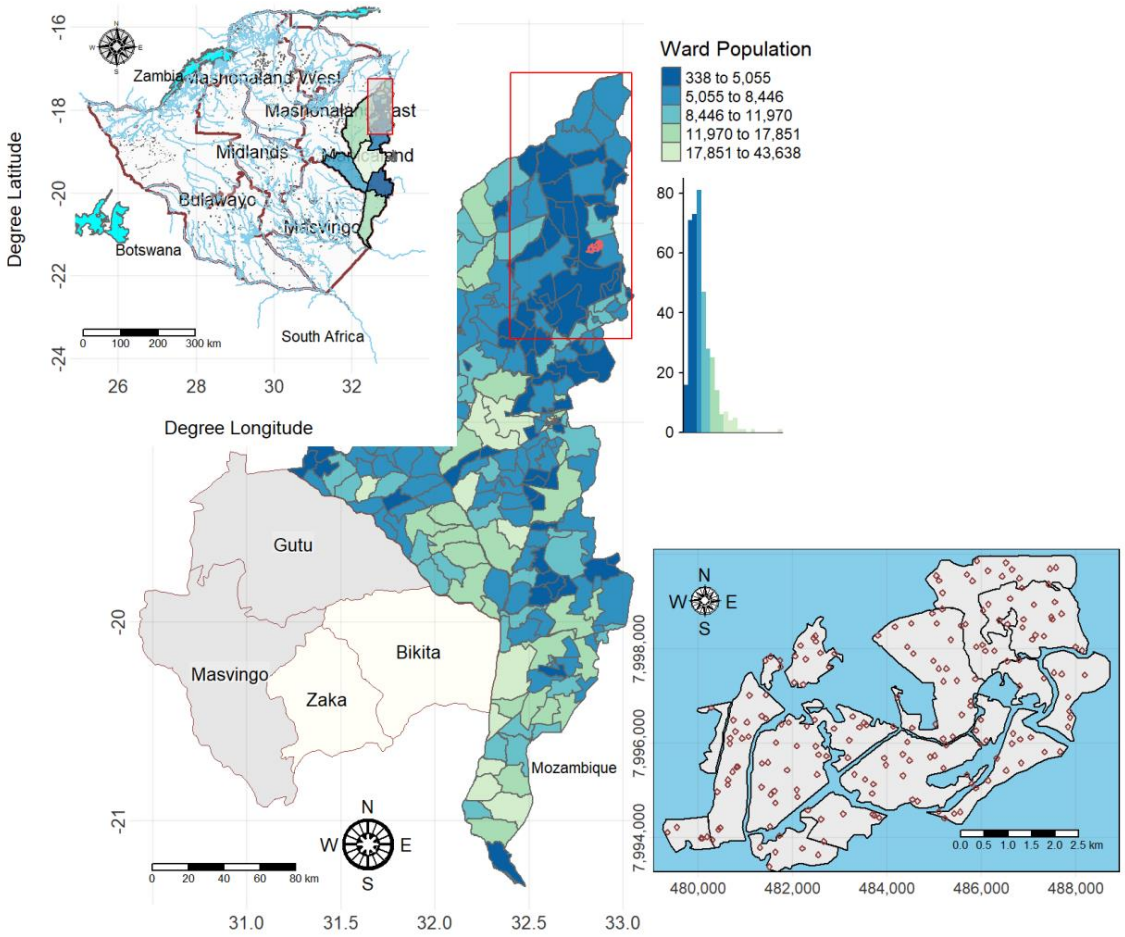


Figure 1-2: Study area in the eastern highlands of Zimbabwe in Manicaland province

1.9.1. General Overview and Contextualisation

1.9.2. Chapter One

Chapter one forms the introductory section of the study as it gives highlights of the various linkages between forest C stock estimation, prediction and climate change issues. Assessing and mapping key vegetation structural parameters like tree crown diameter, tree height and C stock contributes to a better understanding of the universal C cycle. Further elaboration regarding the various methods of estimating C stock and their possible limitations is also presented in chapter one of the study. This chapter culminates with a detailed description of the research problem, aims and objectives.

1.9.3. Chapter Two

Chapter 2 assesses the performance of a Bayesian hierarchical geostatistical framework using related new generation multispectral remote sensing derived metrics of Landsat-8 and Sentinel-2. Remote sensing metrics derived from the aforementioned sensors were used to predict C stock in a disturbed plantation ecosystem.

1.9.4. Chapter three

Chapter 3 evaluates how the Bayesian hierarchical approach performs in predicting C stock under spatially varying coefficients (predictors) (that is, the mean and model parameters are allowed to be non-stationary).

1.9.5. Chapter four

Chapter 5 assesses how a Bayesian hierarchical geostatistical approach performs in predicting C stock when the basket of predictor variables is broadened through incorporation of multiple sources of remote sensing data sets.

1.9.6. Chapter five

Chapter 4 evaluates how a Bayesian geostatistical framework compares with related frequentist geostatistical methods of Kriging with External Drift (KED) utilising the same basket of predictors from the Landsat-8 and the Sentinel-2 data.

1.9.7. Chapter six

Chapter 6 evaluates how the best C stock multispectral remote sensing driven Bayesian hierarchical model characterises C stock density and distribution under a future climate warming scenario in managed plantation forest ecosystems in Zimbabwe.

1.9.8. Chapter seven

Chapter seven presents the overall summary and synthesis of the study.

1.10. Linkage of chapters

The first chapter of this thesis presents a detailed introduction and description of the study by outlining the objectives and hypotheses of the research. This chapter therefore presents a snapshot of all the research objectives investigated in each of the subsequent chapters 2 to chapter six. Chapter two of the study evaluates the performance of the Bayesian hierarchical geostatistical approach using Sentinel-2 and Landsat-8 derived spectral indices influence C stock prediction uncertainty in disturbed plantation forests. This chapter therefore aims at coming up with significant and influential C stock predictors that influence prediction uncertainty of the modelled variable. Chapter three evaluates how significant predictors derived from both Sentinel-2 and Landsat-8 handles C stock prediction uncertainty when the model coefficients are assumed spatially varying. As such, this chapter is an extension of chapter 2 as it assesses how possible spatial heterogeneity in C stock in the form of spatially

varying coefficients (SVC) influences the prediction uncertainty of Landsat-8 and Sentinel-2 driven C stock prediction models.

The biggest limitation of SVC models regards their high demand for computational power, making the utilisation of such models with bigger sample sizes and a large suite of model predictors unfeasible with ordinary computers. As such, both Landsat-8 and Sentinel-2 driven C stock predictive models were only restricted to spatial random effects derived from vegetation indices, yet the source of such spatial variation is broad in practice. In order to exhaust possible sources of spatial variation of the modelled variable at local scale, chapter four adopts a multi-source data approach to C stock modelling using the Bayesian geostatistical methodology. Chapter five establishes how commonly utilised approaches within the geostatistical literature for predicting C stock using remote sensing derived ancillary data compares with the novel and emerging Bayesian hierarchical modelling technique. This chapter utilised the same predictors influencing C stock density when a multisource data approach was utilised in chapter four. Chapter six of the study wraps up and builds on the results of the preceding chapters and utilised significant C stock predictors from the best Sentinle-2 sensor to predict C stock density under a projected warming climate scenario in Zimbabwe. Chapter seven presents the conclusions and recommendations for future work coming from the study.

CHAPTER TWO

Hierarchical Bayesian geostatistics for C stock prediction in disturbed plantation forest in Zimbabwe

This chapter is based on

Ecological Informatics 73 (2023) 101934

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Hierarchical Bayesian geostatistics for C stock prediction in disturbed plantation forest in Zimbabwe

Tsikai S. Chinembiri [a,*], Onisimo Mutanga [a], Timothy Dube [b]

[a] University of KwaZulu-Natal, College of Agricultural, Earth and Environmental Sciences, P Bag X01 Scottsville, 3209 Pietermaritzburg, South Africa
[b] University of the Western Cape, Department of Earth Sciences, Private Bag X17, Bellville 7535, South Africa

Abstract

We develop and present a novel Bayesian hierarchical geostatistical model for the prediction of plantation forest carbon stock (C stock) in the eastern highlands of Zimbabwe using multispectral Landsat-8 and Sentinel-2 remotely sensed data. Specifically, we adopt a Bayesian hierarchical methodology encompassing a model-based inferential framework making use of efficient Markov Chain Monte Carlo (MCMC) techniques for assessing model input parameters. Our proposed hierarchical modelling framework evaluates the influence of two but related covariate information sources in C stock prediction in order to build sustainable capacity on carbon reporting and monitoring. The perceived improvements in the spatial and spectral properties of Sentinel-2 and Landsat-8 data and their potential to predict C stock with shorter uncertainty bounds is tested in the developed hierarchical Bayesian models. We utilized the Mean Squared Shortest Distance (MSSD) as the objective function for optimization of sampling locations for equal area coverage. Specifically, we evaluated the models using four selected remotely sensed vegetation indices namely, the normalised difference vegetation index ($NDVI$), soil adjusted vegetation index ($SAVI$), enhanced vegetation index (EVI) and an additional distance to settlements anthropogenic variable that justifies from the history of the studied plantation forest in the eastern highlands of Zimbabwe. We evaluated two models making use of the Sentinel-2 and the Landsat-8 derived predictors using the Root Mean Squared Error ($RMSE$), Mean Absolute Error (MAE), Coverage (CVG) and Deviance Information Criteria (DIC). The Sentinel-2 driven C stock model gave a RMSE of 1.16 $MgCha^{-1}$, MAE of 1.11 $MgCha^{-1}$, CVG of 94.7% and a DIC of -554.7 whilst its Landsat-8 based C stock counterpart yielded a $RMSE$, MAE, CVG and DIC of 2.69 $MgCha^{-1}$, 1.77 $MgCha^{-1}$, 85.4% and 43.1 respectively. Although predictive models from both sensors show great improvement in predictive accuracy when modelling the spatial random effects, the Sentinel-2 driven C stock predictive model considerably outperforms its Landsat-8 based C stock counterpart. The Sentinel-2 driven C stock hierarchical prediction model therefore adequately addresses multiple sources of uncertainty inherent in the spatial prediction of C stock in disturbed plantation ecosystems. It is evident from the results of this study that carbon reporting and monitoring can always be improved by scouting for improved and easily accessible remote sensing data and allow forest practitioners to keep track of error across space in resource environments of interest.

Keywords: Geostatistics**,** Bayesian Inference; Markov Chain Monte Carlo; Hierarchical Modelling, C stock, Bayesian Prediction, Mean Squared Shortest Distance

2.1. Introduction

The importance of different forest types as carbon sinks and the need to preserve, monitor and enhance terrestrial carbon stocks is recognised in the Kyoto Protocol of the United Nations Framework Convention on Climate Change (UNFCCC). This is because dynamics in forest carbon stocks influence the atmospheric carbon dioxide (CO_2) concentration (Millington & Townsend, 1989; Fan, Wang and Yang, 2022). Quantifying the uncertainty associated with forest carbon stock estimation and prediction can be enhanced by the inclusion of historical information such as scales of spatial variability known for characterising carbon stock dynamics in particular forest ecosystems (Do *et al.*, 2022; González-Vélez *et al.*, 2021). Surveys involving the collection of biomass data from forest plantations are time consuming and expensive. Yet, players and practitioners in the timber industry can always capitalise on historical data and expert opinion to quickly save as aids for accurate accounting and monitoring of carbon stock at landscape scales.

Natural forests in the sub-Saharan African region face increasing threats from cultivation, grazing and urban growth (Traore and Tieguhong, 2018). The rising demand of wood for both industrial and household energy and, in recent times, for carbon sequestration, rationalises the adoption of plantations as a viable and sustainable option for meeting these demands. For example, Zimbabwe lost approximately $3 billion in potential revenue and more than 4 000 jobs due to deforestation by settlers at the height of the Fast Track Land Reform Programme (FTLRP), twenty-two years ago (Newsday, 2017). Reports from the Timber Producers Federation (TPF) published in 2014 indicated that the country's timber plantations were nearly facing total collapse with the national timber industry declining by 25 % (Shumba & Marongwe, 2016).

An analysis of a mangrove forest in Vietnam making use of remote sensing data and Artificial Neural Network (ANN) gave mangrove Above Ground Biomass (AGB) predictions which ranged from 6.53 to 368.2 $Mgha^{-1}$ in 2000 and from 13.75 to 320.3 $Mgha^{-1}$ in 2020, respectively (Do *et al.*, 2022). AGB predictions made by Fararoda *et al.* (2021) using machine learning methods including random decision forest, AdaBoost, Bayesian ridge regression and multilayer neural networks endorsed random forest and AdaBoost as the best performing machine learning methods. It is acknowledged in literature that the type of sensor and the method of prediction influence the performance of AGB models. For instance, Fassnacht *et al.* (2014) compared LiDAR and hyperspectral data for modelling AGB and concluded that LiDAR data fused with Random Forest (RF) models offered the best AGB predictive model.

Zimbabwe has a long-established commitment to protection of biodiversity and its sustainable utilisation as 16 % of the country's land was under reserved forest and national parks before the 2000 Fast Track Land Reform Programme (FTLRP) (FAO, 2018). Subsequently, the country signed and ratified the Convention on Biodiversity (CBD) as a recognition of the role of natural resources to the national economy. Most of the forest disturbances being experienced in the forest plantations are blamed on the government's lack of support for the resettled farmers in the post Agrarian Land Reform period. This has resulted in increased interface between communities and forest areas formerly designated for timber plantations. In fact, the amount of additional biomass that can be accumulated in these areas depend much on the forest condition and management practices in place (Whitlow, 1998).

Obtaining useful estimates of uncertainty relating to forest carbon stock with imprecise predictions needed for decision making using model-based geostatistics is an established problem in climate change and carbon inventory studies (Ravindranath and Ostwald, 2008). Design-based estimation methods assume error propagation from sampling design and can therefore be appropriately accounted for if sampling data are probabilistically selected (Cochran, 1977; Thompson, 2002). On the other hand, errors emanate from the underlying process by which the outcome variable, that is, carbon stock, is generated in model-based assessments (Gregoire, 1998; Ver Hoef, 2002). A well-known limitation of error maps associated with classical geostatistical estimation techniques like co-kriging and kriging is their inability to accommodate uncertainty in the variogram-derived spatial covariance parameters (Cressie, 1993). Non-hierarchical implementations of spatial predictions struggle to efficiently deal with uncertainty associated with spatial autocorrelation parameters, that is, spatial decays and spatial variances (Diggle and Ribeiro Jr, 2007). The benefits of choosing a Bayesian mixed methodology to inference over other related techniques include access to the entire posterior predictive distribution (PPD) (Goulard and Voltz, 1992). According to Beloconi and Vounatsou (2020), full access to the PPD facilitates subsequent analysis that informs management and ecological objectives whilst accounting for prediction uncertainty.

Hierarchically constructed Bayesian geostatistical models are a potential remedy to the challenges faced when utilizing the aforementioned approaches (A. E. Gelfand *et al.*, 2004). Incorporation of medium and high spatial resolution remote sensing derived auxiliary information measured at the same geo-location as the independent variable further enhances the ability of hierarchical models in providing resource estimates with a reduced measure of uncertainty. Babcock et al. (2015) utilized LIDAR for predicting forest biomass using the hierachical Bayesian approach and established additional predictive performace after

accounting for spatial random effects to the prediction equation. Significant progress has also been made in mapping forest attributes at broader spatial scales using spatially enabled hierarchical models. Such forest attributes include forest biomass (Finley, Banerjee and MacFarlane, 2011, Johnson *et al.*, 2014), community forest structure (Finley et al. 2009), deforestation rate (Agarwal *et al.*, 2005) and specific attributes of tree structure (Babcock *et al.*, 2012). Making predictions of forest biomass and other forest parameters needed for carbon accounting under UNFCC using Bayesian geostatistics comprehensively takes parameter uncertainty into account. However, despite the improvements coming from new generation remote sensing platforms as the Landsat series and the European Space Agency (ESA) driven multispectral Sentinel products, no study has utilised vegetation indices from new generation sensors, particularly Sentinel-2 and Landsat-8, as predictors of C stock under a Bayesian Hierarchical framework in climate change studies.

An in-depth description of the global carbon cycle and reduction in carbon emissions can be done through the international monitoring of AGB forest. Nevertheless, there is still a lot of uncertainty regarding the spatial distribution and quantity of AGB emanating from the difficulties faced when measuring AGB using field measurement standards (Lefsky, 2010; Simard *et al.*, 2011). Space-borne remote sensing techniques can collect data correlated with AGB spatial distribution over large national and global regions in a cost-effective method (Clerici *et al.*, 2016). The limited responsiveness of earth observation sensors to AGB and the lack of adaptive data needed for standardisation at the appropriate scales suited for remote sensing, are some of the challenges of facing these C mapping methods. Getting correct allometric correlations in certain regions of the world is a problematic because of armed conflicts, remoteness of the areas of interest and lack of capacity to generate plausible data. Remote sensing signal saturation due to high density of AGB, clod cover, especially in tropical regions and complexity of signal retrieval due to complicated topography are also some of the major AGB mapping challenges (Song *et al.*, 2010).

Semela, Ramoelo and Adelabu (2020) recommend both Landsat-8 and Sentinel-2 as valid sources of information for grass biomass estimation in environments that are mountainous using machine learning (random) forest modelling. Astola *et al.* (2019) compared Landsat-8 and Sentinel-2 models for retrieving leaf area index (LAI), effective canopy cover (ECC) and canopy cover (CC) using spectral bands common to both sensors and found non-systematic differences between the Landsat-8 and the Sentinel-2 data. Korhonen *et al.* (2017) also compared Landsat-8 and Sentinel-2 imagery for forest variable prediction and found Sentinel-2 to outperform Landsat-8 due to the enhanced spatial resolution in the former compared to the

later. The majority of studies comparing Landsat-8 and Sentinel-2 for predicting AGB favour the Sentinel-2 sensor over its Landsat-8 equivalent. This is further justified from the work of Jha *et al.* (2021) who compared Worldview-3, Landsat-8 and Sentinel-2 for mapping AGB in a forest landscape in Thailand and established Worldview-3 and Sentinel-2 as better predictors than Landsat-8 owing to the red-edge and the higher spatial and spectral properties of Worldview-3 and Sentinel-2. It is therefore perceived that the improvements in the spatial and spectral properties of Sentinel-2 over Landsat-8 can minimise prediction uncertainty on C stock prediction under a fully modelled Bayesian Hierarchical framework.

The research purpose for the current paper is therefore to develop and evaluate the effectiveness of Bayesian hierarchical models employing the Sentinel-2 and the Landsat-8 Operational Land Imager (OLI) derived vegetation indices for carbon stock prediction in disturbed plantation ecosystems in Zimbabwe. The employment of Sentinel-2 derived predictors using a Bayesian geostatistical hierarchical approach is expected to offer better C stock predictions in C stock monitoring and reporting than the same model using predictor information from the Landsat-8 data. Model evaluation is utilised to identify the strengths and deficiencies of candidate models and hence, the trade-off associated with utilisation of different information sources in C stock prediction. We further assess and compare C stock prediction uncertainty intervals generated from the Sentinel-2 and the Landsat-8 OLI using a Bayesian hierarchical framework. This is achieved through the development of separate but related models, integrating Sentinel-2 and Landsat-8 OLI derived spectral indices. These models are constructed using a Bayesian hierarchical geostatistical framework employed for predicting C stock in pertubed plantation forests in Zimbabwe. The paper aims at providing clinical information needed for decision making by governments and Non-Governmental Organizations (NGOs) working towards the fulfilment of the 2030 plan on Sustainable Development and the Paris Agreement on climate change. The government's pursuit for charming private capital into the timber industry in recent years in Zimbabwe can only be taken seriously if C stock estimates in the targeted plantation ecosystems for reforestation and afforestation programmes are provided with an accompanying measure of uncertainty.

2.2. Methods

2.2.1. Study area

This investigation was carried out at lot 75 A of Nyanga Downs in Nyanga district in the Eastern Highlands of Zimbabwe, (**Figure 2-1**). The study area is dominated by *Eucalyptus grandis*, *Eucalyptus camaldulensis* and *Pinus patula* plantation forest species which has some

of its patches being used for agriculture, grazing and gold panning and is located between latitude 32º40' E and 32º54' E and 18º10'12''S and 18º25'4''S longitude as illustrated in Figure 1. Grazing, agriculture and gold panning activities came after part of the commercially owned plantation forests were redistributed to small and medium sized indigenous farmers in 2000. This development has increased interface between settlements and timber plantations in all forests originally designated under forest plantations in Zimbabwe.

The study domain covers an area of approximately 2 767 ha. Rainfall amounts are variable and range from 741 mm to 2,997 mm with a mean annual precipitation of 1300 mm. Annual mean temperatures span from a minimum of 9ºC to 13°C to a maximum of 25ºC to 29°C. The weather is very hot and extensive wild fires occur in the high-elevation grasslands from August to November where the grasses are dry.

2.2.2. Remote sensing derived predictors

The spatial coverage and extent of RS data such as Landsat in addition to their relatively low cost vindicate their usefulness as covariates. Sentinel-2 is comprised of 6 land monitoring bands which are analogous to Landsat-8 and also take on board three more bands encompassing the red-edge (RE) spectrum (Drusch *et al.*, 2012). The Red edge bands are cantered at 704, 740, and 782 nm with bandwidths of 15, 15, and 20 nm, respectively. The red-edge is the major spectral feature of vegetation positioned between the red absorption maximum (680 nm) and the high reflectance in the Near-Infrared (750 nm) (Frampton *et al.*, 2013). In addition, the Sentinel-2's land surface bands carry a spatial resolution of 10 m and 20 m in comparison to Landsat-8's 30 m bands. These sensor differences provide the basis and an inferential framework upon which to learn and to have a better understanding of the sensor that outperforms the other in C stock prediction.

2.2.2.1. Landsat OLI and Sentinel-2 MSI imagery

We obtained the Landsat-8 imagery from the United States Geological Survey Earth Explorer (http://earthexplorer.usgs.gov) as analysis-ready datasets (ARDs). Datasets were filtered with cloud cover and cloud shadow cover thresholds set to less than 10%. Sentinel-2 cloud-free imagery was acquired on 20 September 2020 at the same time as the Landsat-8 OLI data collection covering the entire area of interest at lot 75A Nyanga Downs in the eastern highlands of Zimbabwe.

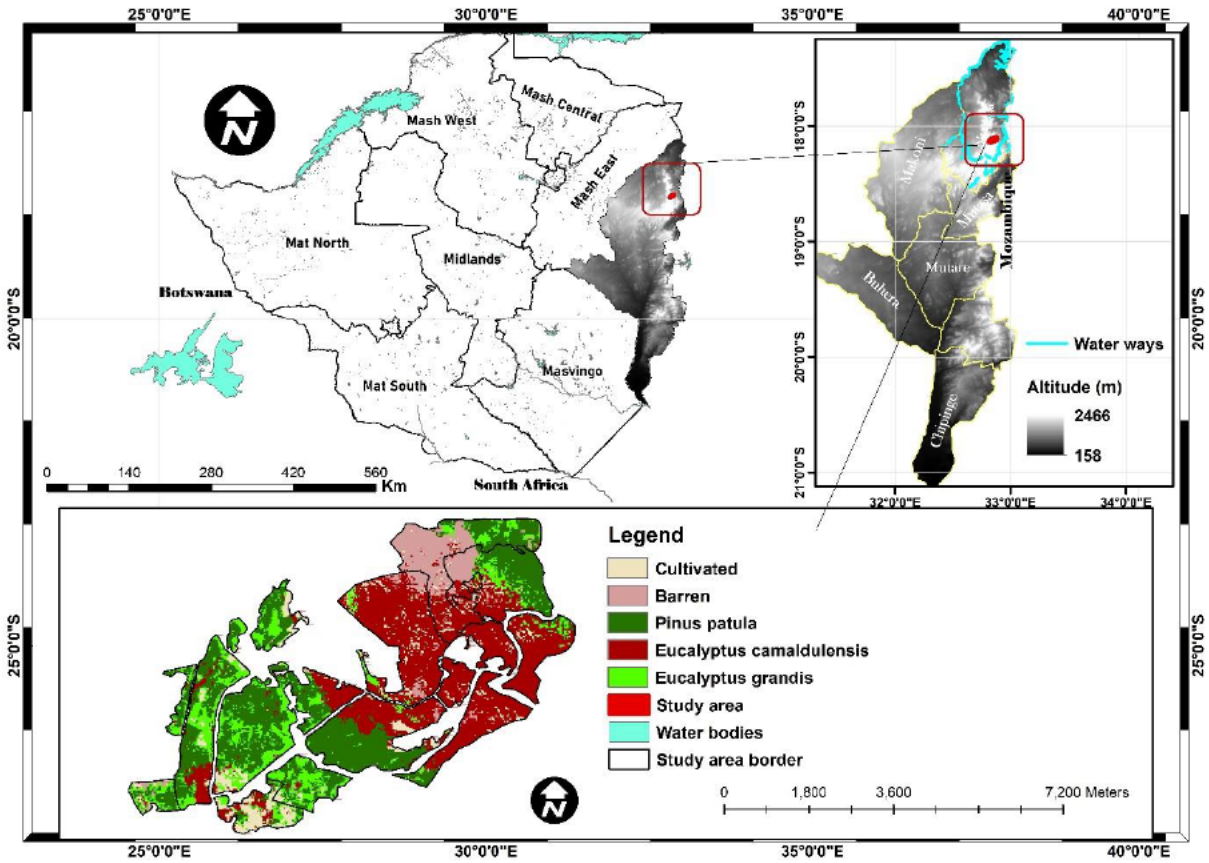


Figure 2-1: Map of the researched area indicating **(a)** the province where samples were derived, **(b)** study area location within the particular province and **(c)** spatial distribution of the sampled regionalised variable in the lower panel.

Sentinel-2 imagery is taken using the multispectral instrument (MSI), a push-broom imaging instrument that measures the Earth's top of atmosphere (TOA) reflected radiance in thirteen (13) spectral bands ranging from 443 nm to 2190 nm. The Sentinel-2 data were derived as level-1C 12-bit pre-set TOA reflectance values. The pre-processing and orthorectification of the level 1-C products was carried out in the R statistical and computing environment using the *sen2r* package (Ranghetti *et al.*, 2020).

We utilized Normalised Difference Vegetation Index (*NDVI*), Soil Adjusted Vegetation Index (*SAVI*), Enhanced Vegetation Index (*EVI*) and an additional distance to settlements anthropogenic variable as covariates for C stock prediction in a perturbed plantation forest ecosystem. The aforementioned vegetation indices have been applied in literature, including, Li and Li (2019) and Bordoloi *et al.*, (2022), as independent variables in AGB biomass estimation. Previous research on climate studies including Wang & Tenhunen (2005) and Fuller *et al.* (2010) have modelled the effects of anthropogenic or biophysical variables on biomass separately. In addition, a limited number of variables for each of these classes of predictors have been applied in biomass estimation. The present study employs a combination

of these classes of independent variables using a Bayesian hierarchical approach for C stock estimation and prediction.

2.3. Sampling design

This section provides details regarding the design of the AGB sampling programme and the spatial geographical coverage of the sampled region.

2.3.1. Spatial coverage sampling scheme

We utilized the Mean Squared Shortest Distance (MSSD) as the objective function for optimization of sample locations in the current study. We therefore employed the k-means clustering algorithm for equal area coverage sampling. Walvoort, Brus and de Gruijter (2010) demonstrated how the mapping of regionalized variables can be enhanced by uniform dispersal of sampling locations within a study domain. Furthermore, Brus, de Gruijter and Groenigen (2006) showed how uniform coverage of the study domain with sampling locations can both be used for resolving the mapping and estimation of spatial means of regionalized variables in the soil, forestry and environmental research.

Figure 2-2: Spatial coverage sampling design. Points represent locations of sampled C stock

Amongst the methods that have been utilized for optimization of the sampling pattern in literature, including, Spatial Simulated Annealing (SSA), the MSSD stands out as the most suited optimization methodology for both prediction and estimation of regionalized variables.

The design is especially appropriate in scenarios where sampling schemes cannot be extended beyond a single phase and where the area has never been sampled before.

2.3.2. Geographical partitioning of the study domain by the *k*-means algorithm

The study area was subdivided into compact subregions by clustering of the raster cells making up the study region using the k-means optimization algorithm (de Gruijter *et al.*, 2006, Walvoort *et al.*, 2010). The *k*-means optimization algorithm makes use of the x and y coordinates of the centre points of the raster cells as classification variables. Centroids of the clusters were then used as sample points where sampling plots for C stock were established as Illustrated in **Figure 2-2**.

2.4. Carbon stock data

Measured tree Diameter at Breast Height (DBH) from the field sampling programme were used to calculate per tree biomass and the final derivation of C stock.

2.4.1. Above ground tree biomass (AGTB) field measurement

We sampled and took measurements of all trees with at least 10 cm diameter at breast height (≈ 1.29 m above the soil surface) using 500 m^2 circular supports using diameter and linear tapes from the 19th of September 2021 to the 24th of October 2021. According to Gibbs et al. (2007), trees with DBH not more than 10 cm have negligible C stocks. Slopes in the study area are generally below 30 % and hence, slope correction was not considered during the tree biomass measurements (Ravindranath and Ostwald, 2008). Optimization of the sample locations that resulted in 200 spatial coverage samples was done in the *spcosa*-package of the *R Statistical and Computing Environment* (de Gruijter *et al.*, 2006, Walvoort *et al.*, 2010). The 200 sampling locations were pre-loaded into a 72 H handheld Garmin GPS before launching the field work exercise. However, 191 observations of forest biomass were obtained during the sampling phase as nine sampling plots fell outside the boundaries of the defined study area (Figure 2). Statistics of sampled C stock data for the measured plantation forest parameters from the field for the studied region are shown in **Table 2-1**.

Table 2-1: Summary statistics of the measured C stock plantation forest parameters

Statistic	*Eucalyptus camaldulensis*			*Eucalyptus grandis*			*Pinus patula*		
	DBH	Height	C stock	DBH	Height	C stock	DBH	Height	C stock
Mean	81.4	60.6	2485.3	67.4	70.6	405.7	56.8	58.6	377.9
Median	77.4	52.7	1470.3	51.4	49.7	327.8	43.5	38.7	295.4
Max	231.9	88.9	8998.2	97.9	90.1	429.8	64.3	66.6	600.3
Min	11.4	23.8	13.7	14.7	27.8	111.3	10.6	19.4	9.7

2.4.2. Biomass calculation and derivation of C stock

Pinus species biomass was calculated using allometric equations from Brown (1997) whilst biomass of the Eucalyptus species was calculated using allometric equations proposed by Zunguze (2012). Tree diameter was measured at approximately 1.3 m above the soil surface with all tree species with at least 10 cm measured. Application of the aforementioned allometric equations was befitting as the same equations were applied to Eucalyptus and Pinus species in Manica province in Mozambique which closely resembles and approximates the weather and climatic conditions of the studied region in Manicaland province of Zimbabwe. The above ground biomass of every individual tree species was then transformed to C stocks per tree species by means of a conversion factor of the IPCC (2006). Per plot derived values were then scaled up to a standard unit area, in this case, a hectare ($MgCha^{-1}$).

We took the position that the association between remote sensing features and AGB apart from being determined by the crown areas of the measured and sampled plantation species in Figure 1, but also determined by other factors like tree age, planting density and plantation conditions (i.e. climate or soils). We therefore assumed that the AGB model captures all these factors indirectly since the remote sensing signal is also a product of these factors.

2.5. Hierarchical Bayesian Modelling

We adopted the Bayesian hierarchical framework in order to fully account for parameter uncertainty in the sampled C stock. The Bayesian hierarchical framework consists of four stages as detailed below:

2.5.1. Observed C stock data

Let $Y(s)$ denotes the observed log transformed C stock data at a spatial location s, $s = 1, \dots, S$, then,

$Y(s) = X^T(s)\beta + w(s) + \varepsilon(s)$ ..... (2-1)

where

$w(s)$ is a spatial random effect term representing the effect of being at location s, and in our case, an exponential autocorrelation function

β is a vector of regression coefficients associated with a $1x\ q$ vector of predictors $X(s)$ sampled at location s

$\varepsilon(s)$ represents the random error (white noise) assumed $i.i.d.N(0, \sigma_\varepsilon^2)$

2.5.2. Spatial random effects specification

We assumed the spatial random effects $w(s) = (w, \dots, w_s)^T$ arise from a multivariate distribution, that is, a Gaussian random field, scaled by a spatial variance σ_w^2 and correlations proportional to the separation distance, d_{ij}, between sites as follows:

$w \sim MVN(0_s, \sigma_w^2 \Sigma_{\boldsymbol{w}})$ …….. (2-2)

Where 0_s is an $S\ x\ 1$ vector of zeros, σ_w^2 representing the between site variance, whilst Σ_w is the $S\ x\ S$ correlation matrix with elements (s_i, s_j) indicating the correlation between sites s_i and s_j, $i, j = 1, \dots, S$. We assumed a stationary model with an isotropic covariance structure where the correlation between sites s_i and s_j is a function of the separation distance between the sites. We therefore computed the covariance matrix Σ_w for the correlation function as follows:

$f(ds_i s, \phi, \kappa) = exp(-\phi d^\kappa s_i s_j)$ …… (2-3)

Because of limited information in the data to be able to estimate both κ and ϕ, we followed common practice and fixed κ to be one (Diggle and Ribeiro Jr, 2007).

2.5.3. Model prediction

A fully Bayesian hierarchical framework entails simultaneous estimation and prediction of model parameters. Uncertainty associated with estimates of model coefficients are considered and supplied through the model to the predictions. The Markov Chain Monte Carlo algorithm was used to make and estimate predictions at a location s' with the form:

$$\hat{Y}(s)' = X^T(s)'\hat{\beta} + \hat{w}(s)'$$

As such, predictions were made using a combination of the overall mean, the spatial effect and the effect of independent variables from medium resolution sensor satellite data. We calculated the spatial component of the predictions using multivariate normal properties. That is, if $\boldsymbol{w} = (w_1, \dots, w_s)'$ are the spatial effects at site of C stock observation, then the conditional distribution at an arbitrary location, s′, is $w_{s'}|\boldsymbol{w}$ that is normally distributed with mean and variance denoted by:

$E[w_{s'}|\boldsymbol{w}, Y] = \sigma_w^{-2}\delta_j'\Sigma_w^{-1}\boldsymbol{w}$ …… (2-4)

and

$var(w_{s'}|\boldsymbol{w}, Y) = \sigma_w^2\left(1 - \delta_j'\Sigma_w^{-1}\delta_j\right)$ ……... (2-5)

respectively, where δ_j is the vector of the effect of distance between new sites and C stock observation locations where each element $\delta_{ij} = f\left(d_{s_i s_j}, \phi\right)$.

2.5.4. Hyperprior specification

We performed Bayesian hierarchical modelling using the *spBayes* package (Finley, Sudipto and Carlin, 2007) in the R Statistical and Computing environment (R Core Development Team, 2008). As articulated in Gelfand (2012), the Bayesian approach treats the vector of model parameters, $\boldsymbol{\theta} = \boldsymbol{\beta}, \sigma^2, \phi, \tau^2$ as random and mutually independent variables and assigns prior distributions to the parameters. By taking $p(\boldsymbol{\theta})$ to be a prior distribution and $p(\boldsymbol{\theta}|y, X)$ to be the posterior distribution of the observed C stock model parameters, we computed the posterior distribution of model parameters, $\boldsymbol{\theta}$, as follows:

$p(\boldsymbol{\theta}|\boldsymbol{y}, X) \propto p(\boldsymbol{\theta})\ x\ N(\boldsymbol{w}|0, \Sigma_w)\ x\ N(\boldsymbol{y}|X^t\boldsymbol{\beta} + \boldsymbol{w}, \Sigma_\varepsilon)$ …… (2-6)

A summarised modelling framework for the hierarchical specification shown from Eqn 2-1 to Eqn 2-7 is illustrated in the flow chart illustrated in **Figure 2-3**. We utilized Eqn 6 to quantify uncertainties in model parameters. Predicted C stock values were then derived from the predictive distribution of C stock at unsampled locations, s_0 using Eqn 2-7, which quantifies uncertainties in C stock predicted values.

$p(y_0|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{x}_0) \propto \int p(y_0|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{x}_0)p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X})d\boldsymbol{\theta}$ ……... (2-7)

Where;

y_0 denotes the predicted C stock at a location s_0 and x_0 are the covariate values at location s_0. We assigned normal prior for the overall mean of measured C stock and a multivariate normal prior for the regression coefficients. We also classified model parameters into two categories of (1) regression coefficients consisting of $\boldsymbol{\beta} = \left(\beta_{\boldsymbol{0}}, \beta_{\boldsymbol{1}}, \ldots, \beta_{\boldsymbol{p}}\right)'$ in the mean function and (2) the partial sill (σ^2), nugget (τ^2) and range (ϕ) in the variance covariance matrix. In respect of our sampled data, we specified two sets of priors for hierarchical modelling of C stock using predictors from Sentinel-2 and Landsat-8 data. An inverse gamma prior distribution was assigned for the data and measurement error variance whilst a uniform prior was assigned for the spatial decay parameter, ϕ as defined by $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ below:

$p(\boldsymbol{\theta}_1) = \boldsymbol{Unif}(\phi|0.38, 0.0012)\ x\ \boldsymbol{IG}(\sigma^2|0.75, 1.76)\ x\ \boldsymbol{IG}(\tau^2|0.1, 1.76)\ x\ \boldsymbol{MVN}\left(\beta\middle|\boldsymbol{0}, \Sigma_{\boldsymbol{\beta}}\right)$

$p(\boldsymbol{\theta}_2) =$

$\boldsymbol{Unif}(\phi|0.38, 0.0012)\ x\ \boldsymbol{IG}(\sigma^2|0.071, 0.021)\ x\ \boldsymbol{IG}(\tau^2|0.071, 0.0028)\ x\ \boldsymbol{MVN}(\beta|\mathbf{0}, \Sigma_{\boldsymbol{\beta}})$

Where;

$p(\boldsymbol{\theta_1})$ and $p(\boldsymbol{\theta}_2)$ are prior specifications for Sentinel-2 and Landsat-8 derived covariates, respectively.

In both sets of priors ($\boldsymbol{\theta}_1$ and $\boldsymbol{\theta_2}$), the shape parameters were chosen in order to guarantee that the prior distribution has finite mean and the distribution is adequately vague that it would not have finite variance (Demirhan and Kalaylioglu, 2015). In the same vein, we adopted scale parameter values in order to expresses the preference that the prior mean of $\sigma_{\boldsymbol{\varepsilon}}^2$ is less than that of $\sigma_{\boldsymbol{w}}^2$. This is vindicated from the fact that we assumed the measurement error variance (nugget, $\sigma_{\boldsymbol{\varepsilon}}^2$) to be much smaller than the data variance, $\sigma_{\boldsymbol{w}}^2$. The spatial dependence parameter ϕ, was assigned a uniform prior with distance over the spatial range of the study domain.

In order to get results from the Bayesian hierarchical modelling, a Metropolis-Hastings algorithm typically used in the *spBayes* R package for Markov Chain Monte Carlo (MCMC) methods was employed (Gelman, 2006). An algorithm of one chain each with 20000 MCMC iterations was specified for the posterior densities of the parameters in which 15000 were discarded as burn-in.

2.6. Variogram modelling and exploratory analysis

We carried out spatial exploratory analysis of the sampled C stock observation data in order to get some insights regarding the choice of scale values on the prior probability distribution of the spatial covariance parameters. Spatial exploratory analysis is an important step towards assessing the strength of the spatial covariance structure of the modelled regionalized variable (Stoyan, 2014; Sahu, 2022; Pascual, Tupinambá-Simões and de Conto, 2022). We also streamlined the modelled C Stock data to the normality assumption through the Box-Cox transformation approach (Box and Cox, 1982). C Stock data was therefore transformed according to the Box-Cox (1982) formula illustrated in Eqn 2-8.

$$\left\{ Y(\lambda) = \begin{matrix} (Y^{\lambda} - 1)/\lambda \quad if\ \lambda \neq 0 \\ log\, Y\ if\ \lambda = 0 \end{matrix} \right. \quad \ldots\ldots \text{(2-8)}$$

Y is the C Stock observation and λ is the transformation parameter.

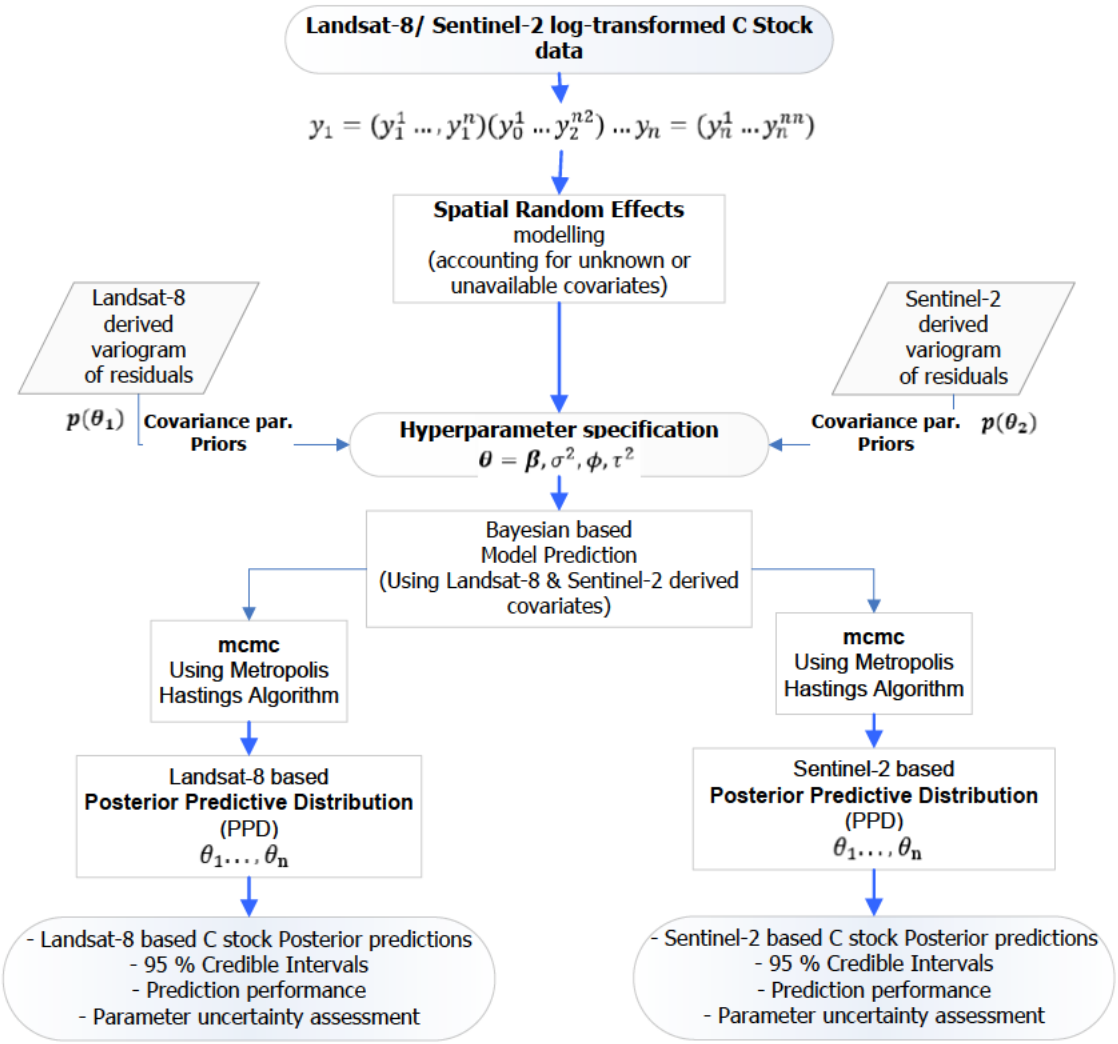


Figure 2-3: Hierarchical Bayesian geostatistical modelling illustrated by Eqns. 1-7 for C stock prediction

Developments made by Box and Cox (1964) regarding the transformation parameter in Eqn 8 are of the normal theory linear model as in Eqn 2-9,

$y(\lambda) = \boldsymbol{X}\boldsymbol{\beta}(\lambda) + \varepsilon,$ ..... (2-9)

where $\boldsymbol{X}$ is an $n\ x\ p$ vector of predictors, $\boldsymbol{\beta}(\lambda)$ is a $p\ x\ 1$ vector of unknown parameters while the standard deviation of the independent errors $\varepsilon_i (i = 1, \dots, n)$ is $\sigma(\lambda)$.

2.7. Model predictions and uncertainty assessment

We tested the statistical significance of the predictor variables by considering the 2.5% and the 97.5% percentiles of the posterior samples. Independent variables in the Landsat-8 and Sentinel-2 based hierarchical models were deemed statistically significant if their 95% Credible Intervals (CIs) excluded zero. Uncertainty of C stock predictions was assessed by employing the method of Hengl et al. (2004). Average predictions of C stock for each grid cell (10 000m^2) were displayed using a colour ramp alongside the associated uncertainty (95% Credible

Interval Widths (CIW)). Posterior predictive distribution CIWs was therefore utilized to assess the precision of model predictions. Highly uncertain predictions have wider 95% CIWs (Babcock *et al.*, 2016). As provided in literature including Babcock *et al.* (2016) and Babcock *et al.* (2018), highly uncertain (less precise) predictions have wider CIW.

2.8. Model comparison and validation

Three different Bayesian hierarchical models constructed from Landsat-8 and Sentineal-2 derived predictors were compared as illustrated in Eqns 2-10, 2-11 and 2-12. The models compared were the independent error model (simple multiple linear regression), the spatial intercept only model and the spatial model, respectively

$$\boldsymbol{y} = \boldsymbol{x}\beta + \varepsilon \quad \text{(2-10)}$$

$$\boldsymbol{y}(s) = \boldsymbol{x}(s)\beta + \varepsilon(s) \quad \text{(2-11)}$$

$$\boldsymbol{y}(s) = \boldsymbol{x}(s)\beta + w(s) + \varepsilon(s) \quad \text{(2-12)}$$

Posterior samples of the spatial and spatial intercept only models' predictions and parameters are gathered via a MCMC algorithm and composition sampling. We further tested the predictive efficacy of each of the candidate models using a k-fold cross validation algorithm, which proceeds by randomly splitting the 191 C stock observations into ten almost equally seized segments (Duchene *et al.*, 2016). Log-transformed C stock for the holdout data block was successively predicted given model parameters derived using data in the remaining nine blocks.

Mean Absolute Error (MAE), Root mean squared error (RMSE) and other validation metrics were calculated from the holdout posterior predicted means and observed C stock data for each of the models (Green, Finley and Strawderman, 2020). We considered the model with the lowest k ($k = 10$)-*fold* RMSE and MAE as the "best" predicting model. Model convergence was assessed using graphical displays in the form of trace plots of the estimated model parameters (Jackman, 2000).

2.9. Results

This section provides specific information on the results of the study, starting with the results of an exploratory analysis.

2.9.1. Exploratory analysis

An objective approach for arriving at the appropriate transformation parameter made through the Box-Cox transformation as in Eqn 2-8 for the C stock data led to a log-transformation as λ was estimated to be 0.012 as illustrated in **Figure 2-4**.

A transformation parameter of 0.5 entails a square root transformation, a parameter of 1 entails no transformation whilst a transformation of 0 supports a logarithmic transformation as illustrated in Figure 3 (Box and Cox, 1964).

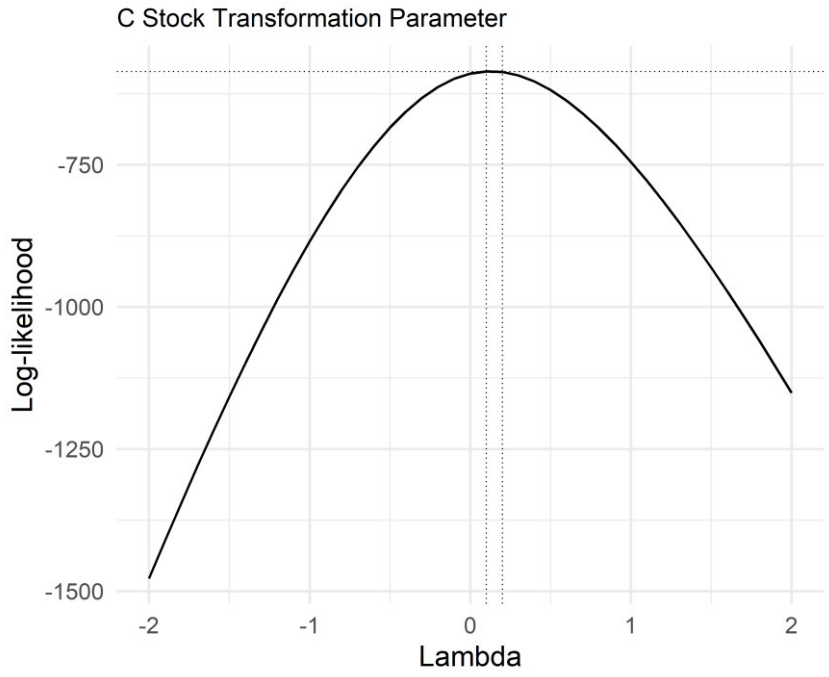


Figure 2-4: Box-Cox Transformation for C stock.

As detailed in Section 2, parameter transformation from this preliminary analysis governed the treatment of the outcome variable in all the subsequent analysis involving treatment of the variable under different satellite derived auxiliary information. Log and square root transformation of outcome variables are common data transformation parameters in natural resources modelling (Chinembiri *et al.*, 2013; Babcock *et al.*, 2016).

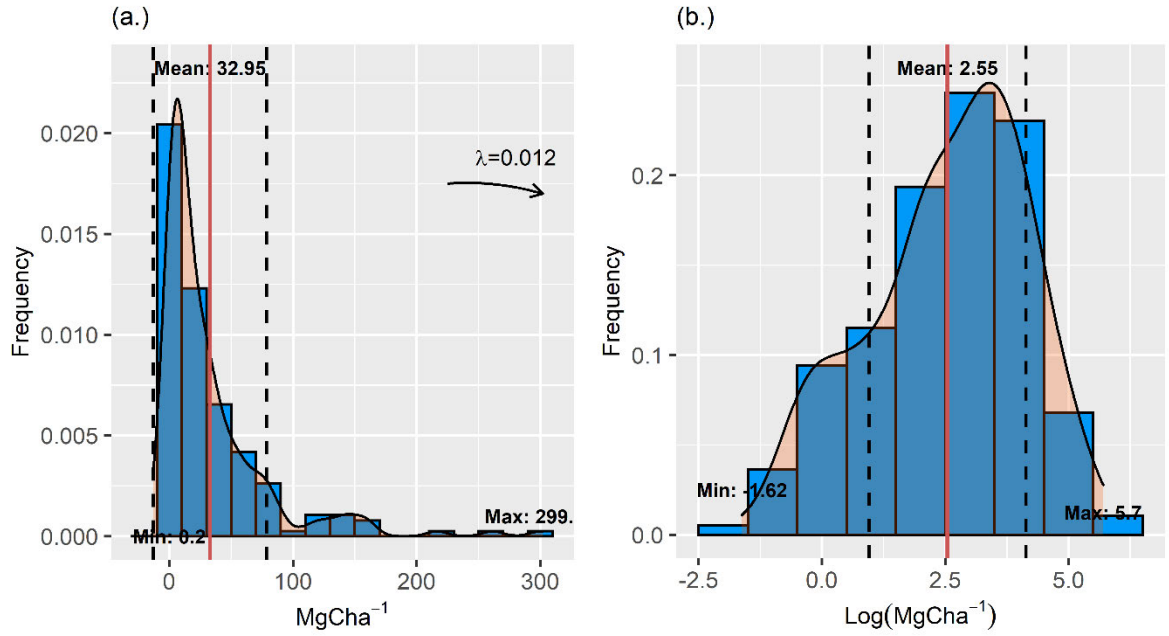


Figure 2-5: Logarithmic transformation for the C stock data (with normal overlay)

The modelled C stock data employed in the Bayesian hierarchical framework displayed a positively skewed distribution as illustrated in Figure 5. Positively skewed C stock data implies that the lower bounds of the sampled data are relatively lower than the rest of the data. A Box Cox transformation with a transformation parameter of 0.012 ($\lambda = 0.012$) entailed a log-transformation for the C Stock variable which led to an approximately normally distributed C Stock data as illustrated in **Figure 2-5 (b)**. The σ_{ε}^2 and the σ_{w}^2 variance hyperpriors for the Sentinel-2 and Landsat-8 based spatial models of the log-transformed C stock data were set according to the variogram of residuals of predictors derived from the respective medium resolution satellite sensors depicted in **Figure 2-6(a)** and **Figure 2-6(b)** respectively.

As illustrated in Figure 6, the σ_{ε}^2 and the σ_{w}^2 variance parameters of the C stock hierarchical Bayesian models derived from the respective medium resolution sensors show greater influence of Sentinel-2 derived predictors on C stock. Subsequent Bayesian hierarchical modelling of the response variable followed the trajectory of scale parameters illustrated in **Figure 2-6** as these provided a constraint on the prior probability distribution function of the parameter space.

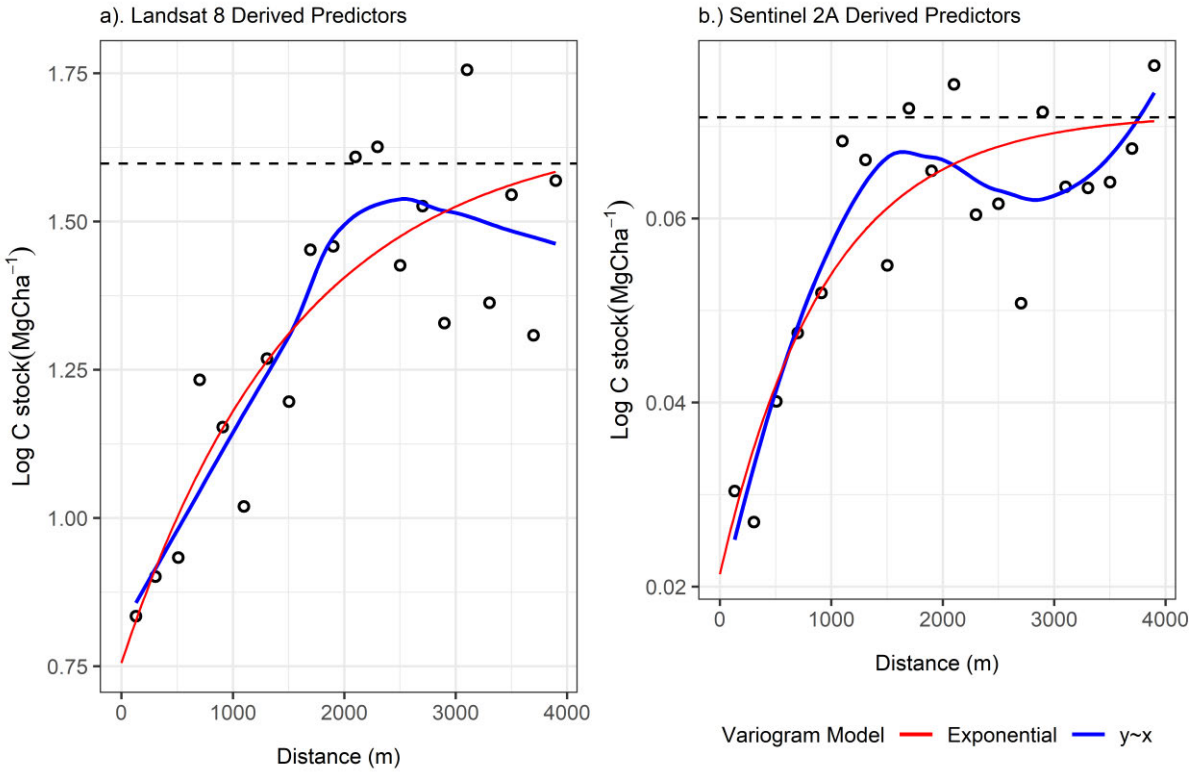


Figure 2-6: Landsat-8 and Sentinel-2 derived variogram of residuals for the modelled C stock. The black dotted line is the asymptote of the theoretical variogram model

2.9.2. Effects of Landsat-8 and Sentinel-2 predictors on model prediction

Covariate information in the form of $NDVI$, $SAVI$, EVI and distance to the nearest settlements from Landsat-8 and Sentinel-2 employed in the hierarchical modelling of C stock showed distance to nearest settlements and $NDVI$ to be significant predictors of C stock. The two predictors are significantly different from zero as their 95% CIs exclude zero. The posterior distribution of predictor coefficients markedly differed between Landsat-8 and Sentinel-2 based spatial models. However, amongst the modelled predictors, $NDVI$ and distance to settlements ($DIST$) were statistically significant in both the Landsat-8 and Sentinel-2 based spatial models as shown in **Table 2-2**.

The carbon stock model derived from Landsat-8 predictors implies a weaker spatial correlation with an effective range estimated as 1764 meters and a 95% CI of (1578.9, 2307.7) meters whilst C stock model developed from the Sentinel-2 independent variables shows a stronger correlation with an effective range of 2142.9 meters with a 95% CI of (2000, 2307.7) meters directly estimated from the CI of the spatial decay parameter, ϕ. However, estimates of the range of spatial dependence from both medium resolution sensor C stock-based models seem plausible as the maximum distance between data locations within the study domain is 2463

meters. As expected, we observe that the spatially structured variance, σ_w^2, is higher than the micro-scale variance in both models (Diggle and Ribeiro Jr, 2007).

Table 2-2: Sentinel-2 and Landsat-8 derived predictors of C stock

Parameter	Lansat-8 OLI C stock Model				Sentinel-2 MSI C stock Model			
	Mean	**s.d**	**2.5%**	**97.5%**	**Mean**	**s.d**	**2.5%**	**97.5%**
Intercept	-2.75	1.01	-4.73	-0.82	-2.41	0.32	-3.04	-1.79
NDVI	2.67	0.99	0.77	4.60	4.97	0.24	4.53	5.45
SAVI	-0.67	0.69	-2.05	0.67	-0.52	0.36	-1.22	0.18
EVI	-0.52	0.53	-1.56	0.53	-0.002	0.097	-0.18	0.20
DIST*	1.47	0.33	0.79	2.09	0.78	0.13	0.52	1.05
σ_w^2	1.24	0.28	0.71	1.78	0.072	0.014	0.045	0.091
σ_ε^2	0.31	0.12	0.079	0.55	0.0047	0.0030	0.0006	0.012
ϕ	0.0017	0.0002	0.0013	0.0019	0.0014	0.0001	0.0013	0.0015

*DIST** is the distance to the nearest settlements used in the Bayesian hierarchical modelling framework

2.9.3. C stock prediction from medium resolution sensor derived covariates

Fitted models with spectral independent variables from the Sentinel-2 and the Landsat-8 sensors were employed in the prediction of C stock at unvisited locations in the study domain. Since spectral variables of $NDVI, SAVI, EVI$ in addition to distance to settlements were on gridded raster with 10 000 m^2 resolution, the predicted log-transformed C stock represents averaged values in every pixel. MCMC sampling from the predictive distribution of the fitted models resulted in a total of 20 000 predicted samples for each grid. Consequently, predicted samples for each pixel provide a comprehensive account of the uncertainty in the predicted C stock values. The specific expressions and ratios of vegetation indices represent properties of green vegetation much better than single bands (Baloloy *et al.*, 2018).

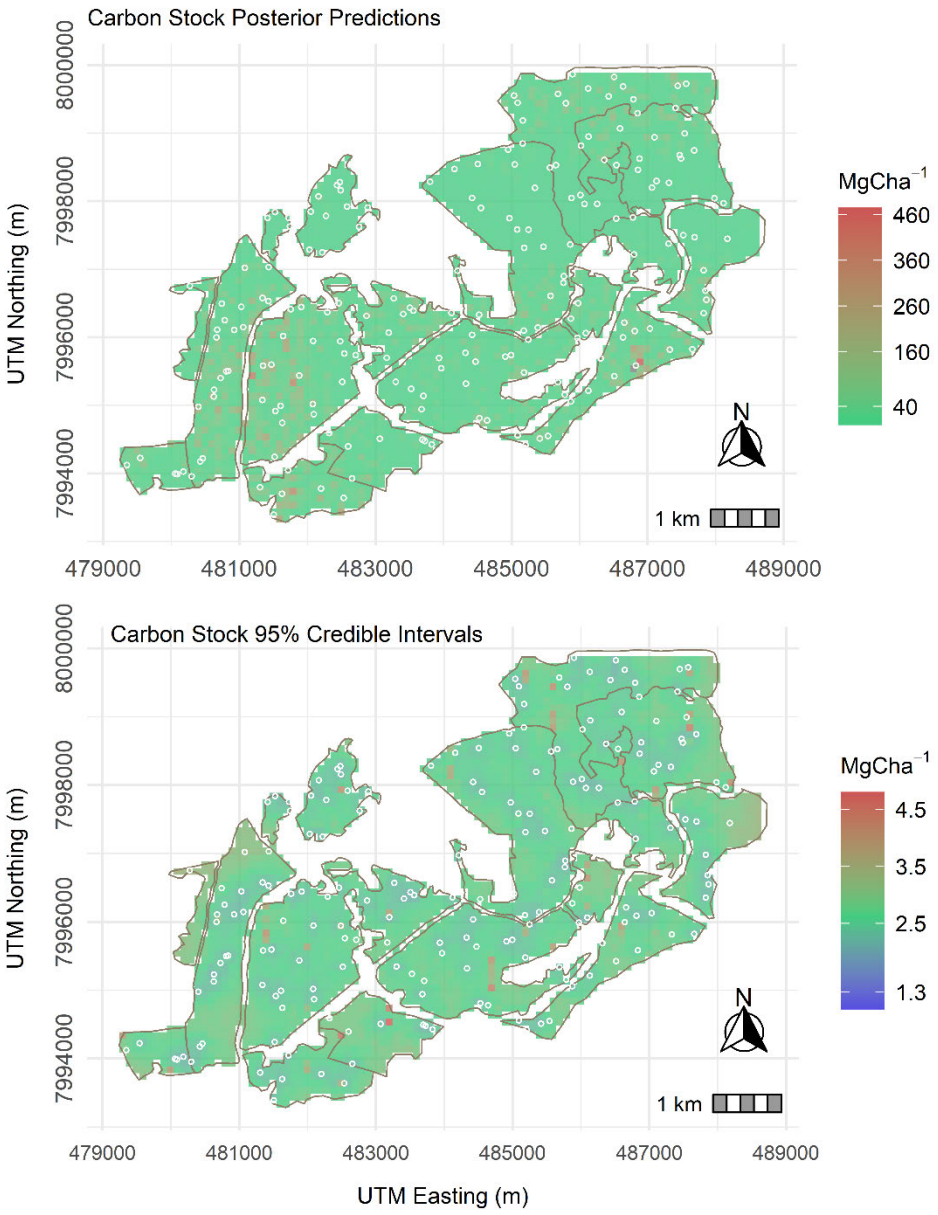


Figure 2-7: Landsat-8 based C stock posterior mean and posterior standard deviation. Points represent locations of sampled C stock

As illustrated in **Figure 2-7**, the 95% CIs for C stock predictions derived from Sentinel-2 spectral variables are much more plausible than predictors supplied from Landsat-8 spectral covariates. The indispensability of *NDVI* as an index well correlated with vegetation biophysical properties, amongst them, leaf area index (*LAI*), and chlorophyll is much established (Baugh and Groeneveld, 2006). Sentiel-2 based C stock predictions have lower uncertainty compared to Landsat-8 OLI C stock aided predictions. This fact is supported by the evidence of the 95% posterior predictions illustrated in **Figure 2-8**, showing C stock 95% CIs orders of magnitude lower than the predicted estimates in Sentinel-2 than those displayed in Landsat-8 based estimates. The accompanying credible intervals for both Landsat-8 and Sentinel-2 C stock-based predictions are highly certain. This is because the entire part of the

study area has 95% CIs orders of magnitude lower than the C stock estimates (**Figure 2-7** and **Figure 2-8**).

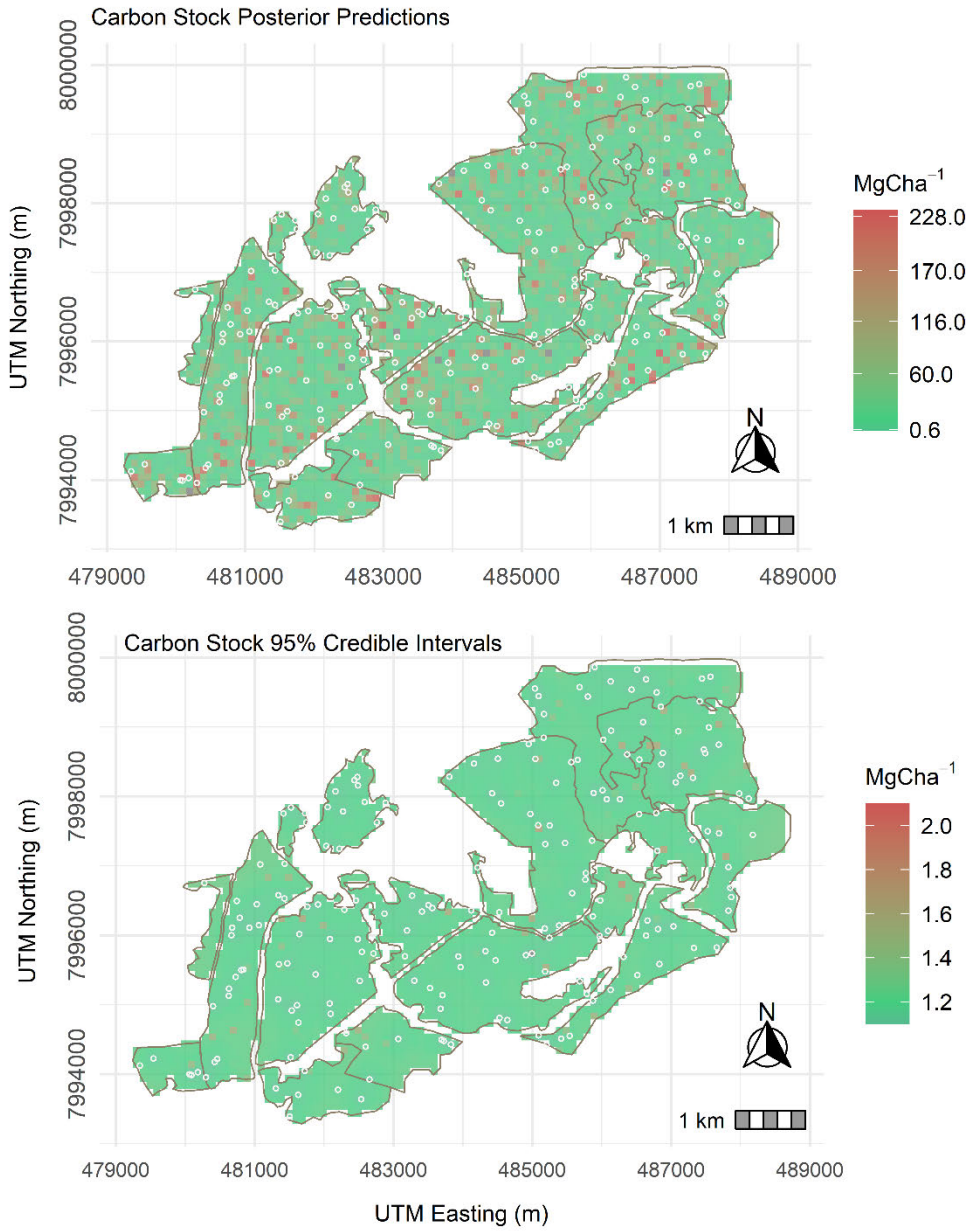


Figure 2-8: Sentinel-2 based C stock posterior mean and posterior standard deviation. Points represent locations of sampled C stock

It is observed that predictions in the central northern region of the studied region (**Figure 2-7**) in Lot 75A of Nyanga Downs are much lower in Landsat-8 than in Sentinel-2 based C stock predictions. We attribute underprediction of C stock in this region in Landsat-8 based predictions to the finer spatial and spectral attributes in Sentinel-2 data in the near-infrared and visible portions of the electromagnetic spectrum (Sovdat *et al.*, 2019; Wang *et al.*, 2020; Ahmed, Atzberger and Zewdie, 2022). The enhanced spectral and spatial resolution of Sentinel-2 also justifies the shorter 95% CIW displayed by the Sentinel-2 based C stock predictions (1.02 - 1.82) $MgCha^{-1}$ compared to Landsat-8 based C stock predictions (2.23 -

4.50) $MgCha^{-1}$. The results from both sensors look attractive compared to AGB reported in literature where Jiang *et al.* (2021) established an average AGB RMSE of 40.92 $MgCha^{-1}$ in northeast China whilst Dang *et al.* (2019) reported a RMSE of 36.67 $MgCha^{-1}$ in Vietnam using Random Forest (RF). To add on, Takagi *et al.* (2015) utilised LiDAR for predicting forest biomass in Hokkaido, Japan, and established a biomass RMSE prediction of 19.10 $MgCha^{-1}$. Differences in the prediction accuracy between the reported results in literature and our study can also be explained by the differences in forest density as the aforementioned studies have been undertaken in tropical and subtropical rainforest biomes.

2.9.4. Performance assessment of C stock prediction models

Table 2-3 presents the cross-validation metrics utilised in the assessment of the predictive efficacy of the Bayesian hierarchical models emanating from the Sentinel-2 and the Landsat-8 derived spectral predictor variables. Different predictive Bayesian hierarchical models from the two medium resolution satellite sensors, with varying levels of richness (independent error, spatial intercept only and the spatial models) favour the C stock spatial predictive model from the Sentinel-2 data (**Table 2-3**).

In addition to the application of RMSE, MAE and CRPS as model validation and performance criteria, coverage (CVG) is employed as an additional criterion of model performance. The Sentinel-2 based spatial C stock predictive model (**Table 2-3**) appears as the top performing model in terms of the RMSE (1.16 $MgCha^{-1}$), DIC (-554.7), MAE (1.11 $MgCha^{-1}$) and coverage (94.7%) thereby making it the best performing model on the premise of the aforementioned validation criteria. The model affords predictive ability little short of the benchmark nominal 95% coverage level (Guhaniyogi and Banerjee, 2019; Sahu, 2022) . We define coverage in this study as defined and interpreted in other studies (Guhaniyogi and Banerjee, 2019) as the percentage of reliably predicted C stock in the study domain. On the other hand, the Landsat-8 based C stock predictive model has 85% (**Table 2-3**) coverage for the 95% prediction intervals and higher RMSE and MAE, making it less favourable as the best performing model within the remote sensing data driven Bayesian hierarchical framework.

Figure 2-9 (a) and **9 (b)** illustrates the scatterplots of observed C stock against the predicted C stock alongside the 95% intervals for both the Sentinel-2 and the Landsat-8 C stock based predictive models. Evidence of the Sentinel-2 driven C stock predictive model performing better than its Landsat-8 C stock-based counterpart is clear from the scatter plot of the model in **Figure 9 (b)**. This fact is further bolstered by the Relative Root Mean Square Error (RRMSE)

and other validation statistics including the MAE for Sentinel-2 based C stock prediction models illustrated in **Table 2-3.**

Table 2-3: Validation metrics for C stock Bayesian hierarchical models

Model Selection criterion	Landsat 8 Derived Predictors			Sentinel-2 Derived Predictors		
	Independent Error Model	**Spatial Intercept only Model**	**Spatial Model**	**Independent Error Model**	**Spatial Intercept only Model**	**Spatial Model**
RMSE (MgCha^{-1})	3.35	2.64	2.69	1.32	3.25	1.16
RRMSE	0.11	0.09	0.09	0.04	0.11	0.04
MAE (MgCha^{-1})	2.44	1.70	1.79	1.25	2.16	1.11
CRPS (MgCha^{-1})	2.18	1.46	1.48	1.85	1.75	1.13
CVG (%)	91.7	85.4	85.4	95.8	89.6	94.7
DIC	210.2	56.9	43.1	-244.4	37.8	-554.7

2.9.5. MCMC convergence diagnostics

Independent variable coefficients displayed as trace plots for the Landsat-8 and Sentinel-2 based C stock predictions are depicted in **Figure 9 (a)** and **Figure 9(b)**.

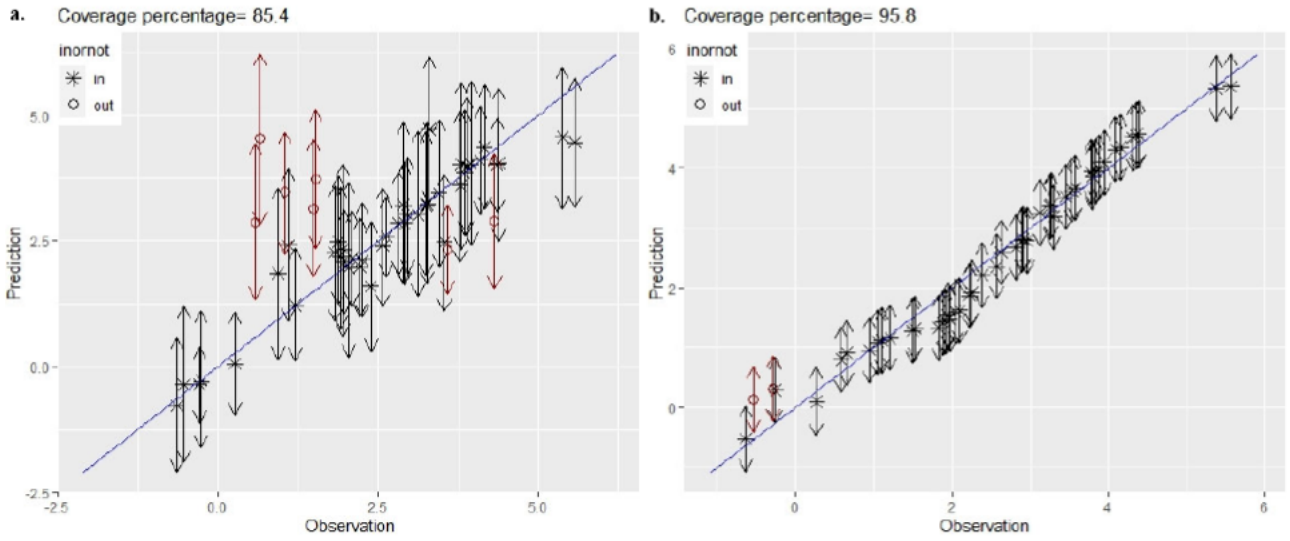


Figure 2-9: (a). Spatial model of the Landsat-8 C stock-based predictions against observed C stock. **(b).** Spatial model of the Sentinel-2 C stock-based prediction against observed C stock alongside 95% intervals.

We utilized the trace plots for the aforementioned models as visual assessment of chain convergence. It is clear from the plots that the extent, N_0, of the burn-in period was adequately large to allow Markov chains to converge to the smooth distribution at the end of the burn-in period (Gelfand *et al.*, 2004; Finley, Sudipto and Carlin, 2007).

For the sake of presentation and improvement of plot aesthetics, we thinned the chains so that only every just 10^{th} iteration is kept. The thinning process speeds up calculations through the reduction of the Monte Carlo sample size and also enhance the display and presentation of the plots**.** In both models, the chains seem to have converged to the smooth distribution at the end of the burn-in and hence, to mix fairly well over 5,000 iterations (2000 after thinning). It can be deduced from the respective presentations (**Figure 10** and **Figure 11**) that the length, N_0 of the MCMC chain should be sufficiently large so that moments and quantiles calculated from the MCMC samples are precise estimates of the resulting characteristics of the posterior (Beloconi and Vounatsou, 2020). A visual assessment of **Figure 9** and **Figure 10** point to a converged sample with MCMC sufficiently mixing well.

2.10. Discussion

This section discusses the results of the study with regards to the performance of the two satellite sensors and their influence on C stock prediction uncertainty.

2.10.1.Landsat-8 and Sentinel-2 derived C stock predictors

The current study juxtaposed findings from the Sentinel-2 and the Landsat-8 medium resoultion passive satellite sensors in the predictive modelling of C stock using a Bayesian hierarchical approach. We beased the modelling and prediction framework on a set of vegetation indices together with antropogeninc factors as predictors of C stock in a disturbed plantation ecosystem in Zimbabwe. Bordoloi *et al.* (2022) and Somvanshi and Kumari (2020) have established $NDVI$, $SAVI$ and EVI to be significant predictors of above ground biomass when using Landsat-8 and Sentinel-2. However, no Bayesian hierarchical approach has yet employed similar predictors from the two medium resolution sensors and established their influence on C stock prediction estimates in a forest plantation set-up.

Apart from the utilization of medium resolution derived covariates, the significance of the anthropogenic predictor variable in the form of distance to nearest settlements in this investigation is significant. As depicted in **Table 2-1**, distance to nearest settlements and $NDVI$ is a significant predictor of C stock in disturbed plantation forest in models estimated from both medium resolution remote sensing sensors. This fact draws logic and support from two main fronts. The plausibility of distance to nearest settlements as a predictor of C stock rationalises from the circumstances prevailing in the plantation forests after the agrarian land reform of 2000. Secondly, studies undertaken elsewhere, including Chinembiri *et al.* (2013) and Nwobi and Williams (2021), have shown distance to settlements as a significant predictor of C stock in perturbed and related environments.

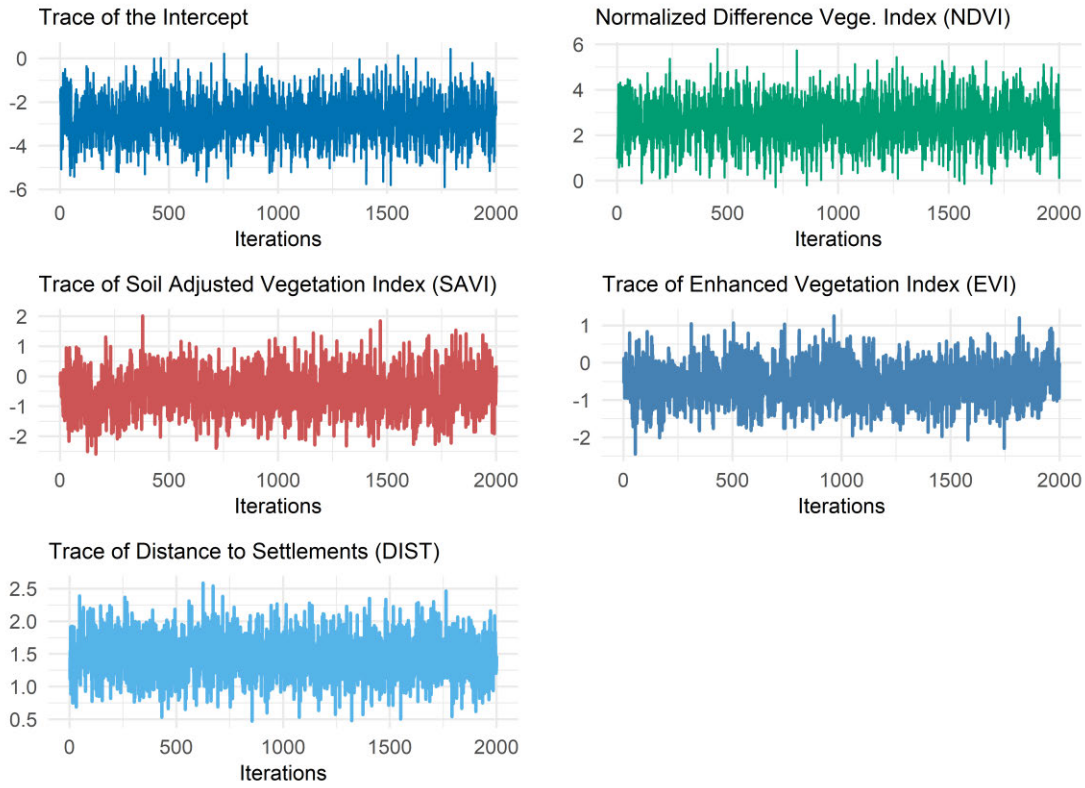


Figure 2-10: Lansat-8 based C stock prediction MCMC trace plots

Previous studies including Finley, Banerjee and McRoberts (2008) proffer interesting exposition on forest biomass over larger scales. The authors mapped AGB using a suite of spatial regression models with Landsat as the source of auxiliary information where residual structure is accounted for by a Gaussian process that has the capacity to account for long and short range spatially structured dependence. Their findings demonstrated effective long-range dependence to the tune of 8000 meters with associated 95% CIs of 4000 to 11000 meters. Similar improvements in spatially structured dependence is also reported in Datta *et al.* (2016) work when modelling AGB in the conterminous United States.

2.10.2. Parameter uncertainty and model performance of new generation sensor-based C stock models

An investigation of the AGB distribution of a mangrove forest in Vietnam employing a combination of remote sensing data and Artificial Neural Network (ANN) gave mangrove AGB predictions that ranged from 6.53 to 368.2 $Mgha^{-1}$ in 2000 and from 13.75 to 320.3 $Mgha^{-1}$ in 2020, respectively (Do *et al.*, 2022).

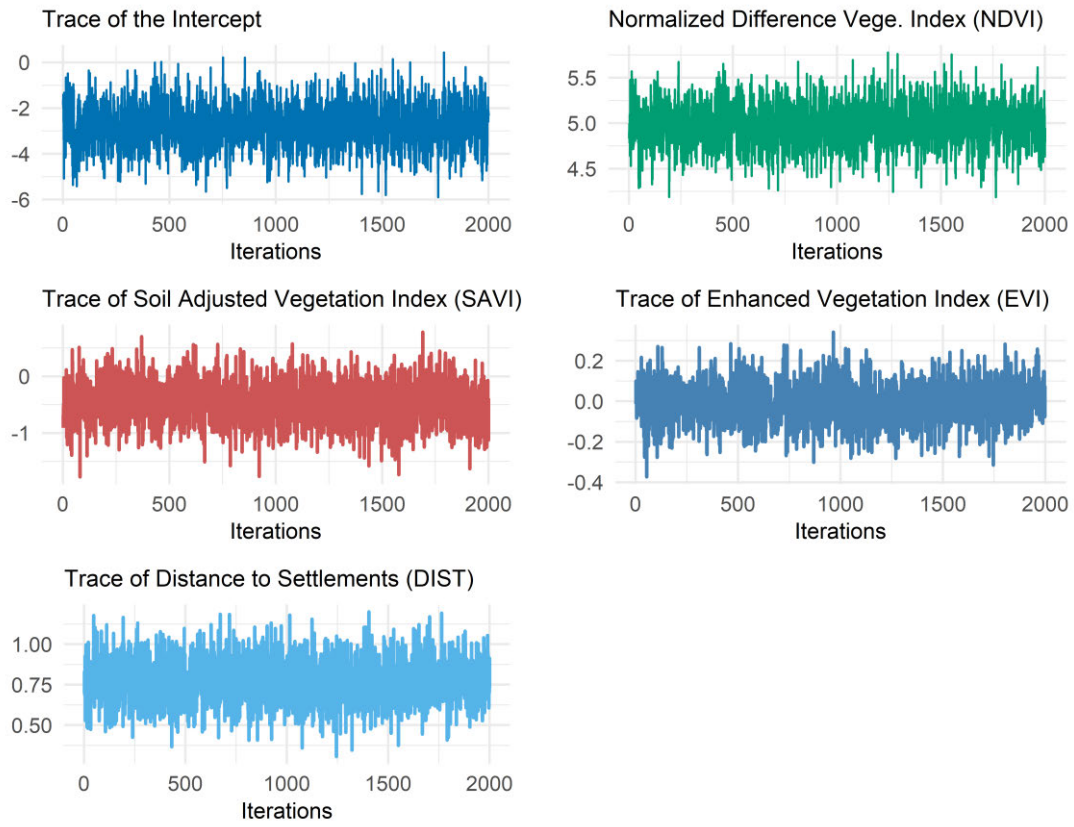


Figure 2-11: Sentinel-2 based C stock prediction MCMC trace plots

AGB predictions made by Fararoda *et al.* (2021) using machine learning methods including AdaBoost, random decision forest, multilayer neural networks and Bayesian ridge regression endorsed random forest and AdaBoost as the best performing methods. The type of sensor and the method of prediction used have great influence over AGB model performance. For instance, Fassnacht *et al.* (2014) compared LiDAR and hyperspectral data for modelling AGB and concluded that LiDAR data fused with Random forest models offered the best AGB model performance.

In terms of the accuracy of prediction, it is vital to compare the results of the present study with recent publications in literature from similar and related terrestrial biomes. Jiang *et al.* (2021) established an average AGB RMSE of 40.92 $MgCha^{-1}$ in northeast China whilst Dang *et al.* (2019) reported a RMSE of 36.67 $MgCha^{-1}$ in Vietnam using Random Forest (RF). Takagi *et al.* (2015) utilised LiDAR for predicting forest biomass in Hokkaido, Japan, and established a biomass RMSE prediction of 19.10 $MgCha^{-1}$. Differences in the prediction accuracy between the reported results in literature and our study can be partly explained by the differences in forest density and largely by the differences in the modelling approaches between results reported in literature and our methodology. First, the aforementioned studies are likely to have significant and higher AGB density than in Zimbabwe as they were undertaken in tropical and

subtropical rainforest biomes. Secondly and most importantly, the present study uses a hierarchical Bayesian approach which has the advantage of having access to the entire posterior predictive distribution (Goulard and Voltz, 1992).

Location based C stock uncertainty maps presented in this study are needed for monitoring ecosystem health and areas of interest in the plantation forests for interventions during times of extreme weather, and in times of disease outbreaks. Such information provides timely decision making and management by forest practitioners. With accuracies for both Landsat-8 and Sentinel-2 based models exceeding accuracies established in other similar environments (Takagi *et al.*, 2015; Jiang *et al.*, 2021; Ahmed, Atzberger and Zewdie, 2022; Fan, Wang and Yang, 2022; Xiong and Wang, 2022), the application of results of the present study can therefore be utilised for comprehensive and location-based ecological interventions. The Sentinel-2 driven C stock predictive model yielded the shortest CIWs, implying more precise C stock predictions than the Landsat-8 based C stock estimates. Hence, the usefulness of Landsat-8 based C stock predictive model is less attractive as it cannot match the standards and qualities inherent in the improved Sentinel-2 for similar modelling framework as the current study. On the other hand, the Landsat-8 based C stock predictive model gave slightly higher RMSE (2.69 $MgCha^{-1}$) and slightly lower coverage (85.4%) than its Sentinel-2 based C stock spatial predictive counterpart. This also means monitoring and conservation for extensive forest ecosystems can be done with greater accuracy and precision, thereby improving forest carbon accounting, especially in inaccessible areas where samples cannot be easily taken. Higher prediction accuracy means mapping and accounting for natural resources can be achieved with great accuracy and save forest species from extinction in the event of disease outbreaks and extreme environmental conditions (Do *et al.*, 2022; Ahmed, Atzberger and Zewdie, 2022; Fan, Wang and Yang, 2022).

As attributed in Frampton *et al.*, 2013 and Gerald *et al.*, 2017, the posterior predictive distribution of Sentinel-2 based C stock predictions are more precise than their Landsat-8 based counterparts as they have far shorter 95% CIWs. This implies that highly certain predictions of C stock in disturbed plantation ecosystems can better be made with Sentinel-2 derived vegetation indices ($NDVI$) and distance to nearest settlements than with similar spectral and anthropogenic variables derived from the Landsat-8 based C stock hierarchical model. The quality of predictors deriving from the tested medium resolution sensors is a critical point of discussion in this case (Ahmed, De Marsily, 1987). More refined spatial and spectral resolution (10 m, 20 m) of Sentinel-2 derived spectral indices offer an opportunity for enhanced and accurate monitoring of forest resources (Mutanga, Dube and Ahmed, 2016; Somvanshi and

Kumari, 2020). Our results fall in the same realm as those of Korhonen *et al.* (2017), Astola *et al.* (2019) and Jha *et al.* (2021) due to the differences in spatial and spectral properties of the two sensors. As such, despite the departure in the modelling methodology used by the present study, comparison of the predictive efficacy of Sentinel-2 based C stock model still outperforms its Landsat-8 counterpart and other machine learning methods reported in literature. This can be attributed to the Bayesian hierarchical modelling approach utilised in the present study, which has the strength of incorporating multiple sources of uncertainty into the prediction modelling framework (Azevedo, 2021). Furthermore, Dang *et al.* (2019) utilised Sentinel-2 together with field based biomass data using random forest (RF) and machine learning algorithm for estimating AGB in Vietnam. Results employing the RF algorithm together with Sentinel-2 derived auxiliary information resulted in more accurate predictions of AGB compared to the machine learning methods. Consequently, findings in literature using Sentinel-2 as a data source have shown the former to be superior than other data sources, thereby making Sentinel-2 medium resolution imagery a data source of choice for forests management and ecological assessment.

Wu *et al.* (2016) and Xiong and Wang (2022) compared artificial intelligent methods for AGB estimation based on Landsat imagery and established Random forest to be the best performing method with a RMSE of 26.44 tons/ha over the other methods like support vector regression, k-nearest neighbour and stochastic gradient boosting. As noted in the present study, hierarchical Bayesian geostatistical methods offer superior C stock predictions than the other methods in literature. Nevertheless, hierarchical geostatistical methods cannot be easily adopted and utilised by forest practitioners for ecological monitoring. Their complexity and the lack of readily available easy to use application packages greatly limit their adoption and application in ecological monitoring. The fusion of medium and high resolution satellite sensors greatly enhance the accuracy of predicting forest biomass (Fararoda *et al.*, 2021). However, differences in model performance reported in literature and the present study are particularly significant as they suggest the selection of appropriate statistical and modelling methods to be more effective than putting more effort and resources in field data collection. Previous research comparing Sentinel-2 and Landsat-8 sensors noted improvements in the spatial and spectral capabilities of Sentinel-2 in distinguishing rangeland management practices (Sibanda *et al.*, 2017), estimating Leaf Area Index (LAI) (Korhonen *et al.*, 2017) and to enhance the quality of classification of built-up areas (Pesaresi *et al.*, 2016). The inclusion of three bands within the red-edge portion of the electromagnetic spectrum is given as the rationale for Sentinel-2's superior performance in all the comparative studies**.** Hence, the performance of the Sentinel-2 driven C stock prediction model in this research can be traced to

the additional and narrower band channels, in particular, the red-edge band coupled with improved spatial resolution. However, studies that have utilized Landsat-8 in isolation to Sentinel-2 for predicting AGB, have noted significant improvement in the spectral capabilities of the imagery.

Predictions in the studied region of Zimbabwe show a much reduced density of C stock compared with other tropical and sub-tropical forest ecosystems in similar environments like the Indian forests, with average biomass density of more than 420 $Mgha^{-1}$ (Fararoda *et al.*, 2021). Implementation of conservation practices needed for ecological restoration of the disturbed plantation forest ecosystems therefore becomes an urgent matter. Depressed C stock concentration in these environments may also imply that the plantation forest may take much longer to return to their former state following perturbations from farming and gold panning activities of settlers in the plantations.

2.11. Conclusions

The objective of our investigation was to develop and test the performance of a Bayesian hierarchical modelling approach using C stock as the outcome variable, under the influence of different but related medium resolution satellite sensor data sources. By comparison of models with varying levels of richness from Landsat-8 and Sentinel-2, we were able to demonstrate that the Sentinel-2 spatial hierarchical model has the best predictive characteristics in terms of the $RMSE$, MAE, $CRPS$ and DIC, and is more attractive as compared to the Landsat-8 based C stock hierarchical model. For the sake of our goal, the Sentinel-2 derived C stock spatial prediction model is the best as it is applicable to domains beyond the training sample. The Sentinel-2 spatial prediction model is more useful than the Landsat-8 based C stock based predictive model in assisting plantation forest practitioners. This is critical as forest practitioners should correctly advise national governments on the possibility of resuscitated plantation forests in sequestering carbon in the country. The findings of this study will aid in the understanding of dynamics between vegetation indices and C stock concentration and determining area specific sites for intervention for ecological restoration and monitoring. The high cost of obtaining data for ecological monitoring and validation of biomass models by forest practitioners and managers can be substantially reduced as readily available and high-resolution Sentinel-2 remote sensing data is cheap to acquire.

There is however, a need to extend the present modelling framework to a broader coverage of auxiliary information by integrating broadband vegetation indices and bioclimatic variables in

a comprehensive Bayesian hierarchical framework. The high residual spatial dependence demonstrated in Sentinel-2 based C stock hierarchical modelling implies the effectiveness of covariate information from remote sensing. Over and above the significant reduction in C stock prediction uncertainty when modelling with covariates from finer spectral and spatial resolution, our model offers the possibility of robust modelling if we adopt a framework encompassing a much broader suite of independent variables to offer better predictive maps of C stock distribution in disturbed plantation ecosystems. We conclude that Sentinel-2 data can be proposed as the primary earth observation data source in the management and monitoring of managed ecosystems with history of ecological fragility.

2.12. Summary

This chapter presented an analytical discourse of how Landsat-8 and Sentinel-2 driven Bayesian hierarchical geostatistical models influence C stock prediction uncertainty. The Sentinel-2 led C stock Bayesian predictive model, with $NDVI$ and $DIST$ as significant predictors, gave lower C stock prediction uncertainty than the Landsat-8 predictive model counterpart. This result is novel and important in two important aspects. Firstly, it opens new avenues on the usefulness of the Sentinel-2 satellite sensor as a cheap and readily available data source for C stock monitoring. Secondly, the result sets itself for evaluation within literature studies employing similar variables and methodologies. The Bayesian approach only considered vegetation indices and the distance to settlements as sources of spatial variation for C stock. It is also possible that the modelled C stock variable and model parameters from the vegetation indices are non-stationary, thereby exerting more constraints on the outcome variable. Accurate accounting and reporting of C stock for climate change and mitigation studies for particular forest types and areas requires an exhaustive modelling approach that takes care of all possible sources of uncertainty into the modelling framework. Hence, the next chapter is an extension of the present chapter as it assesses how possible spatial heterogeneity in C stock in the form of spatially varying coefficients (SVC) influences the prediction uncertainty of Landsat-8 and Sentinel-2 driven C stock prediction models.

CHAPTER THREE

remote sensing

MDPI

Article

Landsat-8 and Sentinel-2 Based Prediction of Forest Plantation C Stock Using Spatially Varying Coefficient Bayesian Hierarchical Models

Tsikai Solomon Chinembiri [1], Onisimo Mutanga [1,*] and Timothy Dube [2]

[1] College of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, South Africa
[2] Institute of Water Studies, Department of Earth Sciences, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa
* Correspondence: mutangao@ukzn.ac.za

Abstract

This study sought to establish the performance of Spatially Varying Coefficient (SVC) Bayesian Hierarchical models using Sentinel-2 and Landsat-8 derived auxiliary information in predicting plantation forest carbon (C) stock in Manicaland province of Zimbabwe. The development and implementation of Zimbabwe's land reform program undertaken in the year 2000 and the subsequent redistribution and resizing of large-scale land holdings are hypothesized to have created heterogeneity in aboveground forest biomass in plantation ecosystems. The Bayesian hierarchical framework, accommodating residual spatial dependence and non-stationarity of model predictors, was evaluated. Firstly, SVC models utilizing Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), and Enhanced Vegetation Index (EVI), developed from Sentinel-2 and Landsat-8 data and 191 sampled C stock observations were constructed. The SVC models built for each of the two multispectral remote sensing data sets were assessed based on the goodness of fit criterion as well as the predictive efficacy using a 10-fold cross-validation technique. The introduction of spatial random effects in the form of Sentinel-2 and Landsat-8 derived covariates to the model intercept improved the model fit and prediction performance where residual spatial dependence was dominant. For the Landsat-8 C stock predictive model, the RMSPE for the non-spatial, Spatially Varying Intercept (SVI) and SVC models were 8 $MgCha^{-1}$, 7.77 $MgCha^{-1}$, and 6.42 $MgCha^{-1}$ whilst it was 7.85 $MgCha^{-1}$, 7.69 $MgCha^{-1}$ and 6.23 $MgCha^{-1}$ for the Sentinel-2 C stock predictive models, respectively. Overall, the Sentinel-2-based SVC model was preferred for predicting C stock in plantation forest ecosystems as its model provided marginally tighter credible intervals, [1.17–1.60] $MgCha^{-1}$ when compared to the Landsat-8 based SVC model with 95% credible intervals of [1.13–1.62] Mg Cha^{-1}. The built SVC models provided an understanding of the performance of the multispectral remote sensing derived predictors for modeling C stock and thus provided an essential foundation required for updating the current carbon forest plantation databases.

Keywords: Bayesian hierarchical modelling; geostatistics; *Eucalyptus grandis*; *Eucalyptus camaldulensis*; *Pinus patula*; spatial random effects; Spatially Varying Coefficient

3.1. Introduction

Since the onset of the Fast Track Land Reform (FTLRP) program in Zimbabwe and the subsequent redistribution and resizing of large-scale land holdings in the year 2000, plantation forests within and around the neighbourhood of resettlement areas continuously faced physical distress (Matose, 2008). Zimbabwe lost close to 224 000 ha of tree cover, which is equivalent to a forest degradation of 17% and 88 Mt of CO_2 emissions from 2000 to 2021 (Newsday, 2017). The steady decline of land under forest over the years has therefore been mainly driven by the activities of the year 2000 taking place on resettled land, which included wildfires, illegal logging, mining, and agriculture expansion. Yet the amount of additional biomass that can be accumulated in these areas depends much on the forest condition and land management practices.

A number of studies modelling the relationship between C stock and remote sensing-derived information adopt a standard approach where the effects of independent variables, including vegetation indices on C stock, remain constant over space in the model (Brown, 1997; Bordoloi *et al.*, 2022). Given the history of Zimbabwe's land reform program, it is well known that management and conservation practices of plantation forests have been changing over the years, particularly after the adoption of the FTLRP program in the year 2000. Greater portions of formally designated commercial plantation forests were occupied and replaced by subsistence farmers across the country. It is upon this premise that it cannot be assumed that the impact of covariates on C stock is constant in these ecosystems. The carbon sequestration potential in various regions of the occupied forest plantations, therefore, remains unknown as the management and conservation practices of the managed ecosystems have been largely modified and altered (Zvobgo and Tsoka, 2021). Yet government and other stakeholders in the timber industry are obliged under the 2015 Paris climate agreement to provide accurate accounts of the country's carbon sequestration potential for managed and natural forest ecosystems.

Literature proffers methods used in the estimation of AGB, either as direct or indirect approaches from forest inventory data. These methods either use allometric equations, conversion factors such as wood density, or biomass expansion factors (BEF) (Matose, 2008). In spite of the advantages of using conventional methods given in (Fuller *et al.*, 2010; Liang and Wang, 2020) as generally providing accurate AGB estimates, these techniques are also regarded as time-consuming and environmentally unfriendly. The inaccessibility of some areas due to complicated topography and forest conditions also makes conventional methods less attractive to AGB estimation, especially in extensive areas (Mutanga, Dube and Ahmed, 2016).

However, remote sensing is emerging as a promising technique free of the aforementioned limitations as it offers cost-effective methods of AGB monitoring through stratification of canopy density and forest types. Repeated application of remote sensing leads to the generation of historical data critical for change detection analysis and incorporation into a Geographic Information System (GIS) for integration with other datasets (Hoeting, 2009). However, remote sensing methods are also not immune from limitations as they usually face limitations in areas of bad weather conditions, especially in areas with cloud cover, and require a certain level of training for effective use and application (Fuller *et al.*, 2010).

Remote sensing C stock data derived from satellite imagery has significantly grown over the years. This data has been the basis for informing international policy agreements associated with CO_2 emissions into the atmosphere, mainly from deforestation and other land use land cover changes (LULCC) (Zunguze, 2012). Remote sensing is regarded as a powerful tool for deriving AGB and forest structure as it offers pragmatic methods of acquiring spatially-distributed forest carbon from the local, the continental, and the global scales (Gelfand *et al.*, 2003; Zvobgo and Tsoka, 2021). Forest biomass can generally be measured from three broad categories of remote sensing (RS) data, namely, passive optical remote sensing, radio detection and ranging, microwave (radar), and light detection and ranging (LiDAR) (Zunguze, 2012; Mutanga, Dube and Ahmed, 2016). Passive optical spectral reflectance is responsive to vegetation structural attributes (tree density, leaf area index, and crown size), shadow, and texture (Mutanga, Dube and Ahmed, 2016; Chrysafis *et al.*, 2017). Crown size, tree density, and leaf area index (LAI) are strongly correlated with AGB. Radar remote sensing measure geometrical and dielectric attributes of forests. LiDAR RS methods of biomass measurement characterise forest vertical structure and height. Remote sensing (RS) and ancillary technologies such as Geographical Information System (GIS) are practical and cost-time effective, allowing for imaging and research on extensive and inaccessible areas.

The estimation of plantation forest C stock, together with other structural parameters using new regeneration multispectral remote sensing data, is a relatively new ground in the climate change and carbon monitoring arena. Research in this field has demonstrated the potential for remote sensing data as tools for developing estimates of forest attributes, amongst them, being Above Ground Biomass (AGB), either as a standalone or coupled with other earth observation techniques (Gelfand *et al.*, 2003; Green, Finley and Strawderman, 2020). Spatial regression models applied in mapping C or biomass using new-generation multispectral remote sensing data may fail to clearly make room for residual spatial dependence (Popescu *et al.*, 2011; Tonolli *et al.*, 2011b). Modelling of natural resource variables without explicitly

accommodating spatial variation can be justified if covariates can account for all the spatially structured dependence. Such assumptions do not hold in many practical applications involving georeferenced data.

Failure to account for spatial dependence in the modelling of regionalized variables can lead to inaccurate model parameters and incorrect predictions (Hoeting, 2009). Subsequently, disregarding spatial correlation in electing model choices can result in higher prediction uncertainty for inference of the outcome variable. In addition to the drawbacks highlighted in the foregoing, non-Bayesian spatial modelling can further lead to underestimation of uncertainty (Hoeting, 2009; Babcock *et al.,* 2015) as traditional spatial regression estimation methods assume stationarity of the covariance matrix, $\boldsymbol{\Sigma}$. The ultimate effect is the derivation of standard error estimates that are unable to take all the uncertainties in the parameters into account. Checking for spatially correlated residuals when spatial data are employed in the modeling of above-ground biomass is therefore critical.

Some attention has been given to spatially varying coefficient (SVC) models in the literature (A. Gelfand *et al.*, 2004; Finley, Banerjee and McRoberts, 2009). The Bayesian framework of statistical inference is the bedrock of SVC models in which analyses make use of samples derived from Markov Chain Monte Carlo (MCMC) techniques from the posterior distribution of model parameters (Schabenberger and Gotway, 2004). What makes Spatially Varying Intercept (SVI) and SVC models unique in applied geostatistical and the remote sensing literature is their ability to consider the residual spatial dependence and the non-homogeneity in model parameters differently than ordinary geostatistical approaches. SVI models assume the model intercept is spatially varying, whilst the SVC models assume all the model regression coefficients to be spatially varying (Fotheringham, Brunsdon and Charlton, 2003). A number of applications and methodologies utilizing spatially varying coefficient models are documented in the literature. Amongst them are the geographically-weighted regression (GWR) by (Fotheringham, Brunsdon and Charlton, 2003), who employed the technique of canopy height prediction using remote sensing data. Application of these models resulted in significant improvement in model fit. On the other hand, (Hudak *et al.*, 2002) made provisions for spatial dependence and parameter non-homogeneity by exploring geostatistical kriging variants for forest canopy height prediction. Co-kriging and regression kriging models resulted in significant improvements in model fit.

However, it has been shown in recent times that GWR might not be robust to correlation among predictors and can potentially lead to inaccurate results when complex correlation structures

are involved (Wheeler and Waller, 2009; Finley, Banerjee and MacFarlane, 2011). Again, from an inferential viewpoint, GWR can present problems when drawing inferences regarding prediction uncertainty and model parameters. The lack of valid underlying probability models in GWR makes prediction uncertainty, and standard parameter error estimates difficult to justify. For instance, uncertainty maps produced from kriging and co-kriging techniques make no consideration of the uncertainty in the variogram-based spatial covariance parameters. This is an established and common weakness encountered when using these geostatistical approaches (Cressie, 1993). It is possible to estimate spatial autocorrelation parameters within the SVC and SVI models using a Bayesian hierarchical construction. Such an approach allows the propagation of uncertainty to the prediction of the response variable (Banerjee, Carlin and Gelfand, 2003).

Landsat-8 and Sentinel-2 are multispectral platforms categorized as new generation remote sensing sensors with enhanced spectral and spatial properties than previous missions of the Landsat series. It is this perceived improvement in earth observation properties that we expect to give a dividend to C stock predictive models for applications in carbon accounting and inventorying under the United Nations Framework Convention on Climate Change (UNFCCC) (Mutanga, Dube and Ahmed, 2016). We employed a Bayesian hierarchical framework with spatially varying coefficients (SVC) using predictor information derived from the aforementioned sensors for predicting C stock. The modelling framework considered the non-stationarity and residual spatial dependence of model predictors through the inclusion of spatial random effects into the SVC models. We, therefore, developed and compared the performance of C stock predictive models under spatially varying regression coefficients derived from Sentinel-2 and Landsat-8 predictors. Prediction accuracy and uncertainty quantification for the REDD collaborative program in the developing world is a critical aspect of the Carbon Measurement, Reporting, and Verification Systems (MRVs) of the United Nations (UN-REDD, 2009; CMS, 2014). Thus, in this paper, we explored how spatially varying coefficient models constructed using a Bayesian hierarchical set-up with aiding information from Sentinel-2 and Landsat-8 multispectral satellite data and implemented through Markov Chain Monte Carlo (MCMC), perform in C stock prediction in plantation forest ecosystems.

3.2. Materials and Methods

This section provides a detailed description of the study area and the methods used in the study.

3.2.1. Study Area

The study was carried out at lot 75 A of Nyanga Downs in Nyanga district in the Eastern Highlands of Zimbabwe (Figure 1). The study area is dominated by *Eucalyptus grandis*, *Eucalyptus camaldulensis,* and *Pinus patula* plantation forest species which have some of its patches being used for agriculture, grazing, and gold panning and is located between latitude 32°40′ E and 32°54′ E and 18°10′12″S and 18°25′4″S longitude as illustrated in **Figure 3-1** (Lisboa *et al.*, 2018; Zvobgo and Tsoka, 2021). Grazing, agriculture, and gold panning activities came after part of the commercially owned plantation forests were redistributed to small and medium-sized indigenous farmers in 2000.

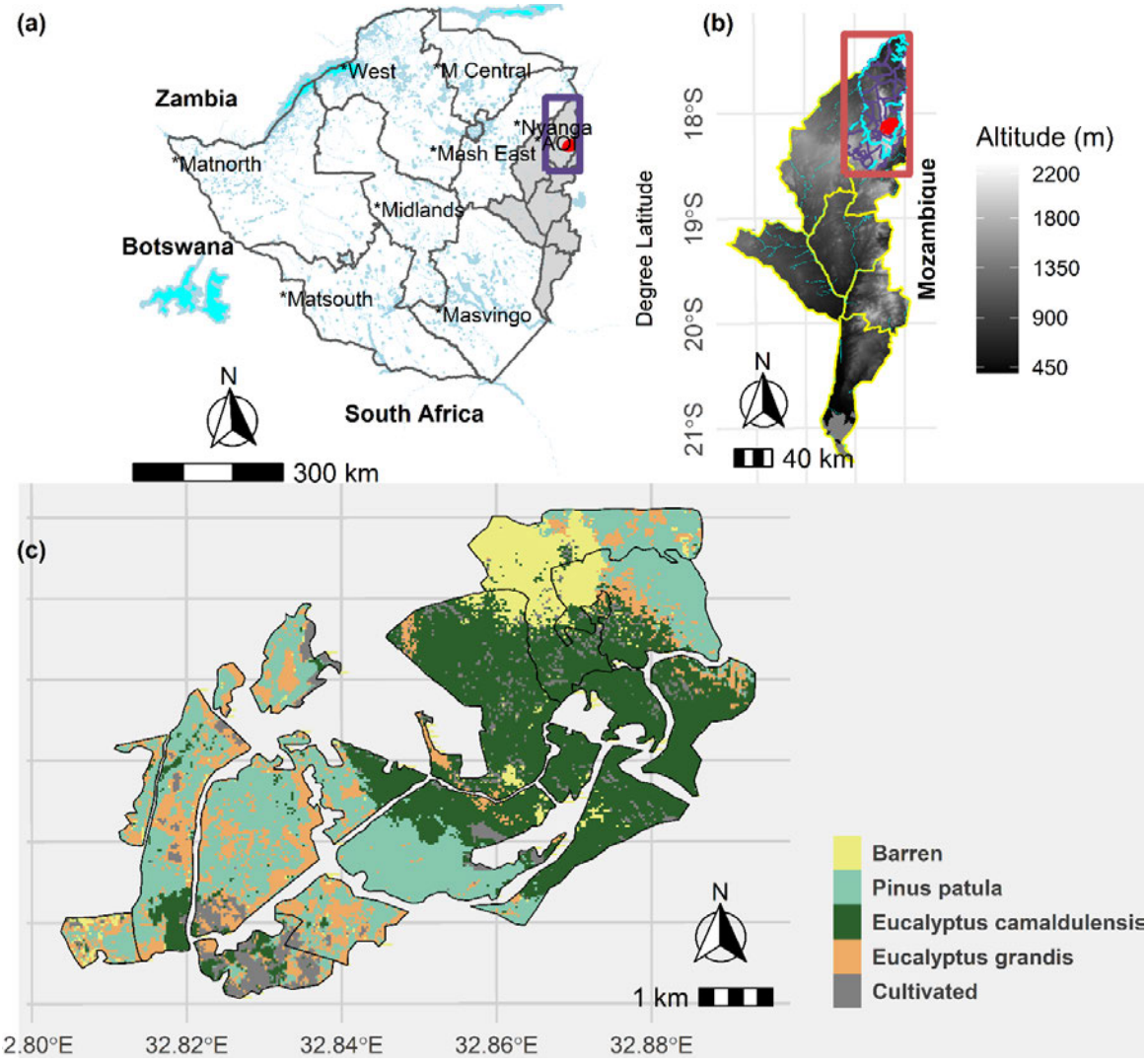


Figure 3-1: Map of the study area indicating (a) the province where samples were derived, (b) the study area location within the particular province, and (c) the spatial distribution of major plantation tree species in the studied region. * Refers to Provinces in

This development has increased the interface between settlements and timber plantations in all forests originally designated under forest plantations in Zimbabwe. The study area covers an area of approximately 2766 ha. Rainfall amounts are varied, with a mean annual precipitation range of 741 mm to 2997 mm. Annual mean temperatures range from a minimum of 9 °C to 12 °C to a maximum of 25 °C to 28 °C. The weather is very hot, and extensive wildfires occur in the high-elevation grasslands from August to November when the grasses are dry (Lisboa *et al.*, 2018).

As illustrated in **Figure 3-1 (c)**, *Eucalyptus camaldulensis* and *Pinus patula* are the most dominant species in the study area. Pockets of cultivated land within the plantations are evident and are partly responsible for the present biomass density in the sampled region. Greater portions of former plantation vegetation have been cleared by pockets of resettled small-scale and medium-scale farmers venturing into coffee and tea plantations and, in some cases, for dairy farming. Patches of unprotected forest plantations are still present but remain vulnerable to attack for agriculture by settled farmers in the area.

3.2.2. Sampling Design

We employed the spatial coverage sampling scheme that exploits the Mean Squared Shortest Distance (MSSD) for the optimization of data locations in the study domain.

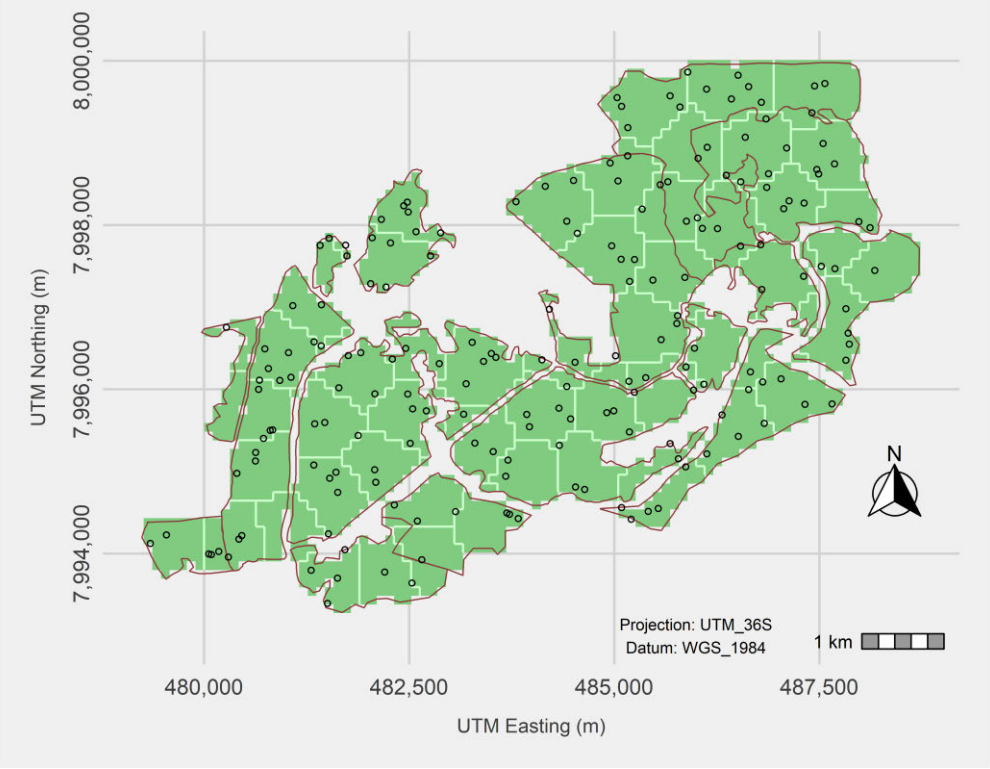


Figure 3-2: Study area sampling design. Points represent locations of sampled C stock

The *k-means* clustering scheme for equal area coverage was therefore used (Walvoort, Brus and de Gruijter, 2010) for obtaining a representative sample from the studied region. The work of (Brus, de Gruijter and van Groenigen (2006) demonstrated how the mapping and estimation of mean spatial problems could be resolved through the employment of a uniform coverage sampling scheme. MSSD is particularly suited for areas where sampling campaigns cannot be extended beyond a single phase. As illustrated in **Figure 3-2**, the smallest separation distance between samples was approximately 8 m whilst the largest separation distance between samples was 2500 m.

3.2.3. Above-Ground Biomass Measurements and C Stock Derivation

We sampled and collected measurements of Diameter at Breast Height (DBH) for Above Ground Biomass (AGB) estimation for all trees with a DBH of more than 10 cm (1.3 m) using 500 m^2 circular plots from the 19 September 2021 to 24 October 2021 in Manicaland Province of Zimbabwe as illustrated in **Figure 3-1**. The sampling program meant to measure DBH for subsequent AGB and C stock estimation utilized the MSSD optimization function, resulting in 191 sampling plots being measured from the study area. The 191 sample plot measurements of DBH were then transformed into per plot C stock data using allometric equations of Brown (1997) for the Pinus species, whilst the allometric equations of were used for deriving C stock for the Eucalyptus species. The aforementioned allometric equations were also applied by Zunguze (2012).

Allometric equations shown in Equations (3-1) and (3-2) were used for the calculation of Above Ground Biomass (AGB) for the *Pinus patula and the Eucalyptus grandis* and *Eucalyptus camaldulensis* species, respectively (Matose, 2008). A default conversion factor of 0.47 used by the IPCC was applied to derive AGB to C stock.

$$tDw = e^{(-1.170+2.119*ln(dbh))} \quad (3\text{-}1)$$

$$tDw = 0.39 * (dbh)^{2.142} \quad (3\text{-}2)$$

3.2.4. Modelling Framework

It is a common geostatistical practice to assume at location $S \in D \subseteq \mathbb{R}^2$ where s is a vector of observed x, y coordinates within the domain D. A Gaussian response variable $y(s)$ is therefore modeled through the regression model as in Equation (3-3):

$$y(\boldsymbol{s}) = x(\boldsymbol{s})'\beta + \widetilde{\boldsymbol{x}}(\boldsymbol{s})'\boldsymbol{w}(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) \quad (3\text{-}3)$$

$x(\boldsymbol{s})'$ denotes a set of p covariates in the model. In this case, the linear mean structure accounting for wide-scale variation in the response is comprised of $px1$ vectors of $\boldsymbol{x}(\boldsymbol{s})$ which

include an intercept and spatially varying georeferenced predictors together with an associated column vector of model coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$. The $\tilde{\boldsymbol{x}}(\boldsymbol{s})$ in the model represents a $q\ x\ 1$ vector accommodating the intercept and those predictors from $x(s)$ whose impact on the response is assumed to vary spatially. The space-varying impact is obtained from the vector of spatial random effects $\boldsymbol{w}(\boldsymbol{s}) = \left(w_1(\boldsymbol{s}), w_2(\boldsymbol{s}), \ldots, w_q(\boldsymbol{s})\right)'$. The specification of different combinations of $\tilde{\boldsymbol{x}}(\boldsymbol{s})$ and the associated $\boldsymbol{w}(\boldsymbol{s})$ leads to the formation of different sub-models. We model $\varepsilon(\boldsymbol{s})$ as an independent white noise process that takes care of the micro-scale (measurement error) variation. As such, with the collection of n C stock locations, $S = \boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_n}$, we assume the $\varepsilon(\boldsymbol{s_i})'$s are independent and identically distributed (iid) as provided by $N(0, \tau^2)$ where τ^2 is the nugget.

The spatial structure of this model is generally introduced by way of a multivariate Gaussian process (GP) (Cressie, 1993; Gelfand *et al.*, 2003; Wackernagel, 2003) in which a cross-covariance function clearly models the covariance of $\boldsymbol{w}(\boldsymbol{s})$ within and among data points. The added flexibility in this model is provided in the literature (A. Gelfand *et al.*, 2004; Banerjee *et al.*, 2010). We assume in this study that elements of $\boldsymbol{w}(\boldsymbol{s})$ emanate from q independent univariate GPs. Precisely, the process associated with the $k-th$ prediction is $w_k(\boldsymbol{s}) \sim GP(0, (.,.; \boldsymbol{\theta}_k))$ where $C(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_k) = Cov(w_k(\boldsymbol{s}), w_k(\boldsymbol{s}^*))$ is a sound spatial covariance function modelling the covariance related to a pair of observations $\boldsymbol{s}$ and $\boldsymbol{s}^*$. The process outcomes are gathered into an $n\ x\ 1$ vector, say $\boldsymbol{w}_k = (w_k(\boldsymbol{s_1}), \ldots, w_k(\boldsymbol{s}_n))'$ which permits a multivariate normal distribution $MVN(0, \Sigma_k)$, in which Σ_k is an $n\ x\ n$ covariance matrix of $\boldsymbol{w}_k$ with the $(i,j)-th$ element provided by $C(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta}_k)$. Evidently, $C(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_k)$ cannot just be a function, but guarantees that the resulting $\Sigma_k$matrix is positive definite and symmetric. Functions of this type are regarded as positive definite and are referred to as the characteristic function of a symmetric stochastic variable (Cressie, 1993; Banerjee, Carlin and Gelfand, 2003; Chiles and Delfiner, 2009).

We denote $C(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_k) = \sigma_k^2 \rho(\boldsymbol{s}, \boldsymbol{s}^*; \theta_k)$ with $\boldsymbol{\theta}_k = \{\sigma_k^2, \phi_k\}, \rho(.; \phi_k)$ to be a valid spatial correlation function in which ϕ_k computes the correlation decay rate and $var(w_k) = \sigma_k^2$. We assumed for all the accompanying analyses an exponential correlation function $\rho(||\boldsymbol{s} - \boldsymbol{s}^*||; \phi_k) = exp(-\phi_k||\boldsymbol{s} - \boldsymbol{s}^*||)$, where $||\boldsymbol{s} - \boldsymbol{s}^*||$ represents the Euclidean distance between location $\boldsymbol{s}$ and location $\boldsymbol{s}^*$. Completion of the Bayesian modeling framework requires specification and assignment of prior probability distributions to the parameters of the model, where inference progresses by sampling from the posterior probability distribution of the modelled parameters. As standard practice, we assume $\beta \sim MVN(\boldsymbol{\mu}_\beta, \Sigma_\beta)$ prior where $\boldsymbol{\mu_\beta} = 0$

and $\Sigma_\beta = 10000\boldsymbol{I}_p$, whilst the spatial variance aspects, σ_k^2's and the measurement error variance, τ^2, are assigned inverse-Gamma, $IG(a,b)$ priors. The spatial decay parameters $\phi_k \sim Unif(a,b)$ with the upper and lower bounds of the distribution covering the geographic domain of the sampled study region.

Applying notation similar to the ones used by Wackernagel (2003), we can specify the model parameter posterior distribution as $p(\Omega|\boldsymbol{y})$ where:

$$\Omega = \{\boldsymbol{\beta}, \boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_q}, \sigma_1^2, \sigma_2^2, \ldots, \sigma_q^2, \phi_1, \phi_2, \ldots, \phi_q, \tau^2\}$$

$\boldsymbol{y} = (y(s_1), \ldots y(s_n))'$ and is proportional to:

$$\prod_{k=1}^{q} Unif\left(\phi_k | a_{\phi k}, b_{\phi k}\right) x \prod_{k=1}^{q} IG(\sigma_k^2 | a_{\sigma k}, b_{\sigma k})\ x\ N\left(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta\right) x IG(\tau^2 | a_\tau, b_\tau)$$
$$\prod_{k=1}^{q} N(\boldsymbol{w}_k | \boldsymbol{0}, \Sigma_k) x \prod_{i=1}^{n} N(y(\boldsymbol{s}_i) | \boldsymbol{x}(\boldsymbol{s}_i)'\boldsymbol{\beta} + \tilde{\boldsymbol{x}}(\boldsymbol{s})'\boldsymbol{w}(\boldsymbol{s}), \tau^2) \quad (3\text{-}4)$$

An effective Markov Chain Monte Carlo (MCMC) function for the estimation of Equation (3-4) is derived through updating of $\boldsymbol{\beta}$ from its full conditional and utilizing Metropolis procedures for the remainder of the parameters. Reparameterization of the model is an alternative way of ensuring the unmeasured covariates, $\boldsymbol{w}$, do not require direct sampling (Finley and Banerjee, 2008). The spatial random effects could represent other independent variables that are spatially structured and not taken into consideration in the current modelling approach. Nonetheless, the MCMC process produces posterior samples of the parameter space, $\boldsymbol{\Omega}$.

From a prediction point of view, if $S_0 = \{\boldsymbol{s}_{0,1}, \boldsymbol{s}_{0,2}, \ldots, \boldsymbol{s}_{0,m}\}$ is a set of m new sites, the spatial random effects posterior predictive distribution corresponding to the $k-th$ regression coefficient is provided by:

$$p(\boldsymbol{w}_{k,0}|\boldsymbol{y}) \propto \int p(\boldsymbol{w}_{k,0}|\boldsymbol{w}_k, \Omega, \boldsymbol{y}) p(\boldsymbol{w}_k|\Omega, \boldsymbol{y}) p(\Omega|\boldsymbol{y}) d\Omega \boldsymbol{w}_k \quad (3\text{-}5)$$

where $\boldsymbol{w}_{k,0} = (w_k(\boldsymbol{s}_{0,1}), w_k(\boldsymbol{s}_{0,2}), \ldots, w_k(\boldsymbol{s}_{0,m}))'$.

Since we are making use of MCMC sample-based inference, the integral in Equation (3-5) does not need to be determined precisely. Instead, given *L* posterior samples for the parameter space (Ω), that is, $\{\Omega^{(l)}\}_{l=1}^{L}$, composition sampling can be used to derive this distribution by first drawing w_k^l followed by $\boldsymbol{w}_{k,0}^{(l)}$ for each l from $p(\boldsymbol{w}_{k,0}|\boldsymbol{w}_k^{(l)}, \Omega^{(l)}, \boldsymbol{y})$. The last distribution is multivariate normal as it is a derivative of a conditional distribution from a multivariate normal distribution. More specifically, the process realizations over the new sites are conditionally independent of the measured outcomes given the values over the sampled locations and the process parameters. Expressed differently, $p(\boldsymbol{w}_{k,0}|\boldsymbol{w}_k, \Omega, \boldsymbol{y}) = p(\boldsymbol{w}_{k,0}|\boldsymbol{w_k}, \Omega)$ is a multivariate distribution with mean and variance given by

$$E[\boldsymbol{w}_{k,0}|\boldsymbol{w}_k, \Omega] = Cov(\boldsymbol{w}_{k,0}, \boldsymbol{w}_k) var^{-1}(\boldsymbol{w}_k)\boldsymbol{w}_k = R_0(\phi_k)' R(\phi_k)^{-1}\boldsymbol{w}_k \quad (3\text{-}6)$$

and

$$var[\boldsymbol{w}_{k,0}|\boldsymbol{w}_k, \Omega] = \sigma_k^2\{R(\phi_k) - R_0(\phi_k)'R(\phi_k)^{-1}R_0(\phi_k)\} \quad (3\text{-}7)$$

where;

$R_0(\phi_k$ is an $n\ x\ n$ matrix with $(i,j) - th$ element specified as $\rho(\boldsymbol{s}_{0,i}, \boldsymbol{s_j}; \theta_k)$ and $R(\phi_k$ is an $n\ x\ n$ matrix with $(i,j) - th$ element provided by $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j; \phi_k)$. Repetition of the same procedure results in the generation of samples for all the $\boldsymbol{w}_{k,0}$'s. Lastly, for a set of predictors at unsampled locations $\boldsymbol{s}_0$, posterior predictive distribution samples regarding the outcome variable $y(\boldsymbol{s}_0)^l$, are derived from $N(\boldsymbol{x}(\boldsymbol{s})_0{}'\boldsymbol{\beta}^{(l)} + \widetilde{\boldsymbol{x}}(\boldsymbol{s}_0)'\boldsymbol{w}_0{}^{(l)}, \tau^{2(l)})$ for $l = 1,2, \dots, L$.

We evaluated 95% credible interval widths ($CIWs$) using the posterior predictive distribution of Landsat-8 and Sentinel-2 C stock models by calculating the difference between the 2.5% and the 97.5% quantile bounds. The 95% $CIWs$ were therefore used as summaries of the C stock prediction uncertainty for the Landsat-8 and Sentinel-2 C stock-based spatially varying coefficient models.

3.2.5. Competing Models

We derived five candidate models for each of the two multispectral remote sensing-based C stock models using Equation (1) using $NDVI$ and $SAVI$ as predictors of C stock (Mansfield and Helms, 1982). The models included the non-spatial where w_k's is fixed to zero; the spatially varying intercept (SVI) in which we only included the spatial random effects related to the model intercept; the complete SVC model in which all predictors have associated spatial random effects; the $SVC - \beta_1$ which included the spatial random effects for the intercept and $NDVI$ predictor variables and the $SVC - \beta_2$ which included the spatial random effects for the intercept and the $DIST$ independent variables.

We utilized empirical semi-variograms modelled for the residuals of the independent error model as guidelines for candidate model IG and $Unif$ hyperprior specifications. Precisely, for the variance parameters, the Inverse Gamma hyperprior a was set equal to 1.76, which would result in a mean prior distribution equal to b and infinite variance (Gelman *et al.*, 1995). To add on, the models' b hyperpriors for the τ^2 and σ^2's were calibrated in accordance with the sample variograms of the simple linear regression model residuals derived from the Sentinel-2 and the Landsat-8 sensor's nugget and partial sill. We programmed the prior for the spatial decay parameter ϕ's to $Unif\ (0.38, 0.0012)$ which, adopting the exponential covariance function, equates to support an effective range spanning between ~ 8 and $2\ 500\ m$. We define

the effective range as the distance where the correlation equates 0.05 (Finley, Banerjee and MacFarlane, 2011).

Three Markov Chain Monte Carlo (MCMC) chains were run for 30,000 iterations, each with the computationally demanding model requiring approximately 30 min to complete a single MCMC chain. We diagnosed convergence using the CODA library in the R Statistical and Computing environment by regulating the mixing of chains and the Gelman–Rubin statistic (Gelman and Rubin, 1992). Acceptable convergence was established within 11,000 iterations for all the models. The posterior inference was premised on a post-burn-in subsample of 15,000 iterations, that is, every third sample from the last 15,000 iterations of each chain. SVC and SVI models were fit using the *spBayes* R Statistical and Computing Library version 0.4.3. We, therefore, utilized the *spBayes R statistical package* for fitting all the predictive models for both the Sentinel-2 and the Landsat-8 SVC models.

3.2.6. Sentinel-2 MSI and Landsat OLI Imagery Derived Covariates

Landsat-8 has a revisit time of 16 days and offers nine spectral bands with a spatial resolution of 30 m for Bands 1 to 7 and 9 (Korhonen *et al.*, 2017; Wang *et al.*, 2020). The panchromatic band, Band 8, has a spatial resolution of 15 m. On the other hand, Sentinel-2 has thirteen spectral bands where four bands are configured at 10 m spatial resolution, six bands at 20 m, and three bands at 60 m spatial resolution (Korhonen *et al.*, 2017). Common vegetation indices utilized in the estimation of biophysical variables of Absorbed photosynthetically active radiation (APAR), Leaf Area Index (LAI), and biomass are Normalised Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and Soil Adjusted Vegetation Index (SAVI) (Zvobgo and Tsoka, 2021). We utilized the same Vegetation Indices (VIs) in this study. Readily available and cost-effective multispectral sensors (Sentinel-2 MSI and Landsat-8) with improved spectral and spatial resolution were therefore utilized in the modelling of C stock under the spatially varying coefficients assumption.

We acquired the Landsat 8 imagery from the United States Geological Survey Earth Explorer (http://earthexplorer.usgs.gov) as data ready for analysis (ARDs) on the 20th of September 2020. All the datasets were riddled with cloud cover and shadow cloud cover limits set to smaller than 10%. We acquired Sentinel-2 cloud-free images on 20 September 2020 at the same time as the Landsat 8 OLI data collection covering the entire area coinciding with the dates Landsat-8 OLI was collected covering the domain of interest. The multispectral instrument is the main imaging instrument used for Sentinel-2, a push broom scanner that measures the terrestrial Top of the Atmosphere (TOA) reflectance in thirteen spectral bands, that is, 443 nm

to 2190 nm. We derived Sentinel-2 spectral data as level-1C 12-bit automated TOA reflectance values. Orthorectification and pre-processing of the TOA-derived products were performed using the sen2r package of the R Statistical and Computing environment (Ranghetti *et al.*, 2020).

Soil Adjusted Vegetation Index ($SAVI$), Enhanced Vegetation Index (EVI), and Normalized Difference Vegetation Index ($NDVI$were used as Sentinel-2 and Landsat-8 derived covariates for C stock prediction in the plantation forest of the eastern highlands of Zimbabwe. The literature is replete with studies that have utilized the aforementioned vegetation indices (Li and Li, 2019; Bordoloi *et al.*, 2022) as ABG predictors in biomass estimation. In this study, predictor variables supplied by both Landsat-8 and Sentinel-2 are used specifically for comparing SVC Bayesian hierarchical geostatistical models predicting C stock in a disturbed environment in Zimbabwe. Given information regarding C stock distribution at each location, we fit the model in Equation (1) SVC with $p = 2$. Hence, we have two models, an intercept and one slope process model relating to $NDVI$.

3.2.7. Model Fit and Prediction Accuracy Evaluation

We assessed the performance of the models using the commonly used Deviance Information Criterion (DIC) to categorize models in terms of how good they fit the data (Spiegelhalter *et al.*, 2002). The sum of the Bayesian Deviance and the effective number of model parameters make up the DIC criterion. The Bayesian deviance measurement, which assesses model goodness of fit, and the effective number of model parameters which penalizes model complexity, are measured by D and pD, respectively. Attractive models have lower DIC values.

Predictive performance was evaluated through a $k - fold$ cross-validation technique. C stock was predicted from observations within each of the subsets, given the estimated parameters from the remaining subsets. We employed the Root Mean Square Prediction Error (RMSPE) as a metric from the R Statistical and Computing environment to calculate the sampled C stock data values and the accompanying average of the posterior predictive distribution (PPD). Models with smaller RMSPE indicate more accurate C stock predictions.

3.3. Results

This section provides specific information on the results of the study.

3.3.1. Multispectral Remote Sensing C Stock Derived Predictors

Employed predictors in C stock prediction using a Bayesian hierarchical framework with spatially varying coefficients showed NDVI as a significant predictor of C stock as illustrated

in **Table 3-1**. 95% credible intervals of SAVI and EVI contained zero and hence, rendering them as insignificant predictors of C stock. These predictors were therefore excluded in the final prediction and mapping of C stock distribution.

Table 3-1: Sentinel-2 and Landsat-8 derived predictors of C stock.

Parameter	Lansat-8 OLI C Stock Model				Sentinel-2 MSI C Stock Model			
	Mean	**s.d**	**2.5%**	**97.5%**	**Mean**	**s.d**	**2.5%**	**97.5%**
Intercept	−2.65	1.02	−4.76	−0.83	−2.40	0.31	−3.01	−1.80
NDVI	2.47	0.98	0.78	4.62	4.90	0.23	4.52	5.38
SAVI	−0.57	0.67	−2.04	0.63	−0.55	0.34	−1.25	0.17
EVI	−0.62	0.55	−1.57	0.50	−0.002	0.096	−0.17	0.24
σ_w^2	1.25	0.26	0.72	1.72	0.075	0.016	0.046	0.092
σ_ε^2	0.35	0.17	0.074	0.54	0.0043	0.0030	0.0005	0.011
ϕ	0.0016	0.0003	0.0012	0.0018	0.0015	0.0002	0.0014	0.0014

3.3.2. Candidate Models and Parameter Estimates

Model posterior estimates of the regression coefficients for the Landasat-8 and Sentinel-2-based C stock models are illustrated in **Tables 3-2** and **3-3**, respectively, for the non-spatial, SVI, and SVC models. Spatial autocorrelation is modelled in the residuals in the SVI, SVC, and all the $SVC - \beta_k$ variant models.

Table 3-2: Landsat-8-based SVC model median parameter estimates alongside their 95% credible intervals.

		Non-Spatial	SVI	SVC – NDVI	SVC
Parameter C.I 50% (2.5%, 97.5%)	β_0	−2.5 (−4.1, −0.8)	−2.7 (−4.6, −0.8)	−2.9 (−5.0, −0.9)	−3.5 (−5.3, −1.7)
	$\tilde{\beta}_{NDVI}(s)$	5.4 (3.3, 7.5)	2.8 (0.8, 4.7)	3.3 (1.2, 5.1)	3.6 (1.4, 5.8)
	τ^2	1.5 (1.2, 2.0)	0.5 (0.3, 0.8)	0.12 (0.03, 0.40)	0.1 (0.02, 0.2)
	$3/\phi_0$	-	0.0014 (0.0012, 0.0021)	0.003 (0.002, 0.003)	0.1 (0.07, 0.16)
	$3/\phi_{NDVI}$	-	-	0.13 (0.06, 0.22)	0.03 (0.02, 0.05)
	σ_0^2	-	1.0 (0.5, 1.5)	1.1 (0.8, 1.6)	0.24 (0.12, 0.48)
	σ_{NDVI}^2	-	-	1.0 (0.2, 3.0)	0.45 (0.14, 1.43)
Fit Statistics	D	436	140	41.8	65.9
	pD	4.1	68.9	87.0	108.9
	DIC	207	105.6	−77.9	29.0
	$RMSPE$ (MgCha^{-1})	8	7.77	7.54	6.42

C.I * means 95% Credible Interval.

This may entail differences in the posterior regression coefficient parameter estimates, depending on the C stock model parameter structure. Credible intervals (95% CI) for β_0 and β_{NDVI} including zero for both non-spatial and SVI models would hint at a non-significant relationship between the C stock observations and predictor variables. However, we cannot apply the same reasoning and interpretation for the SVC models as the predictor-specific spatially varying coefficient maps of $\beta_{NDVI} + w_{NDVI}(s)$ should be considered and ascertain whether location-specific CIs include zero. Posterior estimates for the Landsat-8 and Sentinel-2 based $\beta_{NDVI} + w_{NDVI}(s)$ coefficients are shown in **Figures 3-3** and **3-5**, with a significant relationship between the outcome variable and $NDVI$ being evident over the entire study domain.

Table 3-3: Sentinel-2-based SVC model median parameter values alongside 95% credible intervals.

		Non-Spatial	SVI	SVC – NDVI	SVC
Parameter C.I 50% (2.5%, 97.5%)	β_0	−2.5 (−4.1, −0.8)	−2.8 (−4.6, −0.9)	−2.9 (−4.9, −0.7)	−3.5 (−5.4, −1.6)
	$\tilde{\beta}_{NDVI}(s)$	5.4 (3.3, 7.5)	3.0 (1.2, 4.9)	2.9 (0.7, 4.9)	3.7 (1.5, 5.7)
	τ^2	1.5 (1.2, 2.0)	0.5 (0.3, 0.9)	0.32 (0.15, 0.55)	0.2 (0.11, 0.42)
	$^3/_{\phi_0}$	-	0.0015 (0.0014, 0.0016)	0.0016 (0.0015, 0.0018)	0.06 (0.018, 0.076)
	$^3/_{\phi_{NDVI}}$	-	-	0.09 (0.036, 0.13)	0.0015 (0.0014, 0.0017)
	σ_0^2	-	0.7 (0.4, 1.1)	1.1 (0.8, 1.9)	0.23 (0.16, 0.47)
	σ_{NDVI}^2	-	-	0.7 (0.43, 1.1)	0.65 (0.41, 1.05)
Fit Statistics	D	436.6	159	93.3	17.9
	pD	4.1	58.0	89.8	106.1
	DIC	207.6	112.5	65.6	−177.3
	$RMSPE$ (MgCha^{-1})	7.85	7.69	7.46	6.23

C.I * means 95% Credible Interval.

In both cases, there is significant variability in the values of $\beta_{NDVI} + w_{NDVI}(s)$ in the studied region. This could imply more accessibility to plantation resources by communities settling in the plantation areas. The same trend is noticeable for the Sentinel-2-based C stock-based SVC model (**Figure 3-4** and **3-6**) for both predictor coefficient values of $\beta_{NDVI} + w_{NDVI}(s)$ and for the same region as observed in the Landsat-8 based SVC model in **Figure 3-3**. Communities settled within certain areas of the plantation forest have more access to forest resources than in

other areas, rendering low C stock density in these areas as demonstrated by the corresponding low $NDVI$ values in the same region for the Sentinel-2-driven SVC model. Enhanced detail in the spatial resolution of Sentinel-2 based $\beta_{NDVI} + w_{NDVI}(s)$ vindicates variability in $\beta_{NDVI} + w_{NDVI}(s)$ coefficient values in the same region over those derived from generalized Landsat-8 multispectral data (Gerald *et al.*, 2017).

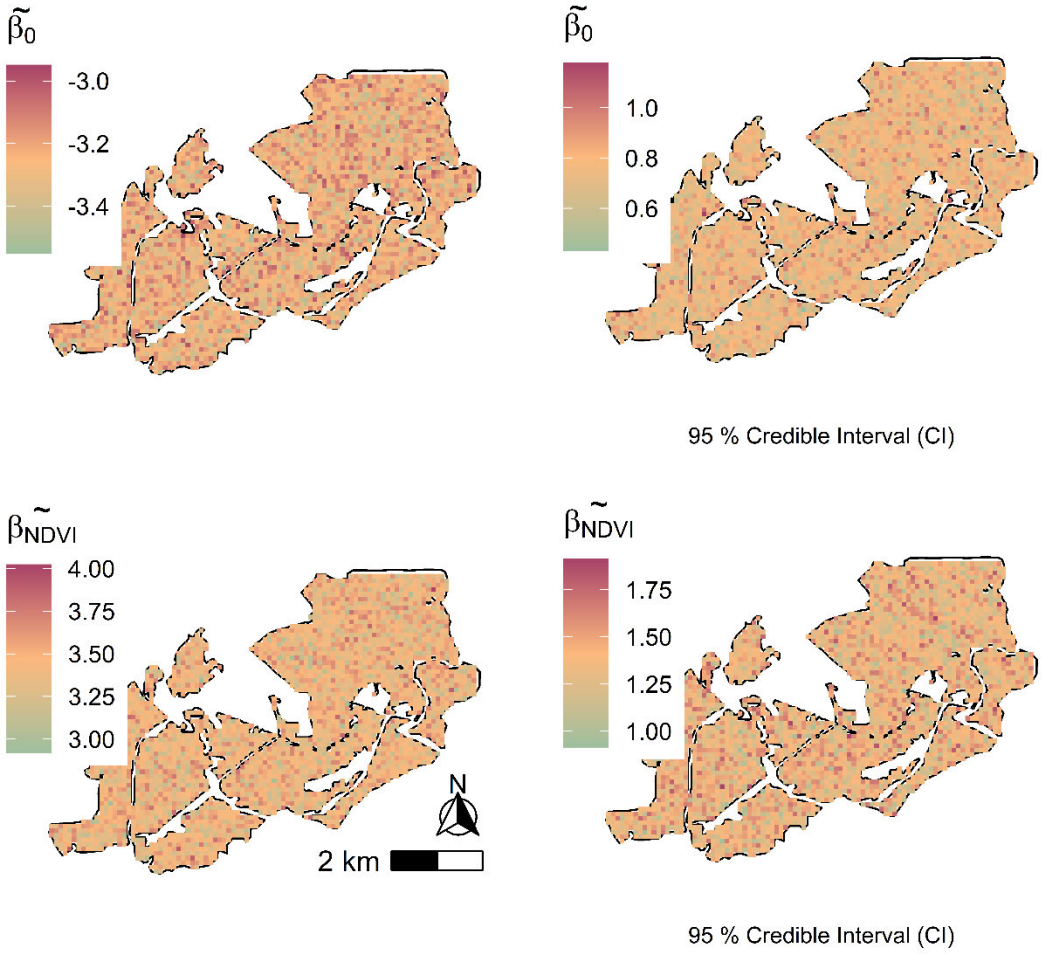


Figure 3-3: Landsat-8-based C stock spatially varying coefficient maps alongside their 95% credible intervals for the SVC model.

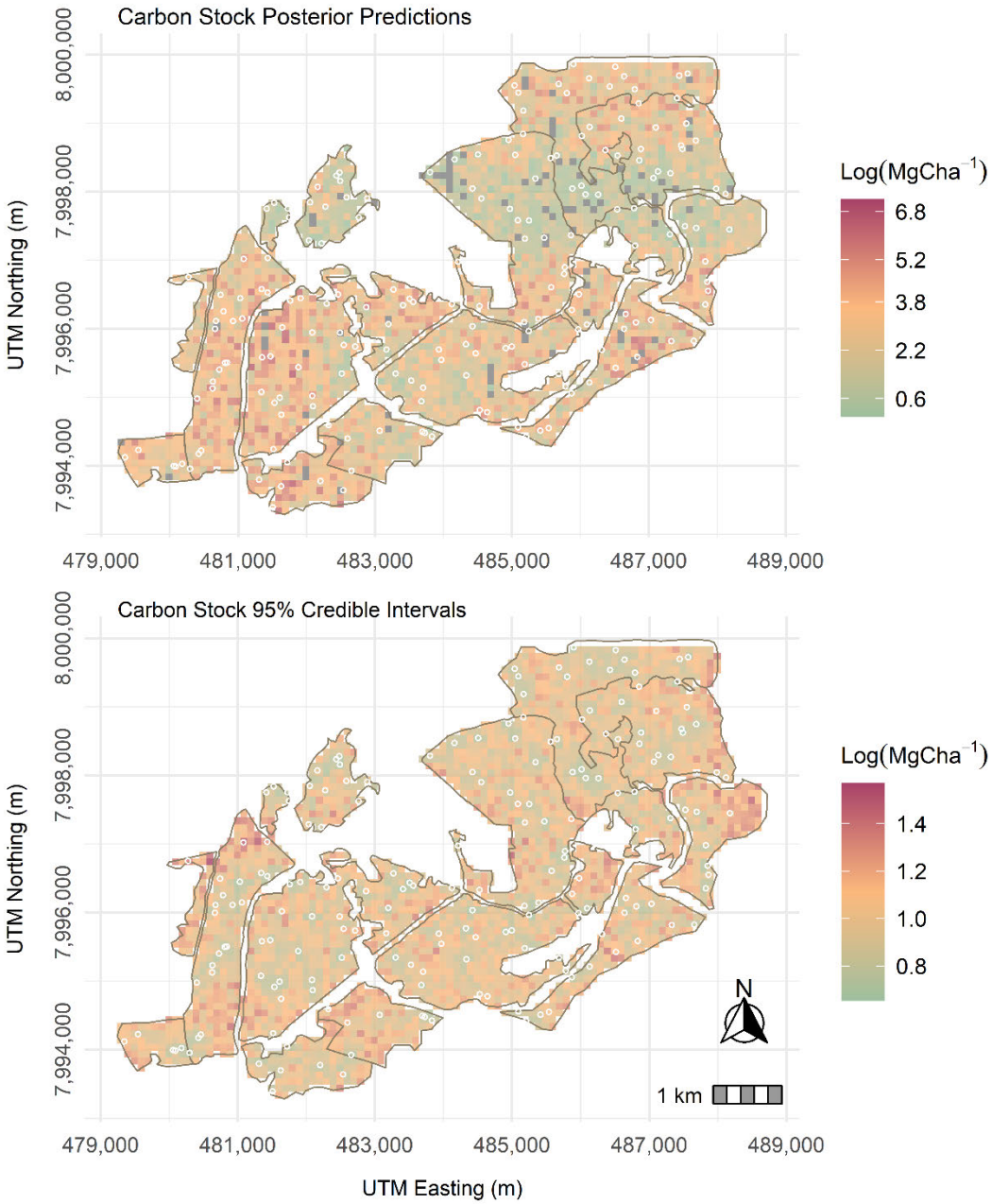


Figure 3-4: Landsat-8-based C stock posterior predictions. Points represent locations of sampled C stock.

Estimates of $\beta_0 + w_0(s)$ for the Landsat-8 and Sentinel-2-based SVI and SVC models are illustrated in **Figure 3-3** and **Figure 3-5**, respectively. The $\beta_0 + w_0(s)$ pattern for the SVI is the same for both Landsat-8 and Sentinel-2 spatially varying coefficient models whilst the $\beta_0 + w_0(s)$ SVC pattern in Landsat-8 is the same for Landsat-8-based SVC ($\beta_{NDVI} + w_{NDVI}(s)$) process. On the other hand, the $\beta_0 + w_0(s)$ SVI pattern for Sentinel-2 based model is different for Sentinel-2 $\beta_{NDVI} + w_{NDVI}(s)$ SVC model. However, the same is not true for Sentinel-2-based SVI as the partitioning of w_0 into w_0 and w_{NDVI} in Sentinel-2 SVC is detailed with enhanced spatial resolution compared to Landsat-8 based w_{NDVI} SVC model.

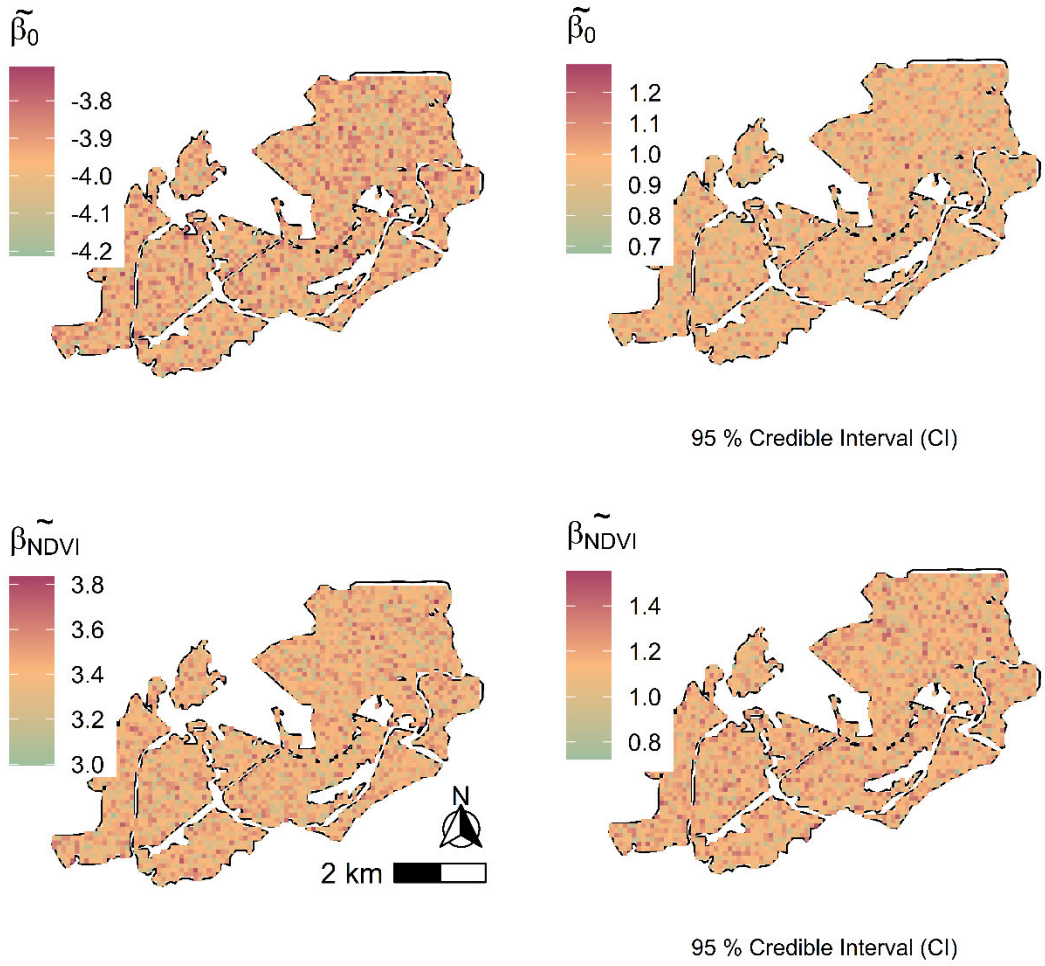


Figure 3-5: Sentinel-2-based C stock spatially varying coefficient maps alongside their 95% credible intervals for the SVC model.

Tables 3-2 and **3-3** illustrate the posterior estimates of the uncorrelated residual variance, τ^2. The uncorrelated residual variance is largest in the non-spatial model, whilst it is small for the SVC-variant models for both Sentinel-2 and Landsat-8 based SVC. The SVI and the SVC variant models incorporate a spatially varying correlated random effect, w_0 with variance σ_0^2. The SVC model variants further incorporate more spatially varying correlated random effects w_{NDVI} and w_{DIST}, with variance σ_{NDVI}^2.. For both Landsat-8 and Sentinel-2 based models, the w_0 and w_k explained much of the residual variability and hence, a reduced τ^2. The implication is higher predictive accuracy for the SVC-variant models in both sensors, making the SVI and SVC more attractive over the error-independent models. This is supported by the goodness of fit diagnostics for the SVI and the SVC-variant models illustrated in **Tables 3-2**, and **3-3** for the Landsat-8 and Sentinel-2 derived regression coefficients, respectively.

In comparison to the SVI model, the SVC model based on both remote sensing-derived covariates reduced the non-spatial residual spatial dependence by incorporating the space-varying impact of β_{NDVI}. Estimates of the spatial process parameters have a big difference. In particular, the spatial process parameters for the Landsat-8-based (**Figure 3-3**) SVC point estimates of $\sigma_0^2 = 0.24$ and the effective range of 30 m (i.e., $\approx -log(0.05/\phi) = -log(0.05/0.1)$ denote a less variable and significantly shorter effective spatial range than the spatial process of the SVI model. The same pattern recurs in the Senitnel-2-based SVI and SVC models, where the effective spatial range reduces from $\approx 1800\,m$ in the SVI to $\approx 100\,m$ in the SVC model (**Table 3-3**). The non-negligible spatial process parameter estimates of σ_{NDVI}^2 and ϕ_{NDVI} in the SVC model denote a potentially space-varying relationship between C stock and multispectral remote sensing derived covariates.

3.3.3. Landsat-8 and Sentinel-2 C Stock-Based Predictions

The entire SVC model based on Landsat-8 and Sentinel-2-based C stock models generated the lowest DIC, D, and RMSPE, as illustrated in **Tables 3-2** and **3-3**, respectively. Landsat-8-based SVC 95% CIW is much shorter when compared to the non-spatial fitted model values. A 10% improvement in the RMSPE is seen in the Landsat-8-based model when moving from the error-independent to the SVC model. In the same vein, the Sentinel-2-based model had a 12% improvement from the SVI to the SVC model (**Table 3-3**). Predictions produced for both Landsat-8 and Sentinel-2-based C stock models mirrored the observed C stock data in the studied region. Observed C stock data values ranged from log (0.2–6) $MgCha^{-1}$ (**Figure 3-1**). Variability in C stock uncertainty is fairly constant across the studied domain for both Landsat-8 and Sentinel-2 SVC-based models.

The almost similar variability in the density of C stock predicted by both sensors could be attributed to the inadequacy of covariates in the modelling framework, which when the range of modelled covariates is broadened, could accurately depict the variability of C stock in these disturbed plantation forest ecosystems (Zvobgo and Tsoka, 2021).

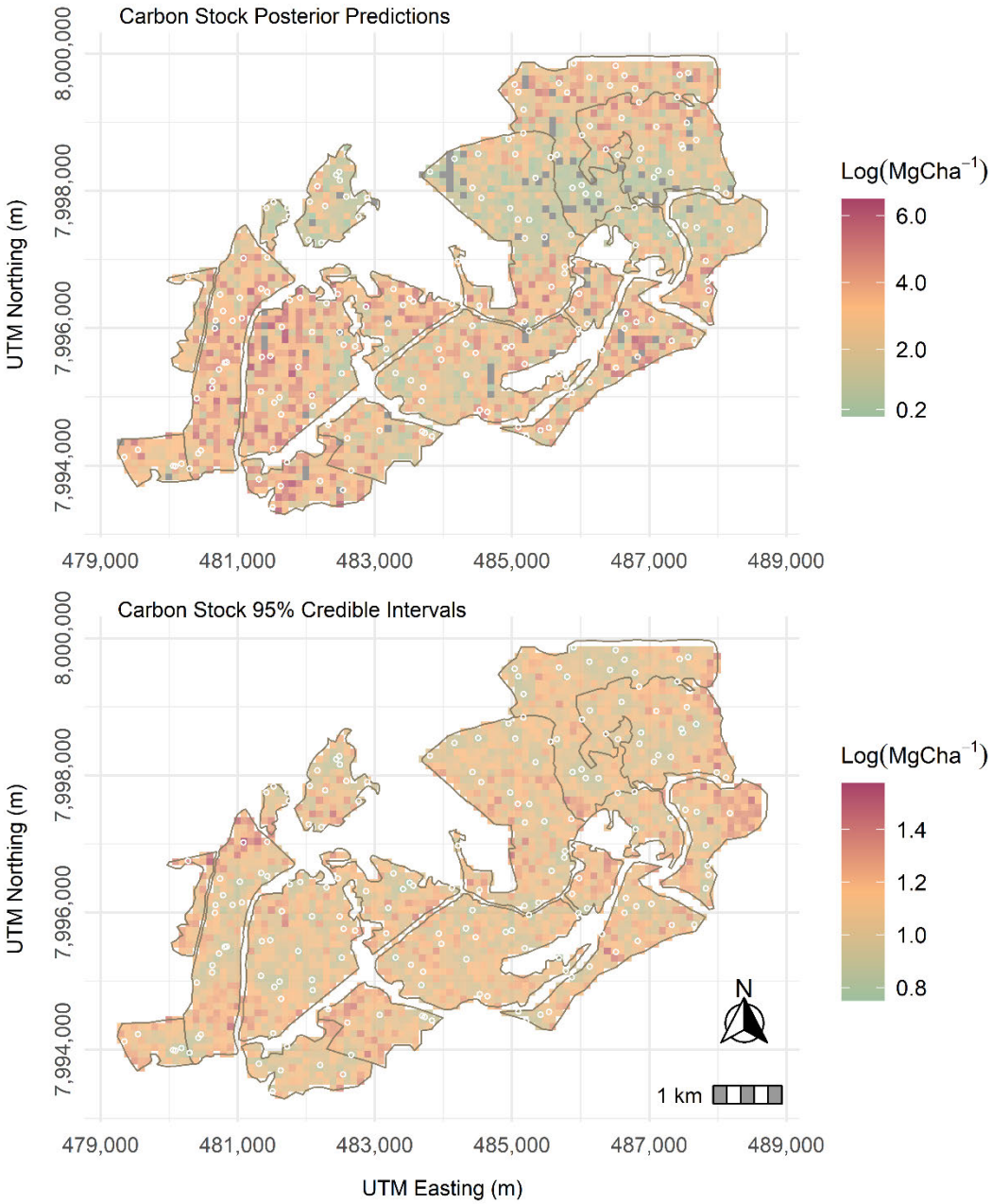


Figure 3-6: Sentinel-2-based C stock posterior predictions. Points represent locations of sampled C stock.

3.3.4. C Stock Model Prediction Assessment

Scatterplots observed against predicted C stock alongside the 95% intervals are illustrated in **Figure 3-7**. Slight improvement in the performance of Sentinel-2 and Landsat-8 C stock-based predictions can be validated through a visual inspection of the results. Estimated Root Mean Square Prediction Error (RMSPE) for Landsat-8 (6.42 Mg C ha^{-1}) and Sentinel-2 (6.23 Mg C ha^{-1}) based C stock prediction further reinforces the model prediction performance for the two sensors shown in **Figure 3-7**.

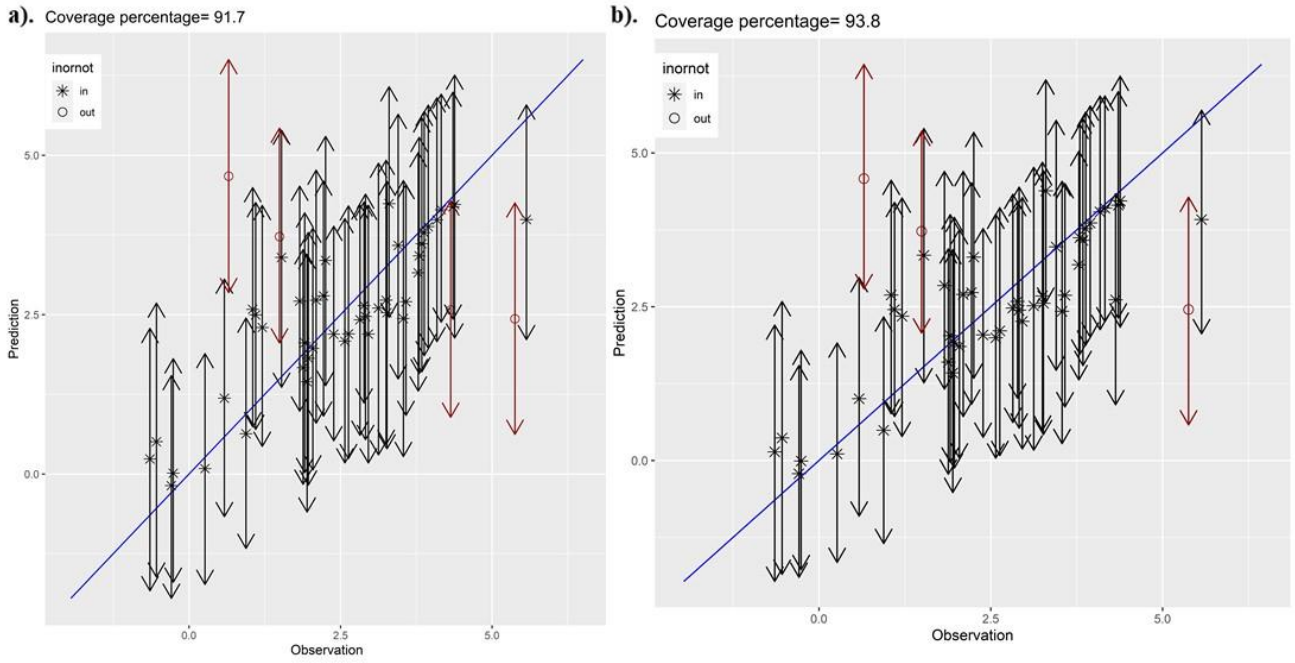


Figure 3-7: (a) Landsat-8 and (b) Sentinel-2 C stock-based predictions vs. C stock observations alongside 95% intervals.

3.4. Discussion

Space varying coefficient models constructed from different but related multispectral remote sensing platforms were used to predict C stock in a disturbed plantation forest ecosystem in Zimbabwe. The RMSPE is marginally higher in Sentinel-2 driven C stock SVC model than in the Landsat-8-based SVC counterpart. Our findings with regard to the performance of Sentinel-2 and Landsat-8 sensors are also congruent with the work of Chrysafis *et al*. (2017) and Korhonen *et al. (*2017*).* SVC models for both new-generation remote sensing-derived predictors showed preference over the error-independent models. However, estimates from the two data sources were marginally different from each other on the prediction of C stock, as illustrated in validation diagnostics (Chen *et al.*, 2012). The adaptable structure of the SVC permitted the residuals spatial variability to be apportioned between the random slope and the random intercept. This provided additional benefits not available in the SVI models from the two predictor sources. The SVC permitted reconnaissance of the C stock observations and the $NDVI$ and $DIST$ predictors. Furthermore, the SVC models in both multispectral data sources had better representation of the processes thereby yielding C stock predictions with reduced variance.

Evaluation of the models utilizing predictors from both remote sensing data sources showed SVC to fit the data better than the SVI models. Kriged maps for C stock using Sentinel-2 and Landsat-8 data were not significantly different from each other, with the Landsat-8 SVC displaying a slightly wider 95% CI compared to the Sentinel-2-based SVC model. Again, this

is partly because the study employed conventional bands (indices) that are calculated in a similar fashion in both Landsat-8 and Sentinel-2, and hence, the differences in prediction only emanated from the spatial resolutions. Explicit accommodation of residual spatial dependence through spatially correlated random effects gathered better predictions as the SVC models using the two data sources as regression coefficients borrowed additional information from neighboring C stock observations (Gelfand, 2012; Babcock *et al.*, 2016). Precise estimates of covariance parameters are not easy to derive with small inventory sample sizes (Baugh and Groeneveld, 2006). Consequently, we could anticipate some impact on the accuracy of the predictions when uncertainty in the covariance ensues to the posterior predictive distributions. Predictions had limited information to borrow from because of the sparsity of C stock inventory observations. This was further worsened by the reduction in the overall sample size through cross-validation.

Similar to previous studies predicting ABG and C stock, our findings establish Sentinel-2 as a better source of RS data for predicting C stock in disturbed environments compared Sentinel-2 and Landsat-8 imagery for forest biomass prediction and showed Sentinel-2 outperforms Landsat-8 because of the enhanced spatial resolution in the former in comparison to Landsat-8. Most studies comparing Sentinel-2 and Landsat-8 for predicting AGB prefer Sentinel-2 over Landsat-8. This is further justified by the work of Korhonen *et al.* (2017), who compared Worldview-3, Sentinel-2, and Landsat-8 for representing AGB in a forest environment in Thailand and demonstrated Worldview-3 and Sentinel-2 as better data sources and, therefore, predictors than Landsat-8 due to the red-edge and the improved spatial and spectral properties of Worldview-3 and Sentinel-2. Furthermore, Green, Finley and Strawderman (2020) utilized LiDAR derived covariates from establishing the prediction performance of SVC models using forest inventory data in North America. The researchers established significant improvement in biomass prediction accuracy in the presence of residual spatial dependence deriving from the finer resolution LiDAR covariates. In most cases, the effectiveness of SVC models in these studies is usually strengthened by the solid non-stationary relationships between the response variable and the independent variables influenced by unobserved ecological factors operating at broad geographical scales. Such ecological factors were also seen in the present study as NDVI was established to be a statistically significant predictor of C stock in the studied region. The biggest drivers to these factors are the presumed activities of forest disturbance due to human encroachment into plantation forests that subsequently impact the density and distribution of ABG biomass.

3.5. Limitations of the study

Apart from the vegetation indices influencing the spatial distribution and density of AGB employed in this study, it is also known and acknowledged in the literature that climate and topographic variables play a part in the distribution of C stock. For example, Korhonen *et al.* (2017) and Ranghetti et al. (2020) have shown elevation and aspect accounting for the bigger portion of the spatial variability of C stock in a mountainous region of Nepal. As such, the limitation of our study is the application of vegetation indices as predictors of C stock, and this may, therefore, not be representative of the general C stock dynamics in the studied region. We, therefore, recommend the integration of topo-climatic factors with new generation remote sensing-derived vegetation indices for future research in order to obtain a more accurate global overview of the C stock density and distribution in the studied region.

3.6. Conclusions

The study presented a hierarchical Bayesian geostatistical spatially varying coefficient model for determining the relationship between sampled C stock data and multispectral remote sensing derived predictors. There was a marginal improvement in model fit, and prediction accuracy in both Landsat-8 and Sentinel-2-based SVC models in comparison to the error-independent models. The SVC model permitted exploration of the observed C stock locations where the models performed well or poorly, which was missing in the SVI models. This provided an understanding of the performance of the multispectral remote sensing derived predictors for modelling C stock and hence, sets the foundation for the updating of the carbon forest plantation database for forest practitioners in the country and utilized as a monitoring tool in the long term. The Sentinel-2-based SVC model was preferred for prediction in the plantation forest ecosystems as its model provided tighter credible intervals compared to the Landsat-8-based C stock SVC model.

The small sample size of the data utilized in the present research enabled the modelling approach to be computationally feasible. When inventory plots are comprised of bigger sample sizes, the matrix operations of immense dimensions are needed for computing model parameter estimates of higher magnitude and may not be achievable through ordinary PCs. Our future work is therefore aimed at exploring algorithms for resolving the dimensionality curse when fitting spatially varying coefficient models. The problem of dimensionality can also get complicated if many predictors are involved in the SVC modelling framework. Resolving dimensionality issues is needed as forest C stock is typically modelled with predictors from

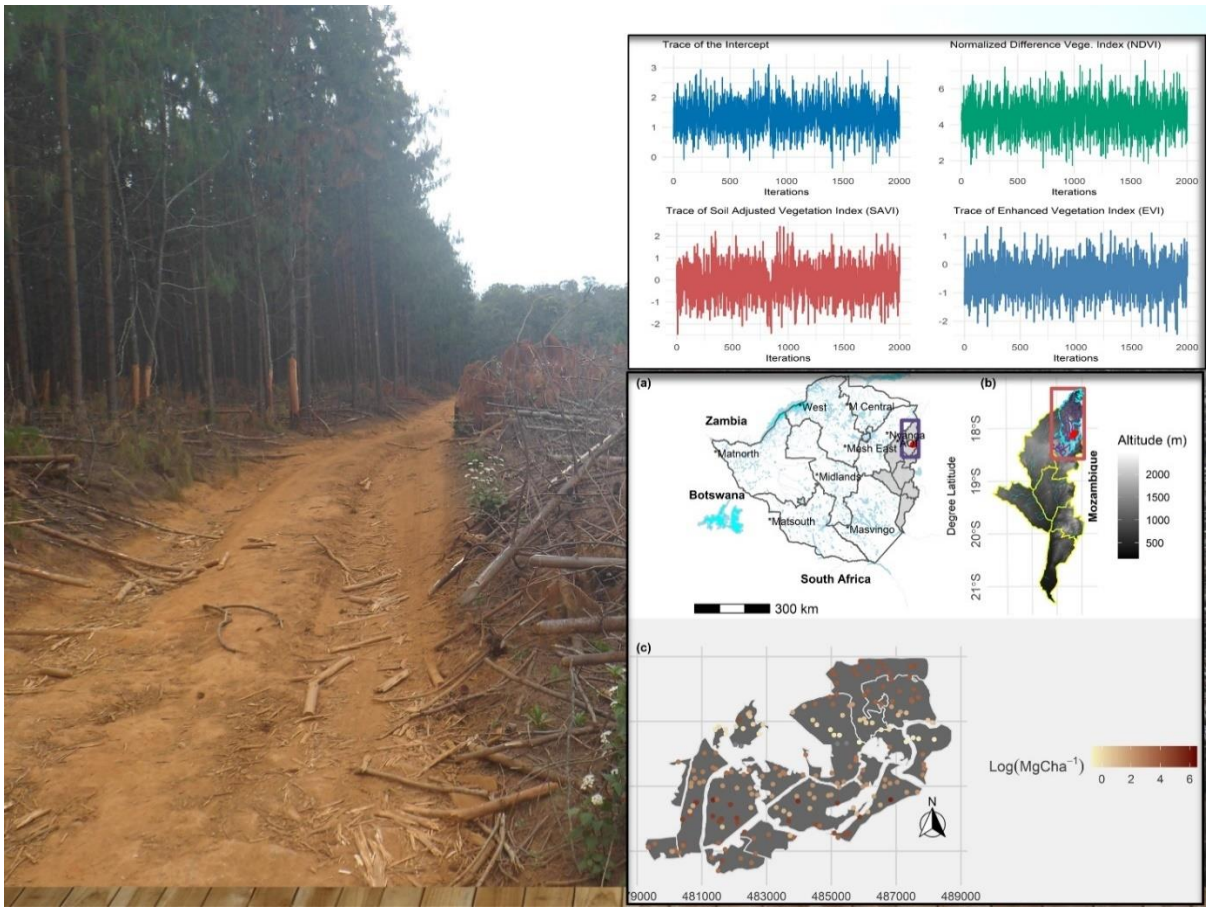
many data sources, chief amongst them being topographic, bioclimatic, and anthropogenic variables.

3.7. Summary

This chapter presented two spatially varying coefficient Bayesian hierarchical models deriving fixed effects parameters from the Sentinel-2 and Landsat-8 satellite new generation remote sensing platforms. Using the same basket of predictors as in chapter two, the sentinel-2 driven C stock predictive Bayesian hierarchical model gives lower prediction uncertainty of C stock than the Landsat-8 model. However, the prediction uncertainty of the best SVC C stock model constructed from the Sentinel-2 derived spatial random effects is not better than the Sentinel-2 driven C stock prediction model without the non-stationarity assumption. The biggest limitation of SVC models regards their high demand for computational power, making the utilisation of such models with bigger sample sizes and a large suite of model predictors unfeasible with ordinary computers. As such, both Landsat-8 and Sentinel-2 driven C stock predictive models were only restricted to spatial random effects derived from vegetation indices, yet the source of such spatial variation is broad in practice. In order to exhaust possible sources of spatial variation of the modelled variable at local scale, the next chapter adopts a multi-source data approach to C stock modelling using the Bayesian geostatistical methodology. The methodology incorporates multi-data sources of predictors ranging from climatic, topographic, anthropogenic and vegetation indices in order to ascertain whether C stock prediction uncertainty can be improved by broadening the range of spatial random effects or taking spatial heterogeneity of C stock into consideration.

CHAPTER FOUR

A multi-source data approach to Carbon stock prediction using Bayesian hierarchical geostatistical models in managed forest ecosystems



This chapter is based on

Tsikai Solomon Chinembiri*, Onisimo Mutanga, Timothy Dube, "A multi-source data approach to Carbon stock prediction using Bayesian hierarchical geostatistical models in managed forest ecosystems, GIScience and Remote Sensing (Manuscript ID: 232584508), 2023, **Under Review**

Abstract

Using a Bayesian hierarchical inferential framework, we employed a multi-source data approach (i.e. remote sensing derived anthropogenic, climatic and topographic set of variables) to model Carbon (C) stock in a managed plantation forest ecosystem in Zimbabwe's Eastern Highlands. We therefore investigated how two related multi-data sources of new generation remote sensing derived ancillary information influence C stock prediction required for building sustainable capacity in C monitoring and reporting. The year 2000 agrarian land reform programme in areas proximal to plantation forests led to significant reduction in forest cover and modification. Two mainstream models constructed from the Sentinel-2 and the Landsat-8 derived vegetation indices coupled with climatic and topographic covariates were used to predict C stocks using forest inventory data collected using spatial coverage sampling. A multi-source data driven approach to C stock prediction yielded slightly lower predictions for both the Landsat-8 ($1 \leq MgCha^{-1} \leq 135$) and the Sentinel-2 ($1 \leq MgCha^{-1} \leq 130$) based C stock models than C stock predictions published in related studies. Distance to settlements ($DIST$) and $NDVI$ appeared to be significant predictors of C stock with the Sentinel-2 based C stock model outperforming its Landsat-8 model variant in terms of prediction accuracy. The Sentinel-2 driven C model resulted in a 1.17 $MgCha^{-1}$ RMSE with a ($1 \leq MgCha^{-1} \leq 4$) 95% credible interval whilst the Landsat-8 based C stock counterpart gave a 2.16 $MgCha^{-1}$ RMSE with a ($2 \leq MgCha^{-1} \leq 4$) associated 95% credible interval. Despite a multi-source data prediction approach to the modelling of C stock in a managed plantation forest ecosystem set-up, the issues of scale still play a major role in modelling spatial variability of natural resource variables. Both climatic and topographic derived ancillary data are not significant predictors of C stock under the present modelling conditions. Landscape scale studies employing a multi-source data approach to Above Ground Biomass (AGB) modelling report mean annual rainfall and mean annual temperature as significant predictors of C stock. However, the results of the present study do not confirm climatic variables as significant predictors of the same. Due to the complexity of the relationship between AGB and climate, the results of this study do not establish an exhaustive relationship between plantation forest C stock and remote sensing based multi-data sources as the temporal aspect of climatic variables also exert a major role in C stock dynamics. Accurate and precise accounting of C stock for climate change mitigation and action can best be done at landscape scales rather than local scale as the scale of variation for climate change related variables vary at larger spatial scales than the ones utilised in the present study.

Key words:

MCMC; geostatistics; hierarchical; Bayesian; prediction; remote sensing

4.1. Introduction

The gradual and steady progression of cutting-edge earth observation technologies and platforms with the capability of imaging the earth's surface at reasonable and acceptable spatial scales in the past two decades has greatly enhanced natural resource modelling methodologies (Finley and Banerjee, 2008). Accurate prediction of forest resources at both local and landscape scales is principally dependent on the availability of pertinent data regarding forest resource distribution and density drivers (Jiang *et al.*, 2022). Consequently, natural resource dynamics are driven by environmental and anthropogenic factors that cannot easily be obtained from a single remote sensing platform, thereby establishing the need for a multi-source approach in the acquisition of data required for modelling resource variables (Baccini *et al.*, 2004). According to Finley and Banerjee (2008), multi-source data refers to different sources of data sharing a common spatial reference system and can be coupled to come up with sets of outcome variables, $y(\boldsymbol{s})$, and predictors, $\boldsymbol{x}(\boldsymbol{s})$, where the $\boldsymbol{s}$ represents a known location in $\mathbb{R}^2$, usually easting–northing or latitude–longitude. Remote sensing is an essential supplementary tool for providing a comprehensive overview of high spectral and spatial resolution data necessary for overcoming some of the time, labour cost and inaccessibility constraints faced in forest biomass measurement (Tran, Reef and Zhu, 2022).

The Sub-Saharan African region has one of the overexploited forest resources in the world, posing great threat to the global climate mitigation strategies under the United Nations Framework Convention on Climate Change (UNFCCC, 2020). The high rate of exploitation of naturally established forests for agriculture and industrialization are some of the biggest human driven factors that have given rise to the expansion of planted forests in Africa (Do *et al.*, 2022). For instance, overall plantation area in Africa was 8 046 000 ha, now subdivided into industrial (3 392 000 ha), non-industrial (3 273 000 ha) and unspecified plantations (1 370 000 ha) in 2000 (FAO, 2003). This figure represented approximately 4.2 % of the global plantation area (FAO, 2003). As highlighted by FAO (2003), the total plantation area of Africa and its yearly planting area represents the lowest planted area among all the continents of the world and action policies are therefore vital to improve the figures. The proportion of planted forest area by region gives an indication of the extent of carbon (C) sequestration and the seriousness of governments in combating climate change impacts through reduction in forest degradation.
Zimbabwe saw a 20 % decline in forest plantation area from 96 835 ha in 2006 to 77 716 ha in 2015 (Forestry-Commission, 2021) due to the government's agrarian land reform policy that started in 2000 and ended in 2004. The agrarian land reform resulted in some former farm workers and families in formerly owned commercial timber and agriculture estates becoming

unemployed. This development saw a growth in population in need of socio-economic resources for survival in these communities. Coupled with prolonged phases of drought and tropical cyclones as a result of climate change in the forest plantation endowed regions of Zimbabwe, most youths and women were left with limited options for survival and started opening up new land for farming, grazing and illegal gold panning activities (Matose, 2008). The potential and seemingly diverse sources of spatial variation of AGB in plantation forest ecosystems in Zimbabwe as driven by the agrarian land reform and potential climate change dynamics require exhaustive modelling for accurately predicting AGB distribution. Precise and accurate prediction of plantation forest C stock under these circumstances is therefore not simple without a comprehensive and multi-source inferential approach to the resource prediction problem. It is also notable that almost all work done on the assessment of topographic and climatic variables as key drivers of forest biomass and forest productivity have been undertaken in forests that do not have any disturbance from human impacts. However, substantial tropical forest land is being modified and reduced through exploitation from human activities, yet it is in these regions where the prospects for REDD+ is greatest (FAO, 2016).

Despite location and existing management regimes of forest resources, climate is well known for exerting a greater influence on the distribution and storage of C in global forests (Pan *et al.*, 2013). Precipitation and temperature are clearly related to biomass storage (Keith, Mackey and Lindenmayer, 2009, Pan et al. 2013) through the limitations they exert on biomass production. Precipitation controls water availability, which successively affects nutrient uptake, leaf area index, stomatal conductance and eventually forest productivity (Eamus, 2003). The rate of carbon dioxide absorption in tree leaves and C losses resulting from respiration needed for maintenance of living tissue is regulated by temperature (Larjavaara and Muller-Landau, 2012). Within ranges of spatial scales under diverse forest conditions, Mean Annual Precipitation (MAP) (Slik *et al.*, 2010, Liu *et al.*, 2014, Timmons, Buchholz and Veeneman, 2016, Poorter *et al.*, 2016) and Mean Annual Temperature (MAT) (Liu et al. 2014, Vieilledent *et al.*, 2016, Gordon *et al.*, 2018) are strongly correlated with forest biomass. Besides, temperature can also limit biomass production through seasonality (Vieilledent et al. 2016), the minimum and maximum temperatures of the coldest and warmest months (Ghasemi, 2015) and the duration of growing days (Ali *et al.*, 2019). On the other hand, precipitation can also limit forest productivity through seasonality (Slik et al. 2010, Poorter et al. 2016), variability (Álvarez-Dávila *et al.*, 2017), and by influencing the extent (SAATCHI *et al.*, 2007) and duration of dry periods (Saatchi et al. 2007).

With regard to the framework of climate change, classification of the temporal and spatial variability of forest AGB is critical for both local and regional C cycle prediction (Wang *et al.*, 2022). A number of studies are documented in literature detailing how AGB is predicted, using earth observation data and in some of these studies, the impacts of topographic features and spectral data on forest biomass were investigated (Hu *et al.*, 2016, Liu *et al.*, 2019, Serrano, Corral-Rivas and López-Sánchez, 2019, Hu *et al.*, 2020, Puliti *et al.*, 2020). Simard *et al.* (2019) established precipitation, temperature and storminess as responsible for accounting for more than 74% of the world trends in the maximum values of above-ground biomass and canopy height. Serrano, Corral-Rivas and López-Sánchez (2019) utilised climatic (rainfall, precipitation), vegetation indices (Normalised Difference Index (NDVI) and Soil Adjusted Vegetation Index (SAVI)) and topographic (elevation, slope, aspect) variables for predicting AGB, using machine learning methods like the Support Vector Regression (SVR) and Support Vector Machines (SVM) in Mexico. The SVM performed better giving a RMSE of 8.20 $Mgha^{-1}$ compared to the RMSE of 13.05 $Mgha^{-1}$ for the SVR for predicting AGB (Serrano, Corral-Rivas and López-Sánchez, 2019). Regions with the highest biomass density corresponded to areas with higher precipitation, lower temperature and higher elevation. The SVM AGB predictive accuracy is better than reported in similar machine learning studies that gave AGB RMSE ranging from 27-40 $Mgha^{-1}$ (López-Serrano *et al.*, 2016). In the same context, Zhang *et al.* (2023) utilised vegetation indices, MAT and MAP to model grassland biomass in Xinjiang using SVM, RF and exponential regression where the RF technique outperformed the other methods with a RMSE grassland AGB of 5.46t/ha. Further, Hu *et al.* (2016) estimated global forest biomass through the integration of ground inventory data, topographic data, climate surfaces, spaceborne LiDAR and optical imagery at 1-km resolution and established AGB global map with 87.53 $Mgha^{-1}$ RMSE.

We employ the hierarchical Bayesian inferential framework that has earned recognition in spatial modelling of resource variables in recent years. Modelling of both the observed data and any unknown predictor effects as random variables makes the hierarchical Bayesian methodology an integrated approach for incorporating sophisticated data models and pre-experimental knowledge to statistical analysis (Gelman, 2006). This is particularly important for reinforcing national and global climate change policy strategies without the need for extensive and expensive forest inventories. The capitalisation of pre-experimental information through the introduction of priors in the Bayesian approach makes it possible to report progress on C sequestration in circumstances where expensive and extensive forest inventories may not be feasible. As with all hierarchical Bayesian modelling techniques, the benefit is full and exact

inference, with accurate assessment of uncertainty (Gelfand, 2012). We therefore employ a multi-source data approach to the prediction of plantation forest C stock using Sentinel-2 and Landsat-8 vegetation spectral indices, climatic and topographic variables within a Bayesian geostatistical hierarchical framework. We present the current study under the postulate that a Sentinel-2 driven multi-source prediction approach offers lower C stock prediction uncertainty than the Landsat-8 driven multi-source C stock model within a Bayesian hierarchical modelling framework.

4.2. Methodology

The study covered one of Zimbabwe's commercial forests at Lot 75 A of Nyanga Downs in Nyanga district in the eastern highlands of Zimbabwe. Apart from *Pinus patula*, *Eucalyptus grandis* and *Eucalyptus camaldulensis* as the major tree species characterising the studied region, substantial portions of the area has patches of land opened for agriculture and gold panning by neighbouring farmers (European-Commission, 2017). Our study area is located between longitude 32º40ˈ E and 32º54ˈ E and 18º10ˈ12ˈˈ S and 18º25ˈ4ˈˈ S latitude as illustrated in **Figure 4-1**. Grazing and gold panning activities in the neighbourhood of the forest plantation started after the redistribution of some formerly owned commercial forest estates to peasant farmers in 2000 (Zvobgo and Tsoka, 2021;). An area of approximately 2 767 ha is covered by the study area with variable rainfall amounts spanning from 740 mm to 2,998 mm and a mean annual rainfall of 1151 mm. The area experiences hot weather characterised by extensive wild veld fires around high-altitude grassland areas from September to November. Annual average temperature range between 9^0 C and 12^0 C and 25^0 C and 28^0 C are experienced during the winter and summer seasons, respectively (Initiative For Climate Action Transparency (ICAT), 2022).

4.2.1. Remote sensing data

The geographic coverage of earth observation data such as Sentinel-2 and the Landsat series, coupled with their easy and cheap availability provides immense opportunities for modelling natural resources of interest at relatively low cost. We make use of Landsat-8, Sentinel-2, WorldClim (Fick and Hijmans, 2017) and SRTM data to derive predictors for carbon stock prediction in a managed forest ecosystem in the eastern highlands of Zimbabwe. We obtained Landsat-8 OLI imagery data on the 20th of September 2020 from the Earth Explorer (http://earthexplorer.usgs.gov) website as a processed product ready for analysis.

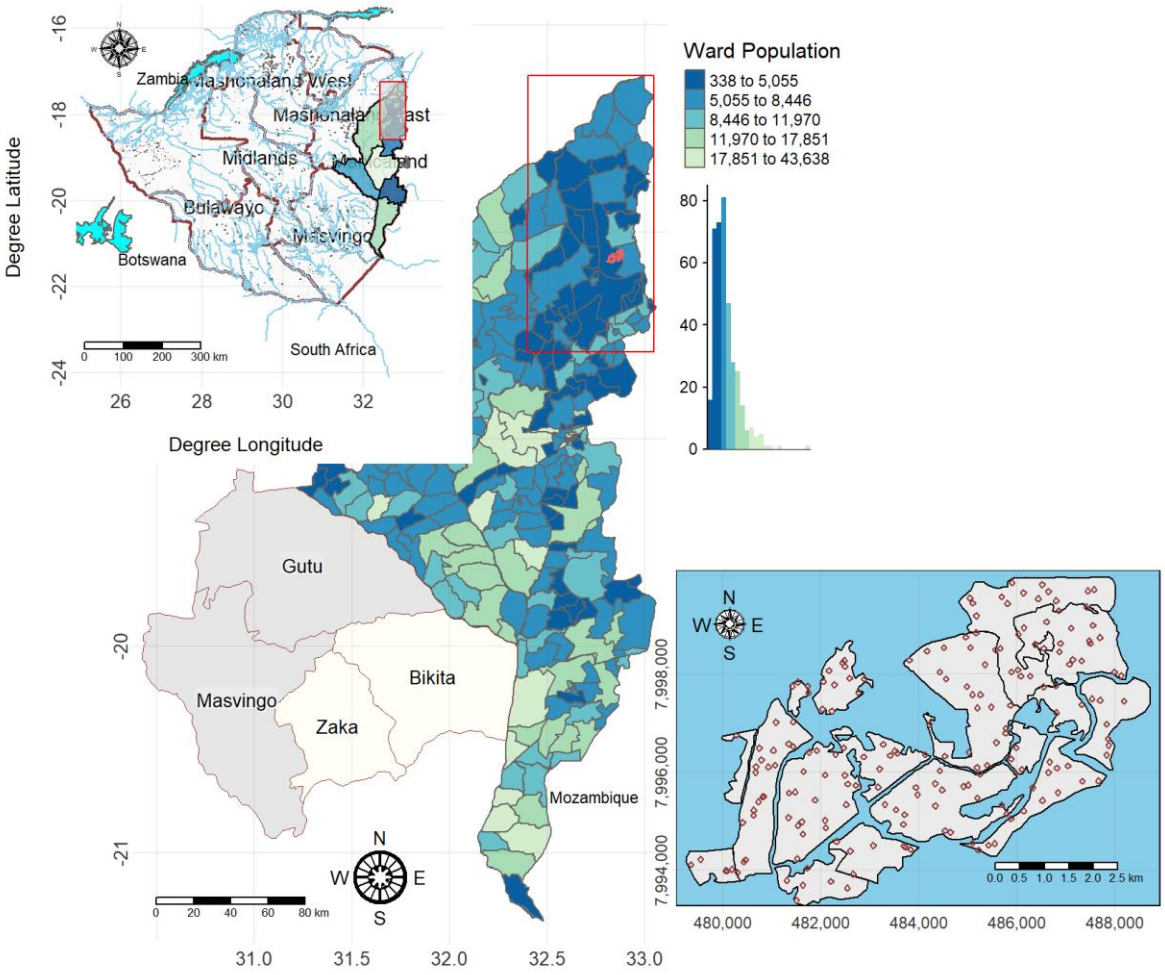


Figure 4-1: Map of the study area indicating (a) the province where samples were derived, (b) study area location within the particular province and (c) the spatial distribution of the sampled regionalised variable in the lower panel.

The dataset was rid of cloud cover with cloud shadow thresholds set to less than 10 %. We also acquired the Sentinel-2 imagery covering the entire study area (lot 75A Nyanga Downs in Manicaland province of Zimbabwe) at the same date and time that we acquired the Landsat-8 satellite data. Sentinel-2 data was obtained as level-1C 12-bit pre-set TOA reflectance values. We pre-processed and orthorectified Sentinel-2 1-C product in the *R Statistical and Computing Environment* using the *sen2r* library (Ranghetti *et al.*, 2020). The Sentinel-2 sensor records data in 13 spectral bands with 4 in the visible range with 10 m spatial resolution, 6 in the near-infrared region with 20 m spatial resolution and 3 bands in the short-wave infrared region with a 60 m spatial resolution (Gerald *et al.*, 2017). The same observation can therefore be imaged by a Landsat-8 and a Sentinel-2 satellite at varying spatial resolutions, spectral bandwidth and spectral response function (Zhang et al., 2018). All the orthorectified infrared and visible bands of both Landsat-8 and Sentinel-2 were assembled, mosaicked and clipped in order to have a subset raster of the sampled area.

4.2.2. Anthropogenic variables

The distance to settlements ($dist$) within the sampled region was utilised as an indicator of accessibility to forest resources by settlers within the plantation forest. Settlements were manually digitized using a combination of high-resolution Google Earth imagery and the revised 1:50 000 topographic maps that covered the study area (Surveyor General, 2018). We then calculated the distance to settlements from each of the sampled C stock plots for the Sentinel-2 and the Landsat-8 satellite image data within a GIS environment. The need for modelling C stock with distance to settlements is based on the principle that accessibility to forest plantations exerts pressure on forest resources and increases the likelihood of indirect impact on the planted forest (Mon *et al.*, 2012; Chinembiri *et al.*, 2013).

4.2.3. Climatic data

Because of the significant effect climatic conditions have on the density of AGB, we considered the Mean Annual Temperature (^{0}C) (MAT) and the Mean Annual Rainfall (mm) (MAP) in the modelling of C stock stretching a 30-year period from 1970 to 2000 from the WorldClim dataset (Fick and Hijmans, 2017). WorldClim version 2.1 climatic data was supplied at a spatial resolution of 30 arc seconds ($\approx$ 1 km). Previous studies including MALHI *et al.* (2006) and Liu *et al.* (2014) have shown a significant association between MAT and MAP to forest biomass.

4.2.4. Topographic data

We acquired slope, elevation and aspect data as the main topographic factors influencing forest biomass distribution (Su *et al.*, 2016) from the Shuttle Radar Topography Mission (SRTM). We employed the 30 m SRTM digital elevation model (DEM) in deriving the aforementioned topographic variables. The SRTM encompasses almost 99.96% of the global earth's surface from 56^0 S to 60^0 N (Rabus *et al.*, 2003; Farr *et al.*, 2007). We re-projected all the datasets to a common grid reference system and rescaled to 1 km spatial resolution using the nearest neighbour interpolation approach before bringing different multi-source data sets together (Christman and Rogan, 2012). This was done in order to ensure consistency with the spatial resolution of the coarser dataset (1 km climatic dataset) and minimise error propagation during the data harmonisation process (Robinson, 1950).

4.2.5. Vegetation indices

We utilized Normalised Difference Vegetation Index ($NDVI$), Soil Adjusted Vegetation Index ($SAVI$), Enhanced Vegetation Index (EVI) derived from the Sentinel-2 and the Landsat-8

multispectral remote sensing data. The vegetation indices used in this study have also been applied as ancillary variables in literature (Li and Li, 2019; Bordoloi *et al.*, 2022) for biomass estimation because of their sound correlation with forest biophysical variables. Hence, we selected the broadband vegetation indices based on simplicity and robustness.

4.2.6. Uniform area spatial coverage sampling

Our C stock forest inventory data collection programme was done in an area with no history of sampling, implying lack of knowledge of scales of spatial variability. We therefore employed the *k*-means clustering algorithm for uniform area spatial coverage sampling that makes use of the Mean Squared Shortest Distance (MSSD) for optimization of sample locations (Brus, de Gruijter and van Groenigen, 2006; Walvoort, Brus and de Gruijter, 2010). As illustrated in **Figure 4-1**, even distribution of sampling locations across the sampling area enhances the mapping and estimation of regionalised variables (Walvoort, Brus and de Gruijter, 2010). Brus, de Gruijter, and van Groenigen (2006) demonstrated how even distribution of sample locations can be applied for estimating both spatial means and the resolution of mapping in environmental, forestry and soil science research. The suitability of the uniform area spatial coverage design also befits application in areas where sampling campaigns cannot be continued beyond a single phase.

4.2.7. Field sampling and tree biomass measurement

Sampling was done and measurements taken for all trees with at least 10 cm diameter at breast height (DBH) (~1.28 m) using circular plots with an area of 500 m^2 between the 19th of September 2021 and the 24th of October 2021. Gibbs et al. (2007) suggest that individual trees with DBH measurements less than 10 cm store negligible amounts of C stock. We did not consider slope correction during the tree biomass measurement exercise as the slopes in the studied region generally fall below 30 % (Ravindranath and Ostwald, 2008). Optimisation of the sampling locations through uniform coverage sampling using the *spcosa*-library of the *R Statistical and Computing Environment* resulted in 200 samples (Brus, de Gruijter and van Groenigen, 2006). We pre-loaded the 200 samples into a 72 H handheld Garmin GPS before launching the field data collection programme and 191 samples of forest biomass measurements were used in the final analysis as the other nine samples fell outside the sampling region.

4.2.8. AGB calculation and C stock derivation

Three major plantation forest species, namely *Pinus patula*, *Eucalyptus grandis* and *Eucalyptus camaldulensis* are found in the study area. We therefore derived the biomass of each of the tree

species using specific allometric equations. Allometric equations by Brown (1997) were used for the calculation of the *Pinus patula* species tree biomass whilst we derived the tree biomass for the *Eucalyptus* species using allometric models by Zunguze (2012). The same allometric equations for the same tree species were utilised in the calculation of the Pinus and the Eucalyptus species biomass in Manica Province in Mozambique. Manica province of Mozambique closely resembles the climatic conditions of the studied region which it shares the same border in Manicaland province of Zimbabwe. We then converted the AGB of each and every individual tree species to C stock using a transformation factor of the IPCC (2006). Per plot value estimates were then calculated to a $MgCha^{-1}$ standard AGB unit of measure.

4.3. Bayesian hierarchical modelling

We modelled the observed C stock data in a hierarchical and spatial regression driven model as:

$$Y(s) = \boldsymbol{x}^T(s)\boldsymbol{\beta} + w(s) + \varepsilon(s) \quad (4\text{-}1)$$

Where;

$Y(s)$ denotes the observed C stock data at a spatial location s,

$\boldsymbol{x}(s)$ denotes a vector of predictors

$w(s)$ denotes a spatial process capturing the residual spatial association

$\boldsymbol{\beta}$ denotes a vector of regression coefficients

$\varepsilon(s)$ captures the random error (nugget effect) that is assumed $i.i.d.N(0, \sigma_\varepsilon^2)$

Simultaneous estimation and prediction of the response variable parameters was done using the Markov Chain Monte Carlo (MCMC) algorithm for deriving and calculating the predictions of the response variable at unsampled sites according to Eqn 4-2:

$$\hat{Y}(s)' = X^T(s)'\hat{\beta} + \widehat{w}(s)' \quad (4\text{-}2)$$

We made the hierarchical modelling using the *spBayes* R Statistical and Computing Environment library (Finley, Sudipto and Carlin, 2007; R Core Development Team, 2008). We treated the vector of all model parameters, that is, $\boldsymbol{\theta} = (\beta, \sigma^2, \phi, \tau^2)$, as random variables and were thus assigned individual prior distributions. The posterior distribution of parameters, $\boldsymbol{\theta}$, was sampled in accordance with Eqn 4-3:

$$p(\boldsymbol{\theta}|\boldsymbol{y}, X) \propto p(\boldsymbol{\theta})\ x\ N(\boldsymbol{w}|0, \Sigma_w)\ x\ N(\boldsymbol{y}|X^t\boldsymbol{\beta} + \boldsymbol{w}, \Sigma_\varepsilon) \quad (4\text{-}3)$$

Eqn 4-3 was therefore employed in order to measure the uncertainties in model parameter and predictions of the response variable at unvisited sites using the derivation in Eqn 4-4 where the y_0 denotes predicted C stock at unsampled sites using covariates x_0 at site s_0.

$$p(y_0|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{x}_0) \propto \int p(y_0|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{x}_0)p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X})d\boldsymbol{\theta} \quad (4\text{-}4)$$

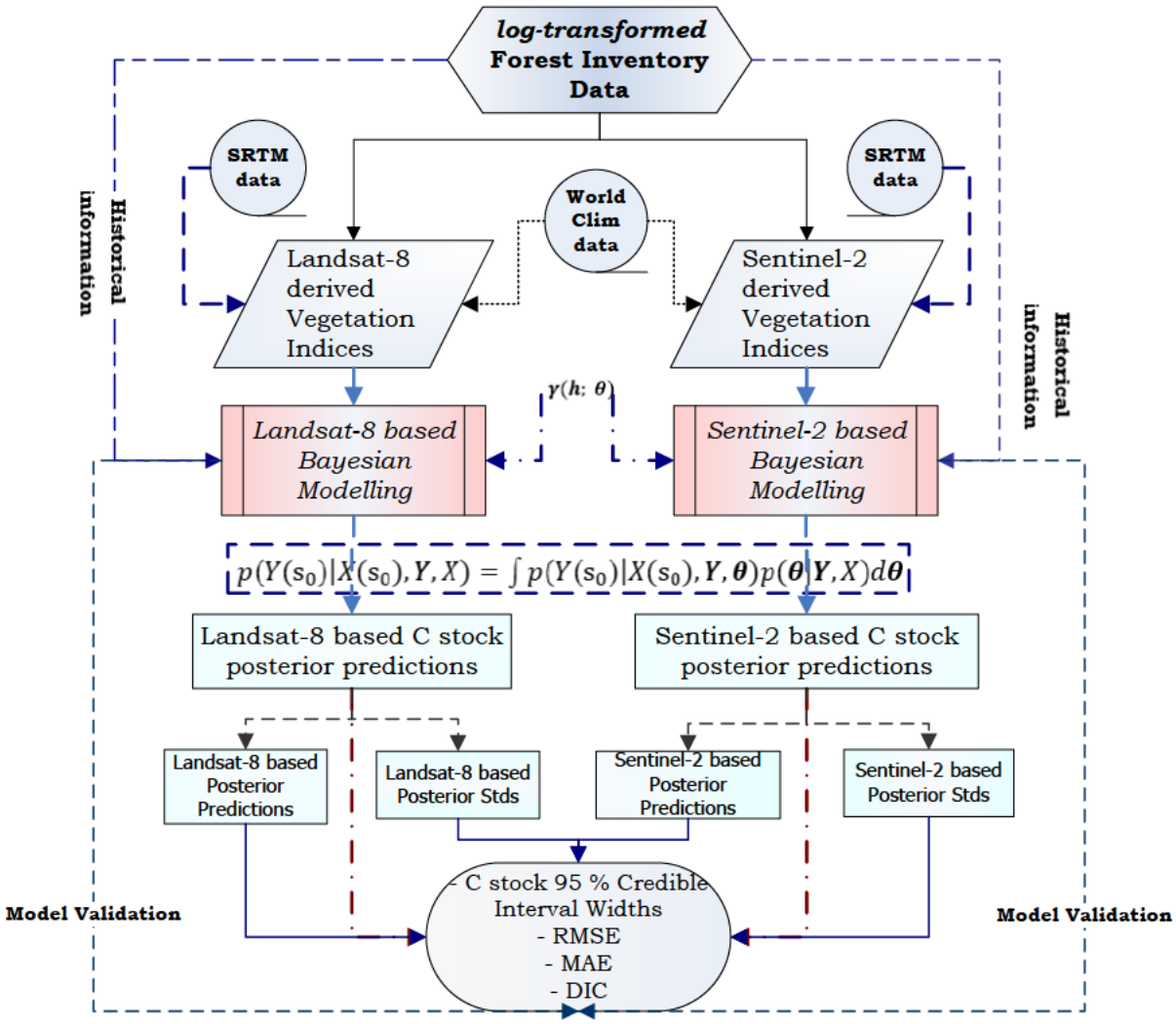


Figure 4-2: Multi-source data model prediction framework

The study relied on a multi-source data supply from Landsat-8, Sentinel-2, climatic and topographic C stock related ancillary variables. Sentinel-2 and Landsat-8 multispectral remote sensing data formed the mainstream Bayesian hierarchical model building blocks where vegetation indices derived from them were integrated with anthropogenic, climatic and topographic variables from WorldClim (Fick and Hijmans, 2017) and the SRTM data (Rabus *et al.*, 2003). Two types of models were therefore constructed with prior specifications designated as $p(\boldsymbol{\theta_1})$ and $p(\boldsymbol{\theta_2})$ for the Landsat-8 and Sentinel-2 based C stock models, respectively.

We assigned the inverse gamma distribution for the data (C stock) and nugget variance whilst the spatial decay parameter was given a uniform prior. The maximum distance between the farthest sample locations $(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \in D$ in the studied area (2413 m) guided the assignment of the prior distribution of the spatial decay parameter. In the same vein, scale parameter values were assigned in order to convey the preference that the nugget error variance $(\sigma_\varepsilon^2) <$ the data

variance (σ_w^2) (Demirhan and Kalaylioglu, 2015). We utilised the Metropolis-Hastings technique for the MCMC (Finley, Banerjee and MacFarlane, 2011). We based the analysis on three chains of 11, 000 samples each where 1, 000 samples were discarded as part of burn-in. Subsequent analysis and parameter estimation therefore depended on the remaining 30, 000 (10, 000 x 3) samples. A summarised flow diagram of the data processing and handling using the Bayesian hierarchical approach is illustrated in **Figure 4-2**.

4.3.1. Bayesian geostatistical model validation

Each of the Landsat-8 and Sentinel-2 based spatial intercept only, independent error and the spatial models were evaluated for their prediction performance using a hierarchical modelling approach with $p(\boldsymbol{\theta}_1)$ and $p(\boldsymbol{\theta}_2)$ being used as guides for prior distribution specification as follows:

$p(\boldsymbol{\theta}_1) = \boldsymbol{Unif}(\phi|0.39, 0.0013)\ x\ \boldsymbol{IG}(\sigma^2|0.53, 1.59)\ x\ \boldsymbol{IG}(\tau^2|0.1, 1.59)\ x\ \boldsymbol{MVN}\left(\beta\middle|\mathbf{0}, \Sigma_{\boldsymbol{\beta}}\right)$

$p(\boldsymbol{\theta}_2) =$

$\boldsymbol{Unif}(\phi|0.39, 0.0013)\ x\ \boldsymbol{IG}(\sigma^2|0.053, 0.0029)\ x\ \boldsymbol{IG}(\tau^2|0.1, 1.53)\ x\ \boldsymbol{MVN}\left(\beta\middle|\mathbf{0}, \Sigma_{\boldsymbol{\beta}}\right)$

A k −fold cross validation algorithm was utilised for evaluating the prediction performance of the spatial intercept only, error independent and the spatial models through random spliting of the 191 sampled C stock data into ten equally sized segments (Duchene *et al.*, 2016). The Root Mean Square Error ($RMSE$) and the Mean Absolute Error (MAE) in addition to other goodness of fit metrics were calculated and used as ranking criteria for models' ability in fitting the data (Spiegelhalter et al., 2002). Models regarded as better fitting have lower values of the Deviance Information criterion (DIC) whilst top performing models have the smallest $k(k = 10)$ fold $RMSE$ and MAE statistics. Graphical trace plots were also used as diagnostics for assessing the adequacy of modelling assumptions (Jackman, 2000; Green, Finley and Strawderman, 2020).

4.3.2. Predictive model uncertainty assessment

Multi-source predictor variables utilised in the Landsat-8 and Sentinel-2 driven C stock models were assessed for significance using the 2.5% and the 97.5% posterior sample percentiles. Independent variables in the hierarchical models were certified significant if their 95% Credible Intervals (CIs) did not contain zero, with uncertainty being assessed according to Hengl et al. (2004) method. We displayed the median predictions of C stock for each grid cell (100 m x 100 m), using a colour ramp together with the associated 95% Credible Interval

Widths (CIWs). Predictions with higher uncertainties have wider CIWs (Babcock *et al.*, 2016, 2018).

4.4. Results

The spatial correlation structure of Landsat-8 and Sentinel-2 based C stock models build from a *WorldClim* and *SRTM* multi-source supply of predictors saw the Sentinel-2 driven C stock prediction model with the longest mean effective range (2000 m) compared to its Landsat-8 (1500 m) counterpart (**Figure 4-3**).

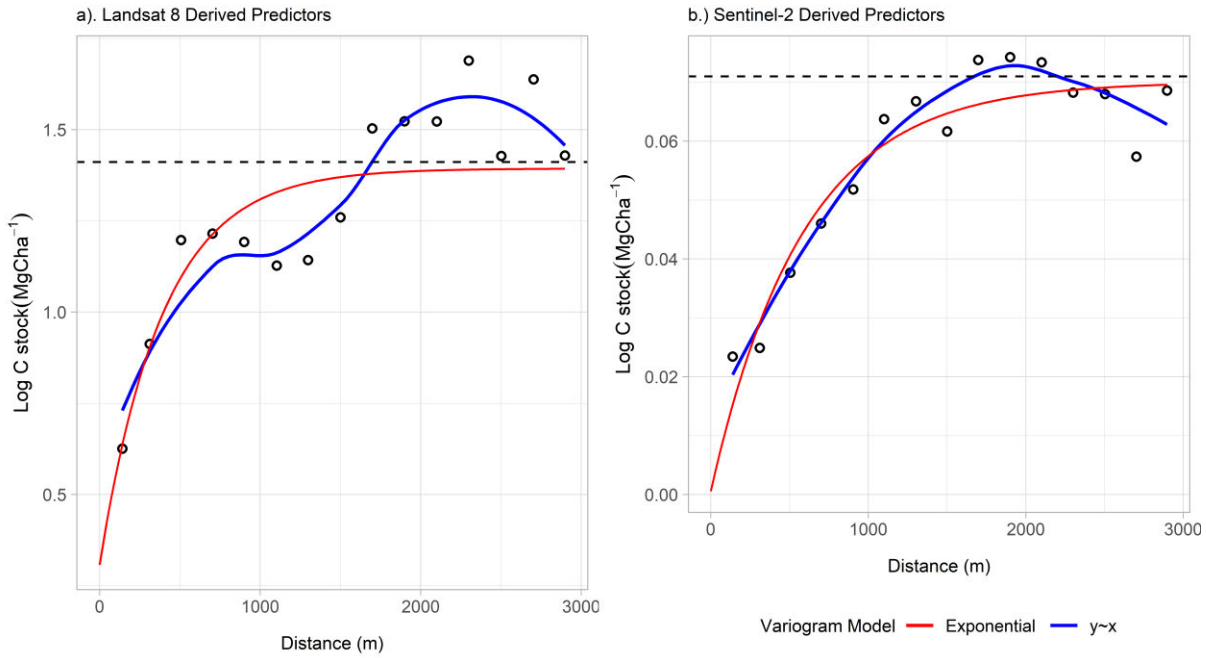


Figure 4-3: Multi-source variogram of residuals for the Landsat-8 and Sentinel-2 driven C stock models. Asymptotes for the theoretical variogram models are illustrated with a black dotted line.

Greater influence of the Sentinel-2 driven C stock model manifests through the spatial covariance structure in which its spatially structured variance (σ_w^2 =0589) is much reduced than that of the Landsat-8 driven C stock model (σ_w^2=1.21). This trajectory also holds true for the micro-scale variance in both models. However, the Landsat-8 C stock-based model has lower effective range (1500 m) when compared to the Sentinel-2 driven carbon stock model due to the vanishing spatial structure in the latter.

4.4.1. The influence of multi-source predictor variables on C stock

A multi-source prediction of C stock using topographic, climatic, anthropogenic and vegetation indices show distance to settlements (*DIST*) and *NDVI* as the only significant predictors (**Table 4-1**) for both the Sentinel-2 and the Landsat-8 carbon stock models. The effect of *NDVI* is much stronger in Sentinel-2 than in Landsat-8 based C stock model, justifying from the

improved spatial resolution of Sentinel-2. The 95% credible intervals of all the topographic (*aspect*, *slope* and *elevation*) and climatic (*MAP* and *MAT*) explanatory variables contain zero and hence, not significant predictors of C stock for both the Landsat-8 and the Sentinel-2 models. However, *SAVI* is a significant predictor in the Sentinel-2 based C stock model whilst its 95% CI (-1.39, 1.32) contains zero in the Landsat-8 based model (**Table 4-1**). Vegetation indices are dominant over other predictor categories like climate and topography in both new generation multispectral remote sensing-based C stock prediction models. The predicted C stock therefore represents averaged pixel values as predictor variables were gridded on a 10 000 m^2 resolution raster surface. A total of 30 000 samples for each grid were obtained from the MCMC sampling from the fitted models' predictive distribution.

Table 4-1: Landsat-8 and Sentinel-2 based C stock model parameter estimates. *DIST* is distance to settlements, *MAT* is mean annual temperature, *MAP* is mean annual rainfall.

Parameter classification	Parameter	Landsat-8 C stock Model	Sentinel-2 C stock Model
		50 % (2.5%, 97.5%)	50 % (2.5%, 97.5%)
Intercept	*Intercept*	-43.8 (16.36, 75.0)	-6.8 (-36.1, 22.3)
Vegetation indices	*NDVI*	2.58 (0.52, 4.76)	4.95 (4.49, 5.44)
	SAVI	-0.07 (-1.39, 1.32)	-0.94 (-1.66, -0.20)
	EVI	-0.49 (-1.59, 0.63)	0.11 (-0.07, 0.30)
Anthropogenic	*DIST*	1.42 (0.75, 2.09)	0.82 (0.55, 1.07)
Topographic	*Elevation*	0.001 (-0.003, 0.005)	0.0004 (-0.0004, -0.0001)
	Slope	-0.069 (1.00, 0.87)	0.08 (-0.27, 0.43)
	Aspect	-0.41 (-0.89, 0.085)	0.05 (-0.04, 0.14)
Climatic	*MAT*	-0.60 (-3.07, 1.84)	-0.12 (-0.62, 0.38)
	MAP	-0.047 (-0.26, 0.16)	-0.02 (-0.07, 0.03)
Spatial covariance	σ_w^2	1.21 (0.71, 1.87)	0.0589 (0.045, 0.075)
	σ_ε^2	0.33 (0.062, 0.65)	0.0024 (0.0007, 0.0076)
	ϕ	0.002 (0.002, 0.003)	0.0015 (0.0014, 0.0021)
	$Effective\ range\ (m)$	1500 (1000, 1500)	2000 (1428, 2143)

There is evidently weaker spatial correlation from the Landsat-8 driven C stock prediction model with a mean effective range of 1500 m and an associated 95% CI of (1000, 1500) m compared to the Sentinel-2 based C stock model showing a much stronger spatial correlation having a mean effective range of 2000 m and an associated (1428, 2143) m 95% CI (**Table 4-1**).

4.4.2. Multi-source-based C stock prediction

Bayesian driven hierarchical prediction of C stock at unsampled sites in the studied area was largely determined by *NDVI* and distance to settlements (*DIST*) as the other climatic and topographic class of predictor variables were not significant (**Table 4-1**).

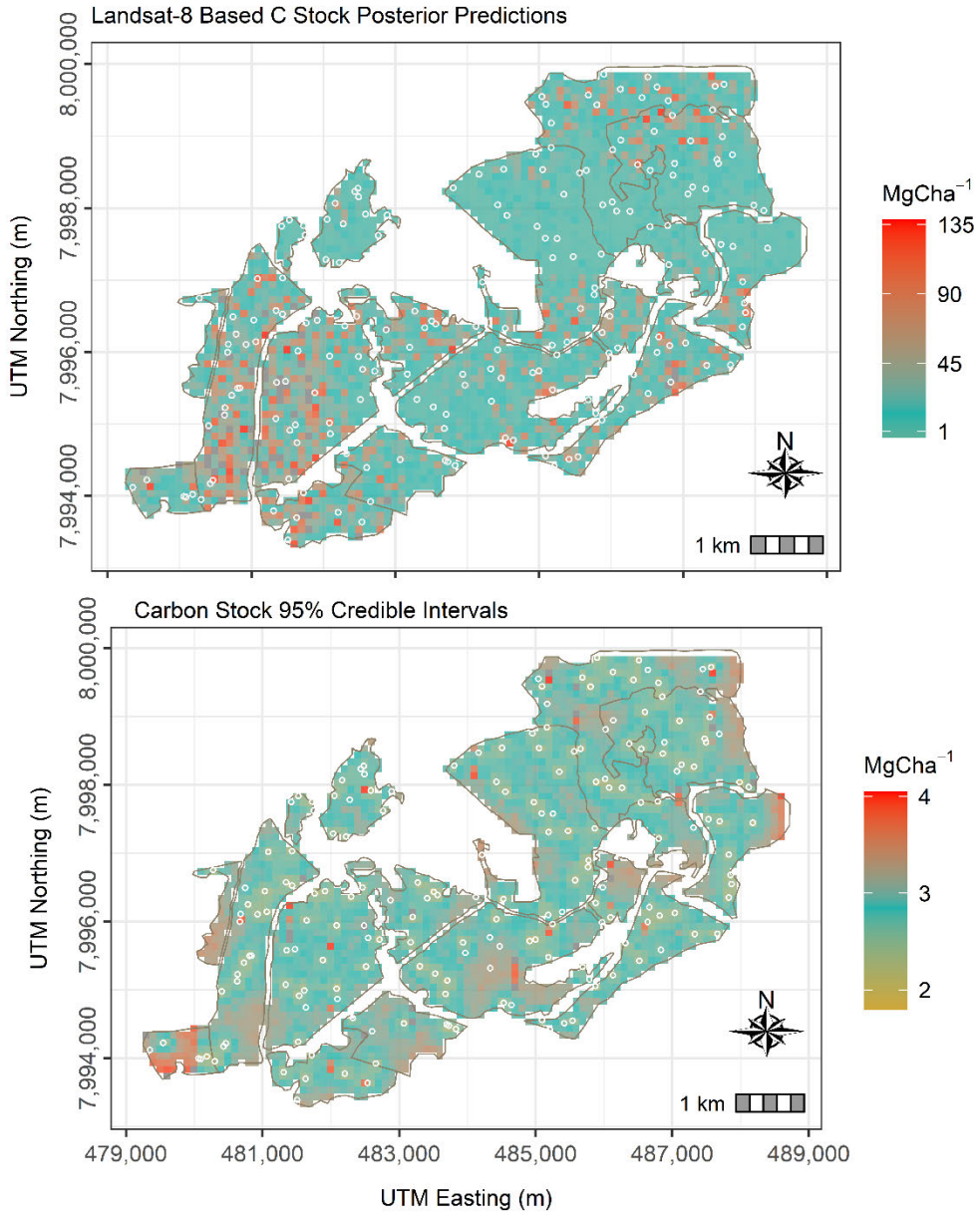


Figure 4-4: Posterior mean and standard deviation of the Landsat-8 based C stock model. Points represent locations of sampled C stock.

4.4.3. Landsat-8 based C stock model

Anthropogenic and vegetation spectral index-based Landsat-8 prediction model gave carbon stock prediction range between 1 MgCha^{-1} and 135 MgCha^{-1} with a 95% CI range of $2 \leq MgCha^{-1} \leq 4$ (**Figure 4-4**). C stock prediction uncertainty is relatively non-uniform across the studied region as can be shown by occurrence of red pixels representing high C stock prediction uncertainty.

There is a notable trend of higher C stock predictions in the western, central and northen parts of the studied region which can be explained by easy accessibility of timber plantations and unlawful logging activities from nearby resettled farmers (**Figure 4-4**).

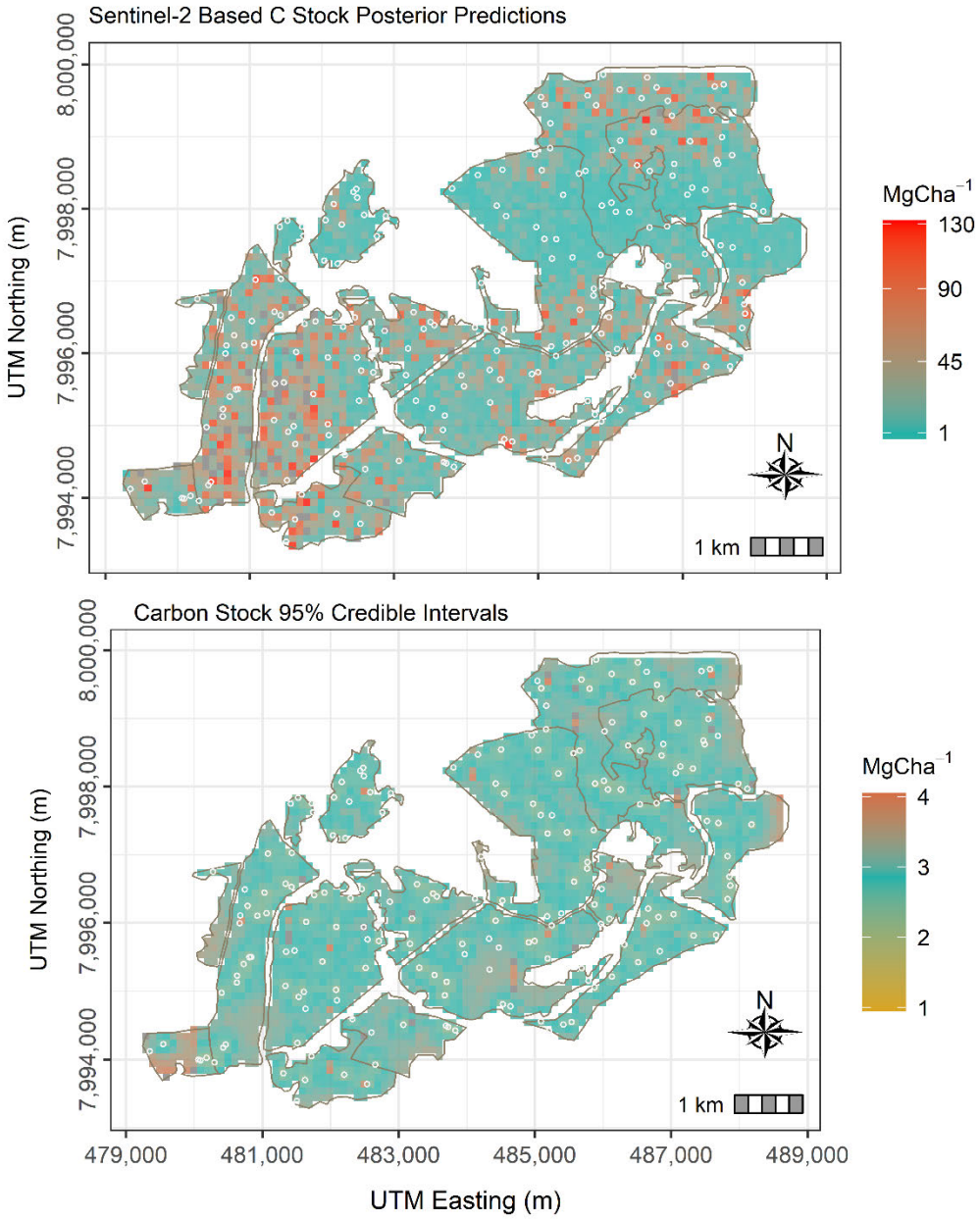


Figure 4-5: Carbon stock posterior mean and standard deviation of the Sentnel-2 based C stock model. Points represent locations of sampled C stock.

4.4.4. Sentinel-2 based C stock predictions

The same trend of higher C stock predictions in the northern, western and central regions of the study area is present in Sentinel-2 based carbon stock prediction model (**Figure 4-5**). In addition to distance to settlements and $NDVI$ as driving predictors of the carbon stock, the Sentinel-2 driven carbon stock model also carries $SAVI$ as a significant predictor, predicting the resource variable with a range between 1 MgCha^{-1} and 130 MgCh^{-1} as illustrated in **Figure 4-5**. However, despite the higher spatial and spectral resolution in Sentinel-2 derived vegetation indices, the Sentinel-2 based C stock model displays higher posterior uncertainty ($1 \leq MgCha^{-1} \leq 4$) compared to that of Landsat-8 (**Figure 4-4**).

4.4.5. C stock model prediction performance assessment

A comparison of the prediction performance of the Sentinel-2 and the Landsat-8 driven C stock models favour the latter as it has lower RMSE (1.17 $MgCha^{-1}$) compared to its Landsat-8 based C stock (2.16 $MgCha^{-1}$) counterpart (**Table 4-2**). The Sentinel-2 C stock model also fits the data well as it displays far much lower Deviance Information Criteria (DIC) (-554.7) than the Landsat-8 based C stock (43.1) model. Different variants of the Bayesian prediction models with different levels of specifications (non-spatial, spatial and error independent models) for both the Landsat-8 and the Sentinel-2 models favour the C stock model with full spatial specifications (**Table 4-2**).

Table 4-2: *C* stock model performance validation statistics

Model Validation criterion	Landsat 8 Based C Stock Model			Sentinel-2 Based C Stock Model		
	Independent Error Model	Spatial Intercept only Model	Spatial Model	Independent Error Model	Spatial Intercept only Model	Spatial Model
RMSE ($MgCha^{-1}$)	3.06	2.08	2.16	1.31	2.89	1.17
MAE ($MgCha^{-1}$)	2.36	1.43	1.55	1.22	1.97	1.13
CRPS ($MgCha^{-1}$)	2.01	1.49	1.52	1.20	1.77	1.14
CVG (%)	91.7	85.4	85.4	95.8	89.6	94.7
DIC	210.2	56.9	43.1	-244.4	37.8	-554.7

The Landsat-8 based C stock model has slightly more under predictions and over predictions and less coverage (93.8%) than the Sentinel-2 based C stock model (97.9%) (**Figure 4-6**). Our explanation of coverage in this investigation resonates with the definition provided by (Guhaniyogi and Banerjee (2019) as the proportion of reliably predicted C stock within the studied region of interest. As such, the Landsat-8 based C stock model's 93.8% coverage provides predictive ability slightly short of the nominal benchmark 95% coverage level whilst the Sentinel-2 C stock model is slightly above (97.9%).

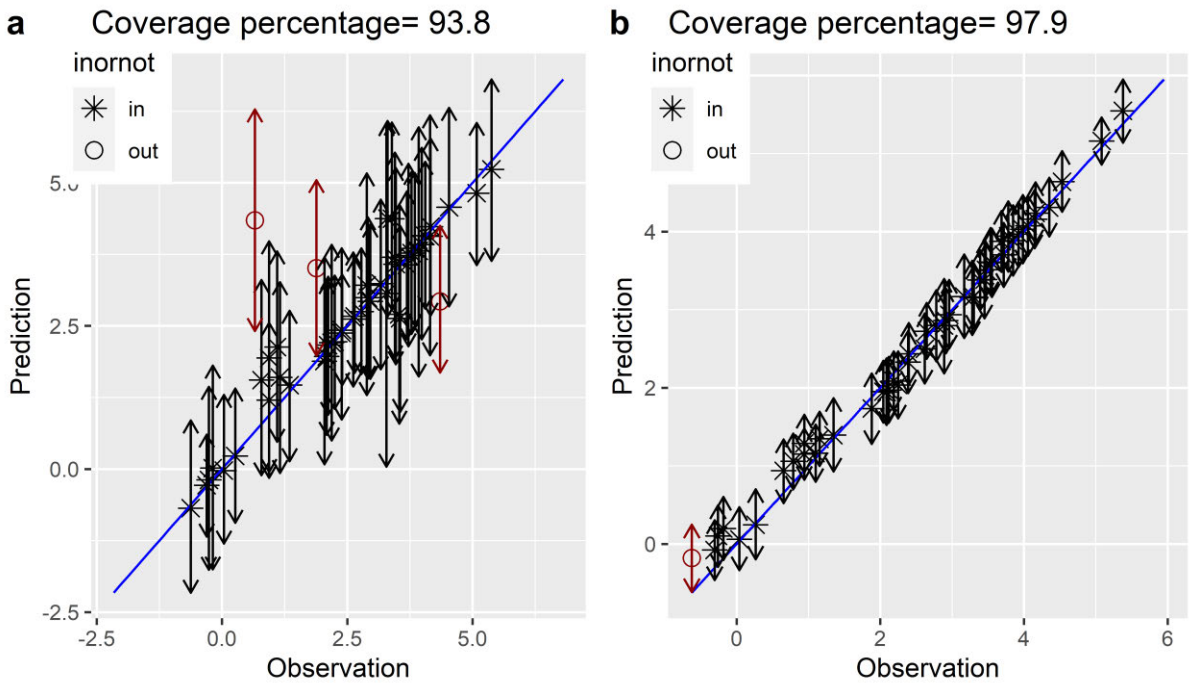


Figure 4-6: Landsat-8-based predicted C stock ($MgCha^{-1}$) against C stock observations ($MgCha^{-1}$) alongside 95% CIs

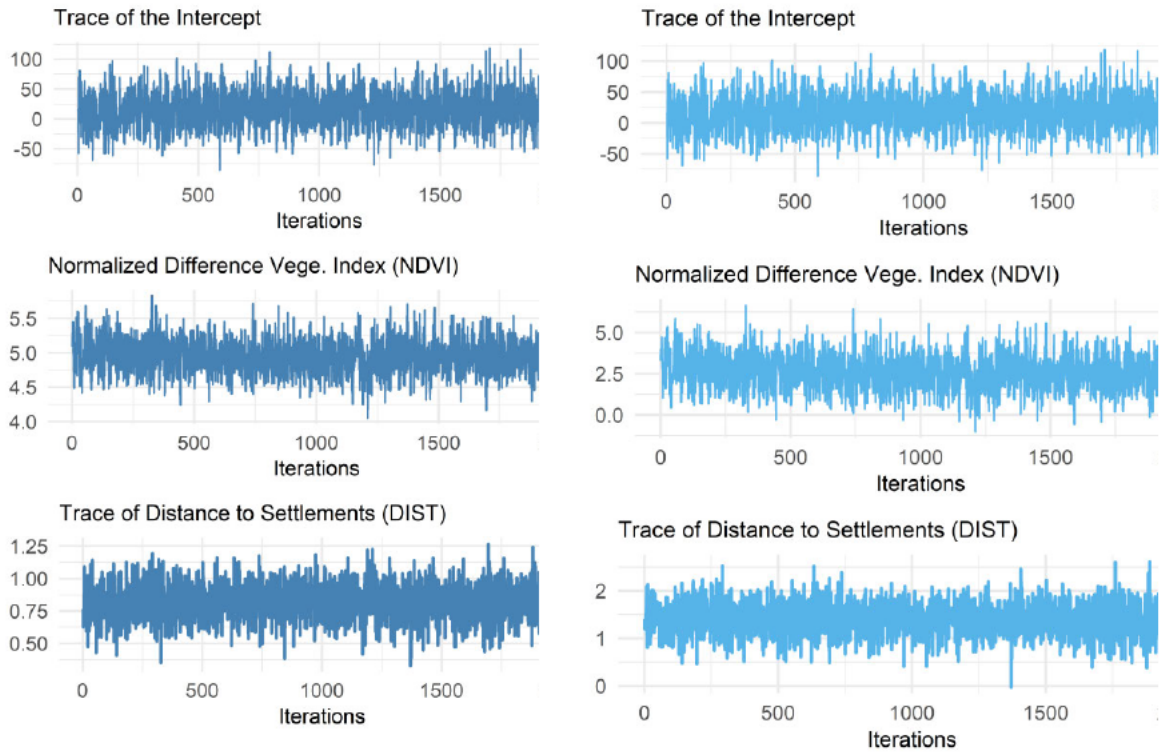
4.4.6. MCMC convergence diagnostic assessment

$NDVI$ and $DIST$ in addition to the model intercept prediction MCMC trace plots are illustrated in **Figure 4-7**. The trace plots are provided as visual aides for assessing MCMC chain convergence and it is evident from **Figure 4-7** that the length N_0, of the burin-in duration allowed Markov chain convergence to the stationery distribution as they were sufficiently large. Presentations shown are as a result of chain thinning for improving the aesthetics of the plots. Thinning speeds up calculated values by decreasing the Monte Carlo sample size and the chains appear to have sufficiently converged to the stationary distribution and mix fairly well after 5, 000 iterations. An assessment of **Figure 4-7** shows a converged sample with MCMC with satisfactory mixing.

4.5. Discussion

The work we present in this research illustrates the significance of coupling different sources of earth observation data in addressing C stock dynamics in managed plantation forest ecosystems. We adopted a multi-source data aproach to the prediction of C stock in a disturbed plantation forest ecosystem using anthropogenic, climatic and topographic derived explanatory variables using Landsat-8 and Sentinel-2 as the mainstream modelling drivers. Literature is replete with AGB and climate change related studies adopting statistical inferencial methods

utilising a combination or stand alone machine learning and remote sensing approaches (González-Vélez *et al.*, 2021; Do *et al.*, 2022; Zhang *et al.*, 2023).



Panel A: Landsat-8 Based Trace plots **Panel B: Sentinel-2 Based Trace plots**

Figure 4-7: Landsat-8 and Sentinel-2 based C stock prediction MCMC trace plots

Sentinel-2 and Landsat-8 are equally comparable new generation remote sensing data sources, but their spatial and spectral properties, coupled with vegetation related climatic and topographic covariates is a good ground for assesing their suitability in C stock assessment for climate change implementation and action plans.

4.5.1. The influence of multi-source predictors on C stock dynamics

The finer spatial resolution of the Sentinel-2 (10 m) satellite data coupled with a supply of other predictors from the WorldClim and the SRTM have stronger explanatory properties than the ones used in Landsat-8 OLI data. Patches of sparsely forested areas are common in the study area due to illegal logging, agricultural activities and overgrazing, which makes *SAVI* significant in Sentinel-2 driven carbon stock model than in the Landsat-8 OLI C stock model (Somvanshi and Kumari, 2020). Modelled topographic variables including elevation are not significant predictors of C stock in the model due to insignificant gradient changes in the study area. Consequently, the spatial variability (Baccini *et al.*, 2004) in the sampled region is not good enough to influence vegetation dynamics. It seems the effect of elevation will be much evident when the study is done on a larger scale (regional rather than local scale) (Su *et al.*,

2016). Both climatic variables, that is, MAP and MAT are not insignificant predictors of C stock in this study. This could be an indication of small to insignificant variability of the factors in the studied area due to its localised spatial extent *vis* regional scale studies done by Simard *et al.* (2019) and Jiang *et al.* (2021). Temperature and rainfall usually vary at landscape rather than at local scales (Zhang *et al.*, 2023). Resettled farmers from the agrarian land reform most likely exploit forest resources closer to their settlements and which can easily be accessed due to closeness to roads and grazing facilities. As the results of C stock predictions show, it is within the same areas of higher C stock predictions that there is higher posterior uncertainty of C stock as a result of the plantation forest possibly responding to multi-sources of environmental variability.

As provided by Isaaks and Srivastava (1989), geostatistical predictions are plausible as prediction uncertainty is lower near sampled C stock observations. On the other hand, the significance of $NDVI$ as an index correlated with forest biophysical attributes, amongst them, chlorophyll and leaf area index (LAI) is much established (Baugh and Groeneveld, 2006). Higher C stock prediction uncertainty associated with the Sentinel-2 driven model indicates modelling inadequacy in terms of the full complement of auxiliary variables needed for capturing C stock variability in the study area. This is because the advanced spectral and spatial resolution of Sentinel-2 derived data suggest that Sentinel-2 driven models perform better than the Landsat-8 models (Semela, Ramoelo and Adelabu, 2020). The fact that all the other topographic and climatic variables are irrelevant in the prediction of C stock in either the Landsat-8 or the Sentinel-2 based C stock models imply $DIST$ and $NDVI$ are the major but inadequate sources of spatial variability for carbon stock prediction in the region. However, prediction uncertainty of Sentinel-2 based C stock model looks uniform and less variable across the study domain compared to that in the Landsat-8 based C stock prediction model. Better spectral properties in Sentinel-2 (13 spectral bands) compared to the Landsat-8 (11 spectral bands) seem to justify the quality of C stock predictions in both models (Jia *et al.*, 2014; Azevedo, 2021).

4.5.2. Multi-source C stock prediction modelling

In spite of coupling different data sources for C stock prediction in a plantation forest ecosystem set-up, vegetation spectral indices and anthropogenic ($DIST$) covariables emerge as significant predictors of C stock. Bordoloi *et al.* (2022) and Somvanshi and Kumari (2020) showed $NDVI$, $SAVI$ and EVI as significant regressors of AGB in a Landsat-8 and Sentinel-2 based remote sensing framework. However, studies coupling new generation remote sensing

information in a multi-source data approach using Bayesian hierarchical modelling are not common. Furthermore, studies that have implemented a multi-source data approach using other aproaches like machine learning (SVM, boosted regression trees and multiple regression splines) have established lower prediction accuracies (8 -13 $Mgha^{-1}$) than the present study (between 1.31 $MgCha^{-1}$ and 1.17 $MgCha^{-1}$) (Shi and Liu, 2017; Serrano, Corral-Rivas and López-Sánchez, 2019). Prediction accuracy differences between similar studies in literature and our study can be justified by the unmatched flexibility of the Bayesian inferential approach in partitioning sources of error and providing access to the entire posterior predictive distribution, which other inferencial techniques cannot do (Goulard and Voltz, 1992; Beloconi and Vounatsou, 2020). The present study went a step further by comparing the performance of C stock models constructed using different multispectral remote sensing platforms within a Bayesian framework and established Sentinel-2 as the best performing C stock model.

A study by Chinembiri et al., (2023) using a combination of anthropogenic and vegetation spectral indices within a Bayesian hierarchical framework and Landsat-8 and Sentinel-2 as the main data sources gave prediction accuracies with RMSE of 1.16 $MgCha^{-1}$ and 2.69 $MgCha^{-1}$ respectively. This study did not comprehensivley deal with all the possible AGB sources of variation including climatic variables. However, adoption of a multi-source data approach using the same multispectral remote sensing data sources in the present study resulted in prediction accuracies with RMSE of 1.17 $MgCha^{-1}$ and 2.16 $MgCha^{-1}$ for the Landsat-8 and the Sentinel-2 based C stock models respectively. It is therefore evident that Sentinel-2 remote sensing is a promising and valuable AGB prediction remote sensing ancillary data source as it provides relatively lower prediction uncertainty (2.16 $MgCha^{-1}$) than its Landsat-8 (1.16 $MgCha^{-1}$) counterpart. The importance of multi-source data coupling within a Bayesian approach for C stock modelling in climate change studies is notable and has the potential for supplying the much needed answers to climate change implementation and mitigation strategies.

4.5.2.1. Climatic variables and C stock prediction

Contrary to global climate change studies, neither MAP nor MAT are significant covariates for C stock prediction in the present study. The influence of climate on C storage and distribution in global forests is well established (Pan *et al.*, 2013). The constraints that precipitation and temperature exert on biomass production clearly link them to biomass storage (Keith, Mackey and Lindenmayer, 2009; Pan *et al.*, 2013). Temperature controls carbon dioxide assimilation rates in the leaves (Lloyd and Farquhar, 1996) and C losses from respiration related to living tissue maintenance (Larjavaara and Muller-Landau, 2012). An inverse relationship between

elevation and AGB is well established, for example, AGB is high at low elevations since these regions have high temperature, thick air and weak ultraviolet rays that promote vegetative growth of plants (Ahmed, Atzberger and Zewdie, 2022). On the other hand, changes in aspect and slope effect on forest biomass is associated with changes in soil nutrients and climatic conditions in small areas (Jucker *et al.*, 2018). Again, areas with gentle slopes have excellent nutrition and soil moisture conditions making the AGB in these environments significantly higher than in poor and steep soils (He *et al.*, 2022).

It appears there is less spatial variability of MAP as one moves from one geo-location to the other within the studied region (Gordon *et al.*, 2018). However, the spatio-temporal influence of MAP also need to be investigated as it plays a pivotal role in AGB dynamics (Simard *et al.*, 2019). This is the same explanation with MAT. This variable does not change dramatically over short distances, but over bigger distances. The maximum distance between the farthest sampling points within the study area is approximately 2500 m, making it inconceivable to expect significant temperature or precipitation changes (Chinembiri, Mutanga and Dube, 2023). However, elevation can dramatically change over short distances, depending on the topography of the area, but it appears the studied area has a generally uniform landscape not complicated enough to influence C stock dynamics of the region.

The multi-source range of predicted C stock values for plantation forest with *Pinus patula*, *Eucalyptus grandis* and *Eucalyptus camaldulensis* as the dominant species for the current study is $1 \leq MgCha^{-1} \leq 135$ for the Landsat-8 based C stock model and $1 \leq MgCha^{-1} \leq 130$ for the Sentinel-2 based C stock prediction model. An assessment of AGB distribution of a Vietnamese mangrove forest using a combination of Artificial Neural Network (ANN) and remote sensing data resulted in a mangrove AGB prediction range of $(6.53 - 368.2)$ Mgha^{-1} in 2000 and from $(13.75 - 320.3)$ Mgha^{-1} in 2020, respectively (Do *et al.*, 2022). AGB predictions established by Fararoda *et al.* (2021) using AdaBoost, random decision forest, multilayer neural networks and Bayesian ridge regression machine learning methods recommended AdaBoost and random forest as the best performing methods. The type of remote sensing platform together with the method of prediction employed have significant sway over AGB model performance (Ahmed, Atzberger and Zewdie, 2022). For instance, Fassnacht *et al.* (2014) assessed and compared hyperspectral and LiDAR data for AGB modelling and noted that Random forest models fused with LiDAR data offers the best AGB model prediction performance. The best model predictions are derived using a combination of the modelling methodology and a near complete coverage of all possible sources of spatial variability of the modelled variable. We endeavoured to meet these two conditions by

employing the novel Bayesian hierarchical approach with a multi-source supply of predictor variables. However, the limitation in our endeavours appear to be the complex spatio-temporal dynamics between AGB and MAP and MAT coupled with scale, as the studied region is rather small to trigger meaningful C stock variability (Baccini *et al.*, 2004; Ahmed, Atzberger and Zewdie, 2022).

Prediction accuracies in our study seem superior than accuracies established in similar studies and environments, implying the results of this study can be applied in comprehensive and targeted ecosystem intervention (Takagi *et al.*, 2015; Jiang *et al.*, 2021; Ahmed, Atzberger and Zewdie, 2022; Fan, Wang and Yang, 2022; Xiong and Wang, 2022). Improvement in the spatial and spectral properties of Sentinel-2 as realised from the model's Credible Interval width (CIWs) makes it the best data source for forest resource monitoring within the United Nations Framework Convention on Climate Change (UNFCCC). This implies that conservation and monitoring of extensive plantation forest ecosystems can be achieved with higher precision and accuracy and consequently, improve forest carbon accounting.

4.6. Conclusions

We undertook to predict C stock in a managed and disturbed plantation forest ecosystem in the eastern highlands of Zimbabwe using a multi-source data approach coupling forest inventory data with earth observation derived information in a Bayesian geostatistical hierarchical framework. Forest productivity related multi-data sources based on anthropogenic, climatic and topographic factors were utilised as predictors of C stock, with Landsat-8 and Sentinel-2 as the mainstream model building blocks. Unlike in other studies that have been done at landscape scale, both climatic and topographic variables were not significant predictors of C stock. An $NDVI$ and $DIST$ driven Sentinel-2 C stock model with a prediction RMSE of 1.17 $MgCha^{-1}$ outperformed its Landsat-8 model counterpart with a 2.16 $MgCha^{-1}$ RMSE. A multi-source data approach of C stock modelling at a local scale does not improve prediction uncertainty in managed forest ecosystems, but the Sentinel-2 constructed C stock model provides lower prediction uncertainty than its Landsat-8 counterpart.

Scale and multi-source data variables are important determinants of C stock prediction needed for climate change implementation plans and strategies. A more practical model for use by forest practitioners and subsequent adoption by central governments is the one based on Sentinel-2 based data with $NDVI$ and $DIST$ as drivers of variability. The need for accurate accounting of C stock density and assessment of C sequestration potential in both natural and managed forest ecosystems by national governments under the UNFCCC requires construction

of credible models. The dynamics between the ability of communities to access forestry resources and services and the rate of C sequestration is critical for site specific interventions needed for ecological monitoring and ecosystem restoration. The Bayesian approach together with free and analysis ready Sentinel-2 data is particularly handy as it nearly eliminates the need for extensive forest inventory surveys that may be needed at frequent intervals for in C accounting for climate change strategies and mitigation plans.

4.7. Summary

This chapter employed a Bayesian hierarchical geostatistical approach using the Sentinel-2 and the Landsat-8 derived vegetation indices in addition to climatic, topographic and anthropogenic factors in the prediction of C stock in a disturbed plantation forest ecosystem. There are mixed results from the literature regarding the effect of both climatic and topographic variables on the spatial distribution of carbon stock at the local scale. Some studies done at the local and landscape scales with climatic and topographic data as ancillary variables established MAT and MAP as significant predictors of C stock. Both climatic and topographic ancillary variables utilised in this study have no influence on the spatial distribution of C stock in disturbed plantation forest ecosystems. Even within a Bayesian driven multi-source data modelling framework, the Sentinel-2 derived $NDVI$ and $DIST$ variables are still predominant C stock predictors. High quality and readily available ancillary data from the Sentinel-2 satellite sensor does not provide a full and easy to use suite of C stock prediction methods for forest practitioners and other decision makers. The Bayesian hierarchical geostatistical approach is computationally challenging and conceptually more involving. However, easy to use frequentist geostatistical methods that provide C stock predictions with associated measures of uncertainty are also available. Thus, the next chapter evaluates how C stock prediction uncertainty obtained using the Bayesian geostatistical inferential approach compares with those derived from a frequentist geostatistical methodology.

CHAPTER FIVE

remote sensing

MDPI

Article

Carbon Stock Prediction in Managed Forest Ecosystems Using Bayesian and Frequentist Geostatistical Techniques and New Generation Remote Sensing Metrics

Tsikai Solomon Chinembiri [1,*], Onisimo Mutanga [1] and Timothy Dube [2]

[1] College of Agricultural, School of Agricultural Earth and Environmental Sciences, University of KwaZulu-Natal, Private Bag X01, Pietermaritzburg 3209, South Africa

[2] Institute of Water Studies, Department of Earth Sciences, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa

* Correspondence: chinembiri24500@alumni.itc.nl

Abstract

The study compares the performance of a hierarchical Bayesian geostatistical methodology with a frequentist, specifically, Kriging with External Drift (KED), geostatistical approach for predicting C stock using prediction aides from the Landsat-8 and Sentinel-2 multispectral remote sensing platforms. The frequentist geostatistical approach's reliance on long run frequency of repeated experiments for constructing confidence intervals is not always practical or feasible, as practitioners typically have access to a single dataset due to cost constraints on surveys and sampling. We evaluated the two approaches for C stock prediction using two new generation multispectral remote sensing data because of the inherent uncertainty characterizing spatial prediction problems at unsampled locations, as well as differences in how the Bayesian and frequentist geostatistical paradigms handle uncertainty. Information on C stock spectral prediction in the form of $NDVI$, $SAVI$, and EVI derived from multispectral remote sensing platforms Landsat-8 and Sentinel-2 was used to build Bayesian and frequentist-based C stock predictive models in the sampled plantation forest ecosystem. Sentinel-2-based C stock predictive models outperform their Landsat-8 counterparts in both the Bayesian and frequentist inference approaches. However, the Bayesian based Sentinel-2 C stock predictive model ($RMSE = 0.17\ MgCha^{-1}$) is more accurate than its frequentist-based Sentinel-2 ($RMSE$=1.19 $MgCha^{-1}$) C stock equivalent. The Sentinel-2 frequentist-based C stock predictive model gave C stock prediction range of $1 \leq MgCha^{-1} \leq 290$ whilst the Sentinel-2 Bayesian based C stock predictive model resulted in the prediction range of $1 \leq MgCha^{-1} \leq 285$. However, both Bayesian and frequentist C stock predictive models built with the Landsat-8 sensor overpredicted the sampled C stock because the range of predicted values fell outside the range of observed C stock values. As a result, we recommend and conclude that the Bayesian-based C stock prediction method, when combined with high-quality remote sensing data such as Sentinel-2, is an effective inferential statistical methodology for reporting C stock in managed plantation forest ecosystems.

Keywords: Bayesian methodology, classical geostatistics, multispectral remote sensing, carbon stock, plantation forest, managed ecosystem

5.1. Introduction

Plantation and natural forests serve as one of the main economic pillars in sub-Saharan Africa because they support both economic growth and human livelihoods. Due to the vulnerabilities caused by climate change, economies in south of the Saharan desert and other regions of the world must balance the need to protect the environment against economic pressures brought on by population growth and poverty (Araki, Yamamoto and Kondo, 2015). Nearly 50% of Africans live without access to electricity, and at least 60% of them still rely on wood for cooking and heating (Shoko and Gara, 2021). This disproportionate reliance on climate-vulnerable industries like energy and agriculture for economic survival and expansion leads to significant increases in the supply of primary energy and greenhouse gas pollutants (Gibbs *et al.*, 2007). Forest biomass is regarded as a crucial part of monitoring forest resources according to the Food and Agriculture Organization's (FAO) 1994 International Forest Resources Monitoring Program (Van Amstel, 2006). Therefore, accurate monitoring and estimation of forest aboveground biomass (AGB) at local and regional scales are essential for understanding how forest AGB contributes to the regional and global carbon cycle (European-Commission, 2017).

Zimbabwe is one of several African countries that, in their Nationally Determined Contributions (NDCs) to the Paris Agreement, have made bold proposals for establishing low-carbon and climate-resilient economies (Gibbs *et al.*, 2007). Yet Agriculture, Forestry and Other Land Use (AFOLU) still remain the biggest contributors of Greenhouse Gas Emissions (GHG), accounting for 54% of GHG in 2017 in Zimbabwe (Government of Zimbabwe, 2021). Deforestation resulting from agriculture expansion, increased stocking levels, fetching of fuelwood, veld fires, harvesting of timber for construction, mining, illegal settlements, tobacco curing, charcoal making and commercial logging are some of the major drivers of GHG emissions in the AFOLU sector (Brown, 1997). The Zimbabwe government also introduced the Greenhouse Gas Abatement Cost Model (GACMO) for establishing a GHG database (Initiative For Climate Action Transparency (ICAT), 2022). The model can also be used as a tool for reporting, monitoring and verification for guaranteeing transparency in mitigation actions for climate change.

Commonly used methods for estimating aboveground biomass (AGB) include mean biomass density, allometric equations, remote sensing, forest identity, geostatistics and biomass expansion factors (Shi and Liu, 2017). The method of statistical inquiry for any of these AGB estimation and prediction methods can be comprehensively categorised as either Bayesian or frequentist, depending on the circumstances of the investigation and assumptions underlying

the inference. Notable differences between the Bayesian and the frequentist statistical methodologies regard the nature of the unknown parameters under investigation (Gelman *et al.*, 2004). The frequentist paradigm treats parameters of interest as unknown and fixed whilst the Bayesian framework regards all unknown parameters as uncertain and therefore, should be characterised by a probability distribution (Cameletti and Biondi, 2019).

The performance of Maximum Likelihood (ML), Least Squares and Bayesian approaches in Bogota, Columbia, was tested by Ghosh & Carriazo (2007) in a hedonic estimation context and concluded that none of the aforementioned approaches are better than the other. However, because of the philosophical differences governing the Bayesian and frequentist statistical techniques, the authors recommend the choice of the estimation technique to be grounded on the peculiarities of the policy challenges at hand (Ghosh and Carriazo, 2007). Some studies in literature dwell at one of the two approaches as the principal methodology of inference. Notable work in the realm of biomass estimation using the Bayesian techniques include (Finley and Banerjee, 2008; T.S. Chinembiri, Mutanga and Dube, 2023). Recent studies assessing AGB distribution established accuracies of 17.52 Mg/ha (Babcock *et al.*, 2015) and 1.16 $MgCha^{-1}$ and 2.69 $MgCha^{-1}$ for Sentinel-2 based and Landsat-8 based Carbon (C) stock predictive models, respectively (T.S. Chinembiri, Mutanga and Dube, 2023). Other remote sensing and machine learning-based efforts towards the estimation and prediction of AGB in rcent times include Do et al. (2022) who established mangrove AGB predictions in Vietnam to range from 6.51 to 368 $Mgha^{-1}$ and from 13.70 to 320.1 $Mgha^{-1}$ for remote sensing and Artificial Neural Networks, respectively.

Addressing a geostatistical research question from the Bayesian view point makes it possible to provide definitions of spatial predictors contributing to uncertainty in the unknown spatial covariance structure (Hudson and Wackernagel, 1994). Kriging provides an optimal geostatistical technique under the frequentist paradigm that is employed in the description of geographic patterns and predicting values of a variable at unobserved locations and, consequently, evaluate the error associated with the predicted values (Wheeler and Waller, 2009).

Research making use of the frequentist geostatistical approach in isolation to the Bayesian technique for the estimation of C stock are also well documented in literature. Sales *et al.* (2007) mapped AGB in Brazilian Amazon using Kriging with External Drift and established prediction accuracies for different sample sizes ranging from 0 to 110 for distances within 300 km radii from the prediction locations. The lowest RMSE for the estimated AGB for a 110-

sample size was 32.8 Mgha^{-1} whilst the lowest accuracy for the lowest sample size of $n > 0$ was 48.06 Mgha^{-1}. To add on, Jiang *et al.* (2022) predicted AGB in Wangyedia forest farm in China using Landsat-8 and the newly launched Landsat-9 and arrived at RMSEs of 16.83 tha^{-1} and 17.91 tha^{-1}, respectively. Wai, Su and Li (2022) coupled remotely sensed derived Sentinel-2 explanatory variables with geostatistics and machine learning algorithms in Mayanmar for predicting above ground biomass and established accuracies of 24.91 Mgha^{-1} and 34.72 Mgha^{-1} for the Random Forest based ordinary kriging and the Random Forest based co-kriging, respectively. As demonstrated by the results of Jiang et al. (2022), Landsat-8 built AGB estimation models can still be superior than AGB models built from the successor Landsat-9 sensor, despite the relative spectral and radiometric improvements in the latest Landsat-9 mission.

It is worthwhile to explore avenues of improving AGB prediction and estimation accuracy through the adoption of befitting statistical methods of inference for production of better-quality reports applicable to climate change mitigation and climate change action. A significantly higher proportion of studies assessing the predictive performance of Landsat-8 and Sentinel-2 in above ground biomass estimation favour Sentinel-2 over Landsat-8, though the differences in prediction performance are insignificant (Korhonen *et al.*, 2017; Jiang *et al.*, 2022). On the other hand, studies that have employed the Bayesian spatial hierarchical technique coupled with remote sensing derived ancillary data have always outperformed similar studies done using the frequentist approach for predicting AGB (Finley, Banerjee and McRoberts, 2009; Babcock *et al.*, 2015; Astola *et al.*, 2019).

Under the frequentist viewpoint, the interpretation of a confidence interval (CI) is hypothesised on Neyman understanding where the CI gives a measure of uncertainty by taking into account the long run frequency of replicated experiments (Neyman, 1937). This suggests that if a practitioner/ forester gathers 500 datasets on tree *dbh* from independent trials for estimating the parameter of a C stock prediction model and constructs a, say 97%, confidence interval for the parameter estimate for each dataset, at least 97 of the CIs would be expected to contain the true (but fixed) unknown model parameter (Neyman, 1937; Babcock *et al.*, 2018). However, in most practical settings, foresters or practitioners would not have access to multiple datasets and usually have a single dataset as it is rather costly and unfeasible to undertake multiple experiments. The constructed CI may or may not contain the true (but fixed) unknown model parameter. On the other hand, the Bayesian credibility interval is a more pragmatic proposition as the credible interval is built in a manner that guarantees that there is a certain probability associated with getting the true (but random) unknown model parameter (Murphy, 2012;

Hazra, 2017). By the same reasoning, if a practitioner estimates the C stock predictive model using a single tree *dbh* dataset and constructs a, say 97%, credible interval, there would be a 97% probability that the true (but random) unknown model parameter is contained within that credible interval.

Since uncertainty is inherent in spatial prediction problems at unsampled locations, the Bayesian approach handles it better than the frequentist approach because it benefits from having access to the full posterior predictive distribution of the modelled variable (Mansfield and Helms, 1982; Hazra, 2017). The principled way in which Bayesian inference incorporates pre-experimental information in the form of priors and experimental data, combined with the real-world benefits of parameters in carbon accounting and climate change action, makes it worthwhile to investigate how the two inferential paradigms compare when using freely available and new generation remote sensing data. As a result, the current study is an extension of current earth observation-based inferential techniques for C stock accounting in climate change adaptation and mitigation under the United Nations Framework Convention on Climate Change (UNFCCC). As previously hypothesized, we set out to determine whether the Bayesian inferential approach can handle the uncertainty inherent in spatial prediction phenomena better than its frequentist counterpart.

5.2. Methods

This section gives a detailed description of the study area and the methods employed in the sampling and analysis of the measured variables.

5.2.1. Study area

We undertook the study at Lot 75A of Nyanga Downs in Manicaland province of the eastern highlands of Zimbabwe. The area of interest is in Nyanga district of the aforementioned province and has dominant tree species comprised of *Pinus patula*, *Eucalyptus grandis* and *Eucalyptus camaldulensis*. Dotted patches of the area lying between latitude 32°40’ E and 32°54’ E and 18°10’12’’S and 18°25’4’’S longitude as illustrated in **Figure 5-1** have undergone changes in land use and are currently under gold panning, agriculture and grazing (Forestry-Commission, 2021).

Changes in land use to grazing and agriculture came after some parts of the commercially owned plantation forests were redistributed by the government of Zimbabwe to communal farmers in 2000. This development removed access barriers that were in place within the commercial plantation forests before the land redistribution, giving settlers more access to forest resources (Zvobgo and Tsoka, 2021). The area of study which covers an approximate

area of 2 767 ha, receives variable amounts of rainfall ranging from 741 mm to 2 997 mm and a mean annual rainfall of 1 250 mm (Whitlow, 1998). Annul mean temperatures are also varied, with a minimum annual range of 9ºC to 12°C and a maximum range of 25ºC to 28°C. Extensive wild fires occur at high altitudes due to the hot weather experienced during the summer season of the year, between the months of August and November (FAO, 2003). Rapid population growth (**Figure 5-1 (b)**) within the jurisdiction of the studied region is also a possible factor driving rapid land use change, leading to veld fires and logging as actions for opening up more land for agriculture.

5.2.2. Remote sensing covariates

The relatively low cost of acquiring Landsat and other freely available earth observation sensors in addition to their spatial coverage makes them indispensable for natural resource modelling. The recently launched Sentinel-2 satellite sensor is made up of six bands which can be compared to the Landsat-8 bands and also carries three additional bands comprising the red-edge (RE) spectrum (Drusch *et al.*, 2012). The Red edge bands are positioned at 704, 740 and 782 nm whose band width are 15, 15 and 20 nm, respectively. The red-edge forms the major spectral feature of vegetation located between the high reflectance in the NIR (750 nm) and the red absorption maximum (680 nm) (Frampton *et al.*, 2013). Sentinel-2's surface bands have 10 m and 20 m spatial resolution as compared to Landsat-8's 30 m bands. These differences in sensor configuration and properties form the basis for assessing their mapping accuracies.

5.2.3. Landsat OLI and Sentinel-2 MSI imagery

We obtained Landsat-8 images from the United States Geological Survey Earth Explorer site as georeferenced and analysis ready data (ARD) (http://earthexplorer.usgs.gov). We filtered the datasets for cloud cover and set cloud shadow thresholds to below 10%. We downloaded Sentinel-2 cloud free images on the 20th of September 2020, coinciding with the time that we acquired the Landsat-8 data which covered the whole area of interest, including Lot 75A of Nyanga Downs in the eastern highlands of Zimbabwe. Sentinel-2 imagery with 13 spectral bands is acquired as level-1C 12-bit fixed Top of the Atmosphere (TOA) reflectance values. We carried out the orthorectification and pre-processing of the Sentinel-2 level 1-C data using the *sen2r* package of the R Statistical and Computing Environment (Ranghetti *et al.*, 2020).

We derived the Enhanced Vegetation Index (EVI), Soil Adjusted Vegetation Index ($SAVI$) and Normalised Difference Vegetation Index ($NDVI$) from each of the two sensors as independent variables for use in C stock prediction in a managed plantation forest in Zimbabwe. Frampton

*et al. (*2013) *and* Ranghetti *et al.* (2020) have employed the aforementioned vegetation indices as independent variables in AGB estimation and assessment. We therefore employ these variables using philosophically different statistical research methodologies in order to assess the best framework that can be applied in C accounting and reporting for climate change studies.

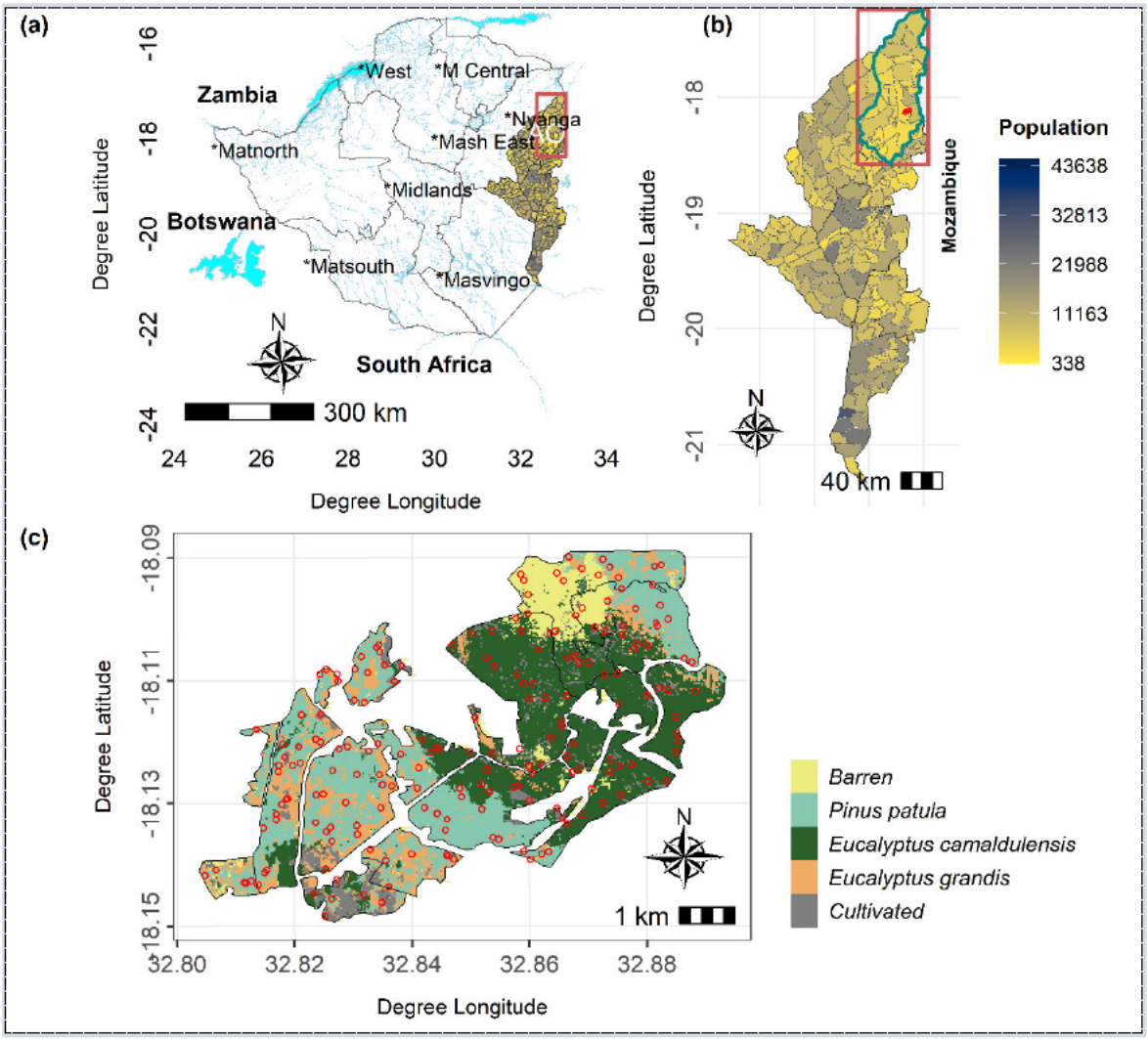


Figure 5-1: Study are map showing (a). location of the province where samples for the study were collected, (b). study area location within the sampled province and (c) plantation forest species distribution in the Area of Interest (AOI) with Points representing locations of sampled C stock.

5.3. Sampling design

The details of the sampling design utilised in the study are provided in the following section.

5.3.1. Spatial coverage sampling and mapping of regionalized variables

We carried out the sampling programme in an area that had not been sampled before and hence, scales of spatial variability were not known beforehand. Under such circumstances, the Mean Squared Shortest Distance (MSSD) is a befitting objective function which we utilized in order to optimize sample locations. We utilized the *k-means* clustering algorithm for uniform area

coverage sampling. According to Walvoort, Brus, and de Gruijter (2010), even distribution of sampling locations within a study domain can enhance the mapping and estimation of regionalized variables. This assertion is further confirmed by Brus, de Gruijter, and van Groenigen (2006) who demonstrated how even coverage of the study area with sampling observations can be utilized for the dual role of estimating spatial means of regionalized variables and resolving mapping in forestry, soil and environmental research. The MSSD remains a dominant methodology for optimizing sampling pattern over other methods like the Spatial Simulated Annealing (SSA) as both a prediction and estimation design for regionalized variables. The suitability of this design for locations where sampling schemes cannot be extended beyond a single phase is well documented.

We subdivided the study domain (D) into compact subunits through clustering of the building blocks making up the sampling domain using the *k-means* optimization procedure (Li and Li, 2019; Bordoloi *et al.*, 2022). The x and y coordinates of the central points of the building blocks are the classification variables for the *k-means* optimization function. We utilised the centroids of the clusters as sample locations where the sampling plots for C stock were set-up.

5.4. Carbon stock data

Measured tree Diameter at Breast Height (DBH) from the field sampling programme were used to calculate per tree biomass and the final derivation of C stock.

5.4.1. Above ground tree biomass (AGTB) field measurement

Measurements of all trees with at least 10 cm diameter at breast height (*DBH*) (at 1.3 m) were taken using 500 m^2 circular plots from the 19th of September to the 24th of October 2021. Diameter and linear tapes were used for the tree measurements and trees with dbh less than 10 cm were excluded as they are generally regarded to have insignificant C stock (Gibbs *et al.*, 2007). As the average slope within the study area was generally less than 30%, we did not consider slope correction for the measured outcome variable (Ravindranath and Ostwald, 2008). We carried out optimization of the resulting 200 sampling points using the *spcosa*-package implemented using the R Statistical and Computing Environment (Li and Li, 2019; Bordoloi *et al.*, 2022). We pre-uploaded the 200 probable sampling observations into a 72 H handheld Garmin GPS before setting out for the field work programme. Actual sampling points of forest biomass obtained during the field exercise were 191 as nine of the pre-loaded sampling locations fell outside the boundaries of the defined study domain (**Figure 5-1**).

5.4.2. Biomass calculation and derivation of C stock

Allometric equations used by Brown (Brown, 1997) were applied in the calculation of AGB *Pinus* species whilst AGB of the *Eucalyptus* species was calculated using allometric equations developed by (Zunguze, 2012). The same allometric equations used for the *Pinus* and *Eucalyptus* species were also applied for the *Eucalyptus* and *Pinus* species of Manica province in Mozambique whose climatic and weather conditions largely resemble those of the studied region in the eastern highlands of Zimbabwe. We then converted the AGB of every individual tree to C stocks per species through the conversion factor by Van Amstel (2006). Estimated per plot AGB values were then expanded to a standardised unit area of a hectare, measured in $MgCha^{-1}$.

Allometric equations shown in Equations (5-1) and (5-2) were used for the calculation of Above Ground Biomass (AGB) for the *Pinus patula and the Eucalyptus grandis* and *Eucalyptus camaldulensis* species, respectively (Matose, 2008). A default conversion factor of 0.47 used by the IPCC was applied to derive AGB to C stock.

$$tDw = e^{(-1.170+2.119*ln(dbh))} \quad (5\text{-}1)$$

$$tDw = 0.39 * (dbh)^{2.142} \quad (5\text{-}2)$$

5.5. Modelling framework

This section provides the modelling approach used for the Bayesian and the classical geostatistical techniques.

5.5.1. The Bayesian geostatistical approach

The modelling framework for the frequentist and Bayesian geostatistical approaches using multi-spectral remote sensing Landsat-8 and Sentinel-2 as data sources is illustrated in **Figure 5-3**. Both approaches culminate in C stock predictions whose qualities were evaluated using cross validation statistics (**Figure 5-3**). We assumed the Bayesian hierarchical methodology in order to have a full account of the parameter uncertainty for the measured C stock as given in Eqn 5-3 (Finley, Banerjee and MacFarlane, 2011).

$$Y(s) = X^T(s)\beta + w(s) + \varepsilon(s) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5\text{-}3)$$

where

$w(s)$ represents the spatial random effects term

β denotes a vector of covariate coefficients

$X^T(s)$ denotes a vector of predictors measured at the same location as $Y(s)$

$Y(s)$ denotes the sampled C stock variable

$\varepsilon(s)$ denotes white noise assumed independent and identically distributed ($i.i.d. N(0, \sigma_\varepsilon^2)$)

We made simultaneous estimation and prediction of C stock parameters by making use of the Markov Chain Monte Carlo (MCMC) technique to derive and calculate C stock predictions at unvisited locations as in Eqn 5-4.

$\hat{Y}(s)' = X^T(s)'\hat{\beta} + \hat{w}(s)'$ …… (5-4)

All hierarchical modelling were made using the *spBayes* package (Finley, Sudipto, and Carlin (Finley, Sudipto and Carlin, 2007) of the R Statistical and Computing Environment (R Core Development Team, 2008). According to Gelfand (2012), the vector of model parameters, $\boldsymbol{\theta} = \beta, \sigma^2, \phi, \tau^2$ is treated as random and mutually independent variables and are assigned prior distributions. We therefore sampled the posterior distribution of the parameters of interest, $\boldsymbol{\theta}$, as in Eqn 5-5.

$p(\boldsymbol{\theta}|\boldsymbol{y}, X) \propto p(\boldsymbol{\theta})\ x\ N(\boldsymbol{w}|0, \Sigma_w)\ x\ N(\boldsymbol{y}|X^t\boldsymbol{\beta} + \boldsymbol{w}, \Sigma_\varepsilon)$ …… (5-5)

Eqn 5-5 was employed in the quantification of uncertainties in model parameters and C stock predictions at unsampled locations derived using Eqn 5-6.

$p(y_0|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{x}_0) \propto \int p(y_0|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{x}_0)p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X})d\boldsymbol{\theta}$ ……... (5-6)

Where;

y_0 represents the predicted C stock at a site s_0 and x_0 are the predictor values at site s_0.

The overall mean of the sampled C stock was assigned a normal prior whilst the regression coefficients were assigned a multivariate normal prior. Since the study made use of ancillary data from Landsat-8 and Sentinel-2 sensors, we specified two classes of priors for hierarchical modelling of the outcome variable. We assigned an inverse gamma distribution for the C stock data and measurement error variance whilst the spatial decay parameter, ϕ, was assigned a uniform prior as indicated by $p(\boldsymbol{\theta}_1)$ and $p(\boldsymbol{\theta}_2)$ for the Landsat-8 and Sentinel-2 set of priors, respectively. The assignment of the prior distribution on the spatial decay parameter was guided by the maximum distance between sampling locations (2 413 m) within the geographic domain ($\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in D$) of the studied region.

As we expected the white noise error variance (nugget, σ_ε^2) to be smaller than the structured variance, $\sigma_{\boldsymbol{w}}^2$, scale parameter values were adopted in order to express the preference that $\sigma_\varepsilon^2 < \sigma_{\boldsymbol{w}}^2$ (Demirhan and Kalaylioglu, 2015). A uniform prior with support covering the geographic domain of the study area was assigned to the spatial decay parameter, ϕ. Prior distributions on modelled parameters were derived from covariance parameters from the exploratory variograms of the two multispectral remote sensing sensors. A Metropolis-Hastings algorithm for MCMC was utilised (Finley, Banerjee and MacFarlane, 2011). We then specified an

algorithm of one chain comprised of 20 000 MCMC iterations for the posterior densities of the model parameters. 15 000 chains were discarded as burn-in.

5.5.2. Bayesian model validation and diagnostic evaluation

We compared the spatial model, the spatial intercept only model and the independent error model (simple multiple linear regression) for assessing the performance of predictions from the Bayesian hierarchical modelling approach with the Uniform (***Unif***) and Inverse Gamma (***IG***) priors on the spatial decay and the spatial random effects as illustrated for θ_1 and θ_2 respectively.

$$p(\boldsymbol{\theta_1}) = \boldsymbol{Unif}(\phi|0.38, 0.0012)\ \boldsymbol{x}\ \boldsymbol{IG}(\sigma^2|0.52, 1.58)\ \boldsymbol{x}\ \boldsymbol{IG}(\tau^2|0.1, 1.58)\ \boldsymbol{x}\ \boldsymbol{MVN}(\beta|\mathbf{0}, \Sigma_{\boldsymbol{\beta}})$$

$$p(\boldsymbol{\theta_2}) = \boldsymbol{Unif}(\phi|0.38, 0.0012)\ \boldsymbol{x}\ \boldsymbol{IG}(\sigma^2|0.052, 0.0028)\ \boldsymbol{x}\ \boldsymbol{IG}(\tau^2|0.1, 1.52)\ \boldsymbol{x}\ \boldsymbol{MVN}(\beta|\mathbf{0}, \Sigma_{\boldsymbol{\beta}})$$

The predictive performance of each of the three models was tested using a *k-fold* cross validation algorithm, which performed the cross-validation through random splitting of the measured 191 C stock observations into approximately ten equally seized segments (Duchene et al., 2016).

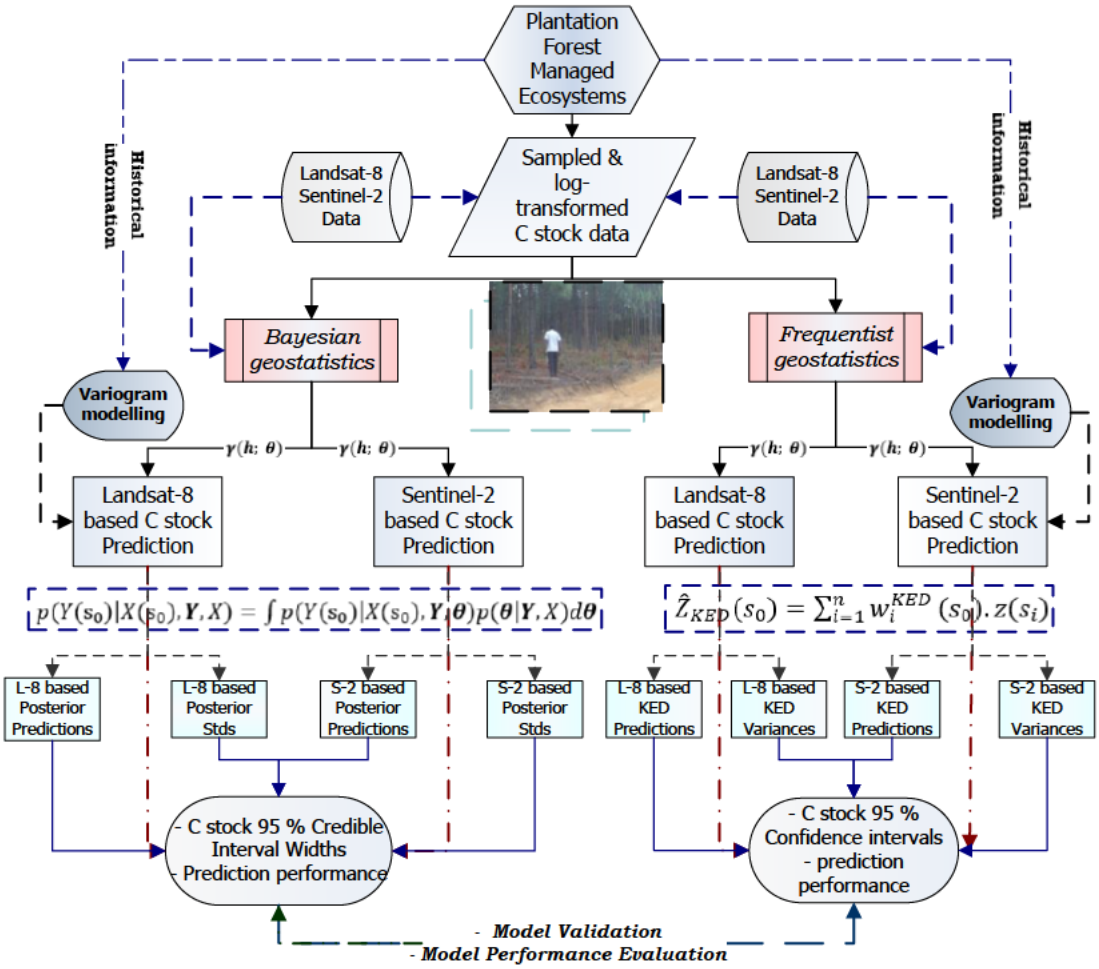


Figure 5-2: Bayesian and frequentist geostatistical modelling framework

We calculated validation metrics of Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and other goodness of fit statistics like the Deviance Information Criterion (DIC) for ranking candidate models on their ability to fit the data (Spiegelhalter *et al.*, 2002). Desirable and better fitting models display lower values of the DIC whilst better performing models would have the lowest $k(k = 10)$ fold MAE and RMSE. We also assessed the conformance of the models to the modelling assumptions using the graphical diagnostic plots of model residuals (Jackman, 2000; Green, Finley and Strawderman, 2020).

5.6. The frequentist geostatistical approach

This section provides information on how the frequentist geostatistical approach was employed in C stock spatial interpolation.

5.6.1. Carbon stock spatial interpolation

We utilised ordinary kriging (OK) as a baseline for assessing how covariates derived from Landsat-8 and Sentinel-2 impact the C stock model using a frequentist geostatistical methodology. Covariates are a means of enhancing the predictive properties of spatial models. The manner in which auxiliary variables lead to better kriging estimates than the ordinary kriging algorithm is shown by the work of (Spiegelhalter *et al.*, 2002; Jackman, 2000). We therefore employed kriging with External Drift (KED) as a geostatistical methodology for modelling the spatial distribution of C stock in managed plantation forest ecosystems using vegetation indices derived from Landsat-8 and Sentinel-2. As Eqn 5-7 shows, the KED algorithm restricts stationarity within a search neighborhood and provides detailed information compared to ordinary kriging (Kupfersberger, Deutsch and Journel, 1998). The standard measurement unit for forest C stock accounting is a hectare (Gibbs *et al.*, 2007) and hence, we utilised a 100 m x 100 m grid resolution for kriging, with the normality assumptions tested with residuals of the selected liner model (Hengl *et al.*, 2004).

$$Z^*_{KED}(\mu) = \sum_{\alpha=1}^{n(\mu)} \lambda^{KED}(\mu) Z(\mu_\alpha) \quad \text{..... (5-7)}$$

Where;

$Z^*_{KED}(\mu)$ denotes the KED estimated value at site μ

$\lambda^{KED}(\mu)$ denotes the KED weights pertaining to n samples at site μ

$Z(\mu_\alpha)$ denotes the sample values inside the search neighbourhood

5.6.2. Frequentist model validation and diagnostics

We followed the method of cross-validation as outlined in Webster *et al.* (2006) for assessing the quality of predictions of each of the C stock predictive models constructed from Landsat-8 and Sentinel-2 satellite sensors. We therefore presented the validation statistical metrics in the form of RMSE, MSE, ME and the Prediction Residual Sum of Squares (PRESS).

5.7. Variogram modelling of the regionalised variable

The Bayesian and frequentist geostatistical modelling philosophies both make use of the variogram as the basis for establishing spatial covariance parameters. We presented the variogram of residuals from the linear modelling of C stock with predictors from both Landsat-8 and Sentinel-2 derived vegetation indices. Variogram modelling was therefore utilised as a basis for assessing the robustness of the spatial correlation structure of the modelled C stock response variable (Stoyan, 2014; Pascual, Tupinambá-Simões and de Conto, 2022). The regionalised variable was also transformed into a logarithmic scale to ensure conformance to the normality of residuals modelling assumptions using the Box-Cox transformation technique (Box and Cox, 1982).

5.8. Results

This section provides specific information on the results of the study.

5.8.1. C stock descriptive statistics

Descriptive statistics of the sampled C stock data relating to the measured forest parameters are illustrated in **Table 5-1**. The mean C stock for the *Eucalyptus camaldulensis*, *Eucalyptus grandis* and *Pinus patula* species were 2 485.3 $MgCha^{-1}$, 405.7 $MgCha^{-1}$ and 377.9 $MgCha^{-1}$, respectively. The *Eucalyptus camaldulensis* had the highest C stock density in the sampled plantation forest as illustrated in **Table 5-1**.

Table 5-1: Summary statistics of the measured C stock plantation forest parameters

Statistic	*Eucalyptus camaldulensis*			*Eucalyptus grandis*			*Pinus patula*		
(MgCha-1)	DBH	Height	C stock	DBH	Height	C stock	DBH	Height	C stock
Mean	81.4	60.6	2485.3	67.4	70.6	405.7	56.8	58.6	377.9
Median	77.4	52.7	1470.3	51.4	49.7	327.8	43.5	38.7	295.4
Max	231.9	88.9	8998.2	97.9	90.1	429.8	64.3	66.6	600.3
Min	11.4	23.8	13.7	14.7	27.8	111.3	10.6	19.4	9.7
n	97	-	-	60	-	-	34	-	-
s.td	57.6	-	-	51.7	-	-	48.9	-	-

5.8.2. Hierarchical Bayesian geostatistical approach

The variogram of residuals of predictors obtained from the Landsat-8 and the Sentinel-2 satellite sensors provided priors of the modelled parameter specifications of the σ_{ε}^2 and the σ_{w}^2 for the geostatistical approach, **Figure 5-3 (a)** and **Figure 5-3 (b)** respectively. Sentinel-2 derived predictors hold more influence on the spatial distribution of the modelled regionalised variable than what we get with its counterpart from the Landsat-8 derived vegetation indices. Spatial dependence is greatly reduced in Sentinel-2 derived and modelled variogram (**Figure 5-3 (a-b)**) as the variable displays a strong spatial structure attributed to the finer spatial and spectral resolution of Sentinel-2 over Landsat-8 data. The modelled C stock using the Bayesian hierarchical approach therefore adopted the scale parameters derived from the variogram exploratory analysis of the outcome variable using ancillary data from Landsat-8 and Sentinel-2 data (Duchene *et al.*, 2016).

In accordance with the foregoing logic, scale parameter values derived from **Figure 5-3** puts on a constraint on the parameter space of the probability distribution function (Goulard and Voltz, 1992). We fitted the variogram of residuals for both sensors using the exponential covariance model as illustrated in **Figure 5-3**.

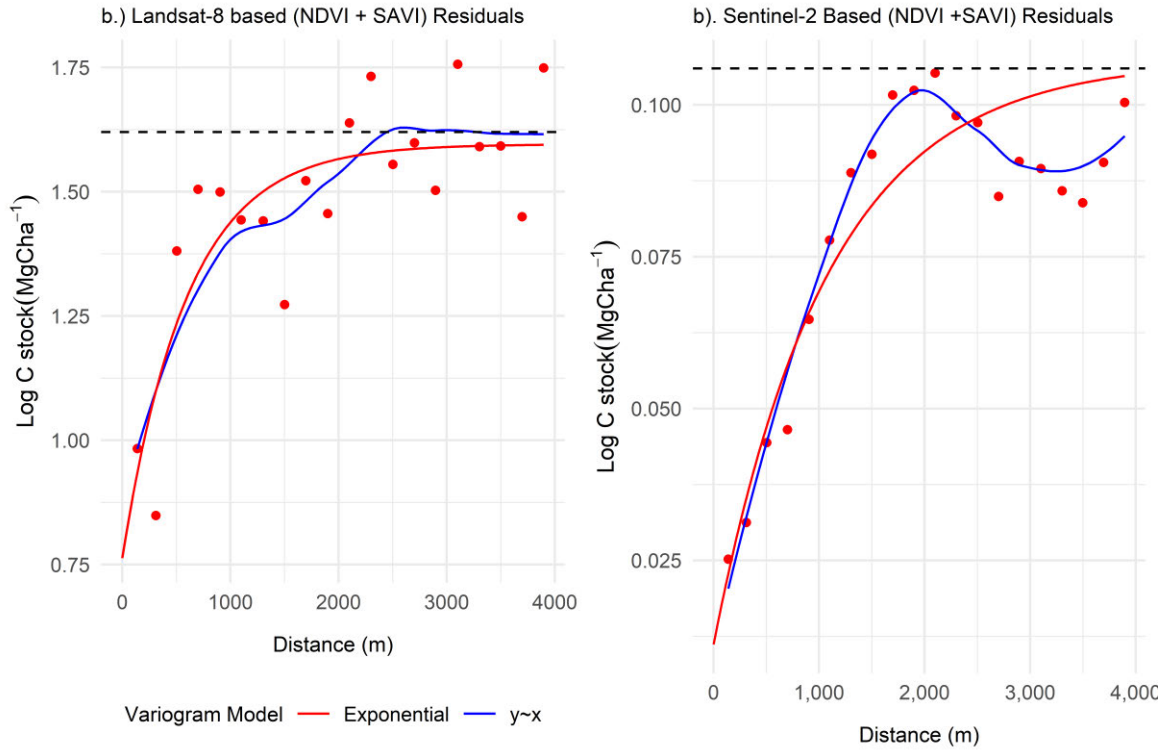


Figure 5-3: Variogram modelling for Landsat-2 and Sentinel-2 derived vegetation indices. The asymptote for the theoretical variogram model is shown by the black dotted line.

5.8.2.1. C stock and medium resolution sensor derived vegetation indices

Medium resolution derived vegetation indices utilised as predictors for C stock modelling in the form of $NDVI$, $SAVI$ and EVI showed different results for the two sensors. Landsat-8 based C stock model demonstrated $NDVI$ as the only predictor for C stock whilst Sentinel-based C stock employing the same covariates illustrated both $NDVI$ and $SAVI$ as significant predictors of C stock. Amongst the tested Landsat-8 and Sentinel-2 derived vegetation indices, only $NDVI$ was significantly different from zero as its 95% Credible Intervals (CI) for both sensors exclude zero as illustrated in **Table 5-2**.

The 95% CI of $2.63 \leq NDVI \leq 6.30$ and $6.06 \leq NDVI \leq 6.51$ for the Landsat-8 and Sentinel-2 based C stock prediction models, respectively, puts Sentinel-2 based predictive model in a stronger position due to the strength in the predictor coefficient. It is evident from **Table 5-2** that the posterior distribution of independent variable coefficients for Landsat-8 and Sentinel-2 displayed marked differences. Larger $NDVI$ coefficients in Sentinel-2 derived C stock model signifies the relative importance of using Sentinel-2 for C stock prediction over Landsat-8 derived vegetation indices.

Table 5-2: Landsat-8 and Sentinel-2 derived predictors of C stock. NDVI =Normalised Difference Vegetation Index, SAVI = Soil Adjusted Vegetation Index, EVI = Enhanced Vegetation Index, σ_w^2= spatially structured variance, σ_ε^2= White noise, ϕ = spatial decay parameter

Parameter	Landsat-8 OLI C stock Model				Sentinel-2 MSI C stock Model			
	Mean	s.d	2.5%	97.5%	Mean	s.d	2.5%	97.5%
Intercept	1.34	0.49	0.37	2.27	0.93	0.24	1.42	-0.49
NDVI	4.49	0.94	2.63	6.30	6.30	0.11	6.06	6.51
SAVI	-0.50	0.72	-1.55	1.26	0.02	0.38	-0.72	0.77
EVI	-0.50	0.55	-1.65	0.53	0.01	0.11	-0.19	0.22
σ_w^2	1.47	0.39	0.76	2.22	0.07	0.01	0.053	0.10
σ_ε^2	0.39	0.15	0.13	0.68	0.005	0.004	0.0005	0.01
ϕ	0.0013	0.000	0.0013	0.0014	0.0012	0.0003	0.0014	0.0023

The Landsat-8 derived C stock model has a slightly stronger spatial correlation compared to its Sentinel-2 C stock counterpart as evidenced from the estimates of the effective ranges of the models. We observed an effective range ($\approx {}^{3}/_{\phi}$) of 2500 meters with a ($2142 \leq \phi \leq 2307$) 95% CI for the Landsat-8 C stock whilst Sentinel-2 based C stock model gave an effective range of 1667 meters with a ($1304 \leq \phi \leq 2142$) 95% CI as illustrated in **Table 5-2**. Sentinel-2 based $NDVI$ takes away most of the spatial correlation structure in C stock than the same predictor derived from Landsat-8 OLI. Differences in spectral properties between Landsat-8 (11 spectral bands) and Sentinel-2 (13 spectral bands) vindicates this observation (El-Askary *et al.*, 2014; Jia *et al.*, 2014). In both models of Landsat-8 and Sentinel-2, the spatially

structured variance, σ_w^2 is higher than the white noise variance (Isaaks and Srivastava, 1989; Diggle and Ribeiro Jr, 2007).

5.8.2.2. Bayesian based C stock predictions

We fitted models with Landsat-8 and Sentinel-2 derived spectral auxiliary variables for predicting C stock at unsampled sites within the studied region. Covariates in the form of $NDVI, SAVI$ and EVI were derived from a 10 000 m^2 gridded raster, thereby making the predicted C stock represent average values in every raster pixel.

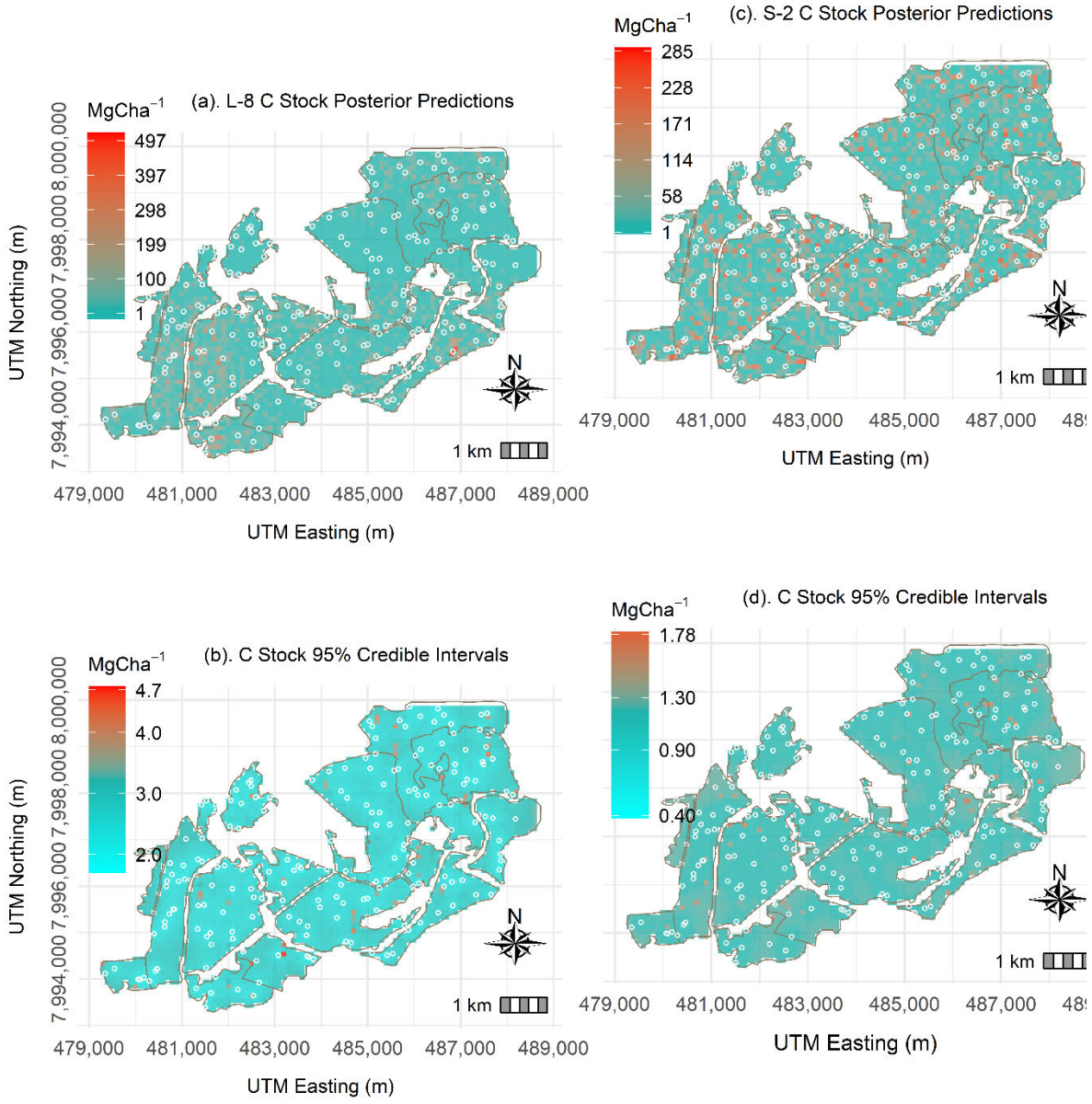


Figure 5-4: Bayesian Landsat-8 (L-8) and Sentinel-2 (S-2) based C stock posterior predictions alongside 95 % credible intervals. Points represent locations of sampled C stock.

$NDVI$ is the only significant predictor for C stock prediction in managed plantation forest ecosystems as the 95% CIs of the other covariates contain zero (**Table 5-2**). The significance of $NDVI$ as a vegetation index correlated with the biophysical properties of vegetation , leaf index being one of them is well established (Baloloy *et al.,* 2018). Sentinel-2 derived C stock predictions are more credible than their Landsat-8 derived C stock predictions. Landsat-8 based C stock predictions have higher uncertainty compared to Sentinel-2 based C stock predictions. This is demonstrated in the 95% posterior predictions illustrated in **Figure 5-4 (a)**, indicating C stock 95% CI to be greater than the Sentinel-2 predicted values. This makes Landsat-8 based predictions highly uncertain and less precise than the Sentinel-2 based C stock predictions. Landsat-8 and Sentinel-2 based C stock predictions alongside their 95% CIs are illustrated in **Figure 5-4 (a)** and **Figure 5-4 (b)** and **Figure 5-5 (a)** and **Figure 5-5 (b)**, respectively. C stock predictions range between 1 $MgCha^{-1}$ to 497 $MgCha^{-1}$ and between 1 $MgCha^{-1}$ and 285 $MgCha^{-1}$ for the Landsat-8 and Sentinel-2 sensors respectively. Higher C stock values are predicted in the southern part of the studied region in the Landsat-8 based predictive models than the Sentinel-2 models. In spite of this trend, Sentinel-2 based C stock predictive model displays higher C stock values uniformly across the study area with a smaller magnitude than Landsat-8 based models (**Figure 5-4 (a-b)** and **Figure 5-5 (a-b)**).

As such, the Sentinel-2 model seems to underpredict C stock at unsampled locations compared to its Landsat-8 based C stock predictive model counterpart. The slight underprediction in Sentinel-2 can partly be attributed to the finer spatial and spectral resolutions of the sensor within the visible and near-infrared segments of the electromagnetic spectrum (EMS) (Sovdat *et al.,* 2019; Wang *et al.,* 2020; Chinembiri, Mutanga and Dube, 2023). Enhancements in the spectral and spatial resoultion of Sentinel-2 confirms the much shorter 95% Credible Interval Widths (CIWs) displayed by the Sentinel-2 based C stock predictive model in **Figure 5-4 (d)** (0.40 - 1.78) $MgCha^{-1}$ than the Landsat-8 based predictive model in **Figure 5-4 (b)** (2.0 - 4.7) $MgCha^{-1}$. Predicted values in both new generation remote sensing based models look better compared to the ones reported in previous research including Jiang et al. (2021) who established a mean AGB RMSE of 40.9 $MgCha^{-1}$ in northeast China and Dang *et al.* (2019) who determined an AGB RMSE of 36.67 $MgCha^{-1}$ in Vietnam using the Random Forest (RF) algorithm.

Furthermore, Takagi et al (Takagi *et al.*, 2015) employed LiDAR for the prediction of forest biomass in Hokkaido, Japan, and determined a RMSE biomass prediction of 19.1 $MgCha^{-1}$. Differences between the prediction accuracy results reported in literature and our study can

also be justified by the differences in forest density since the erstwhile studies were carried out in subtropical rainforest biomes.

5.8.2.3. Model validation and diagnostics

The *k-fold* cross validation metrics employed in the assessment of the Bayesian based predictive models are presented in **Table 5-3**. The Sentinel-2 based C stock predictive model is the top performing model in terms of the RMSE (0.17 $MgCha^{-1}$) and Mean Absolute Error (MAE) (0.13 $MgCha^{-1}$). The Sentinel-2 model presents predictive ability almost at the benchmark nominal coverage of 95% [60, 73]. The Landsat-8 based C stock predictive model has an 85.4% coverage for the 95% prediction intervals coupled with higher RMSE and MAE (**Table 5-3**).

Table 5-3: Validation statistics for C stock Bayesian based C stock predictive models

Model Evaluation criterion	Landsat 8 Derived Predictors			Sentinel-2 Derived Predictors		
	Independent Error Model	Spatial Intercept only Model	Spatial Model	Independent Error Model	Spatial Intercept only Model	Spatial Model
RMSE ($Mgha^{-1}$)	1.23	0.97	0.97	0.31	1.18	0.17
MAE ($Mgha^{-1}$)	0.93	0.53	0.57	0.26	0.77	0.13
CRPS ($Mgha^{-1}$)	0.72	0.38	0.38	0.20	0.56	0.14
CVG (%)	91.67	85.42	85.42	95.83	89.58	100.00
DIC	220.8	48.40	71.0	-201.3	283.5	-564.5

Figure 5-5 (a) and **5-5 (b)** illustrates the scatterplots of observed C stock against the predicted C stock alongside the 95% intervals for both Landsat-8 and Sentinel-2 C stock based predictive models. Evidence of the Sentinel-2 based C stock predictive model performing better than its Landsat-8 C stock-based counterpart is clear from the scatter plot of the model in **Figure 5-5 (b)**. It is evident from the model diagnostics illustrated in **Figure 5-5 (a)** that the Landsat-8 based C stock predictive model tends to overpredict some C stock values as compared to the Sentinel-2 C stock based predictive model (**Figure 5-5 (b)**). This makes the Sentinel-2 C stock-based model favourable compared to the Landsat-8 predictive model.

5.8.3. Frequentist geostatistical modelling

5.8.3.1. C stock density

We assessed C stock density of the sampled in the studied region using species type as a possible source of C stock density variability and established the mean C stock of *Eucalyptus camaldulensis, Eucalyptus grandis* and *Pinus patula* species to be 35.10 $MgCha^{-1}$, 37.38 $MgCha^{-1}$ and 29.45 $MgCha^{-1}$, respectively. Despite *Pinus patula* being the most dominant species in the sampled region (**Figure 5-1**), Eucalyptus *grandis* has the highest concentration

of C stock. An evaluation of the C stock density of the various plantation forest species making up the study area using an Analysis of Variance (ANOVA) shows that the C stock density of the different tree species are not significantly different from each other ($F_{2,188} = 0.21, p = 0.811$). Tukey–Kramer (Kramer, 1956) multiple comparison test conducted in order to avoid the risk of accumulating false positives as a result of multiple tests being carried out at the same time were also not significantly different. An evaluation of C stock density categorised by management style conducted in Nepal showed significant differences between the C stock density stored by different community forests (Chinembiri *et al.*, 2013).

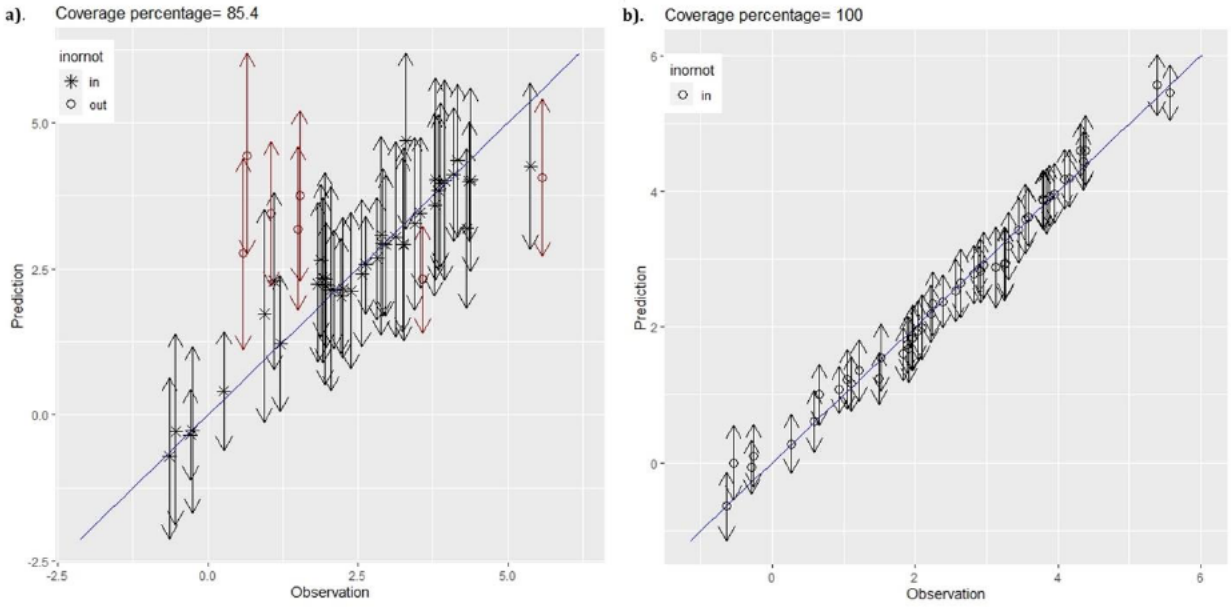


Figure 5-5: (a). Predictions (***MgC ha^{-1}***) against observed C stock (***MgCha^{-1})*** for the Landsat-8 based spatial model. (b). Predictions against observed C stock for the Sentinel-2 based spatial model alongside 95% intervals.

5.8.3.2. Landsat-8 and Sentinel-2 based C stock linear modelling

Linear modelling of C stock using sampled C stock data and multispectral remotely sensed data showed *NDVI* to be a significant predictor for the Landsat-8 derived vegetation indices whilst *NDVI* and *SAVI* were significant predictors for the Sentinel-2 derived vegetation indices (**Table 5-4**).

Table 5-4: Landsat-8 and Sentinel-2 best linear model of feature space

Predictors	**Landsat-8 based linear model**			**Sentinel-2 based linear model**		
	Coefficient	**p-value**	***α=0.05***	**Coefficient**	**p-value**	***α=0.05***
Intercept	0.03	0.93	*Insignificant*	-0.30	0.31	*Insignificant*
NDVI	7.67	0.00	*Significant*	6.69	0.00	*Significant*
SAVI	1.04	0.11	*Insignificant*	1.24	0.01	*Significant*
EVI	0.33	0.57	*Insignificant*	0.20	0.15	*Insignificant*

5.8.3.3. Landsat-8 and Sentinel-2 based KED predictions

Figure 5-6 (a) illustrates Landsat-8 based C stock KED predictions alongside their prediction variances. Results of KED using the spatial dependencies in the outcome variable gave C stock predictions ranging from $1 \leq MgCha^{-1} \leq 327$ and a corresponding standard error ranging from $1.1 \leq MgCha^{-1} \leq 5.1$. As shown in Figure 6 (b), C stock predictions closer to sampled data displayed smaller uncertainty compared to predictions made at remote locations within the study domain. Thus, the Landsat-8 based KED C stock variances illustrated in **Figure 5-6 (b)** are not better than 26.01 (= 5.1^2) $MgCha^{-1}$ for the studied managed plantation forest ecosystem. High C stock density dominate predictions at the original support (500 m^2) where primary data was derived using sampled field data. This is because kriging is a geostatistical method that characterizes values of an outcome variable (C stock) closer to the original data locations which tend to have similar statistical properties to the sampled value at that point than values obtained in remote locations (Tveito, Wegehenkel and Wel, 2008; Chinembiri *et al.*, 2013).

As illustrated in **Table 5-4**, we made Sentinel-2 based C stock KED predictions using $NDVI$ and $SAVI$ as aides of the predictive model and was made after ordinary variogram modelling of the primary variable. Incorporation of $NDVI$ and $SAVI$ as independent variables gave reduced total sill of the modelled variogram **(Figure 5-3(b))** and the subsequent shortening of the range of spatial dependence. Hence, Sentinel-2 derived vegetation indices utilised as predictors in the C stock model predicted C stock with a $1 \leq MgCha^{-1} \leq 290$ range and an accompanying standard error ranging from $2 \leq MgCha^{-1} \leq 11$ (**Figure 5-6 (c)** and **Figure 5-6 (d))**. Finer spatial and spectral characteristics within the visible and NIR of the Sentinel-2 satellite sensor resulted in $SAVI$ being incorporated in the C stock predictive model in addition to the $NDVI$ (Dube and Mutanga, 2015; Gupta, Kamble and Machiwal, 2017).

Locations closer to the margins of the sampled domain show an increasing trend in the prediction standard error as these locations are farther from sampled C stock observations. As **Figure 5-3(b)** and **Figure 5-6 (d)** show, the significant reduction in the range of spatial dependence in the modelled C stock data demonstrate how Sentinel-2 driven model of feature space, carrying both $NDVI$ and $SAVI$ has on the spatial correlation structure of the outcome variable. In line with theory, the KED calculated error variance appear to rely on the sampled data configuration in which uncertainty decays towards the sampling sites (Isaaks and Srivastava, 1989; Li *et al.*, 2017).

As the results in **Figure 5-6 (a-b)** and **Figure 5-6 (c-d)** show, the Landsat-8 based C stock model predicts higher C stock in the southern and northern parts of the sampled region of the sampled domain. On the contrary, the Sentinel-2 based C stock predictive model predicts higher C stock uniformly across the sampling domain. This can be explained by the precision with which the much-improved spatial resolution of Sentinel-2 allows the sensor sensitivity to forest parameter spectral signal within the NIR and visible regions of the electromagnetic spectrum.

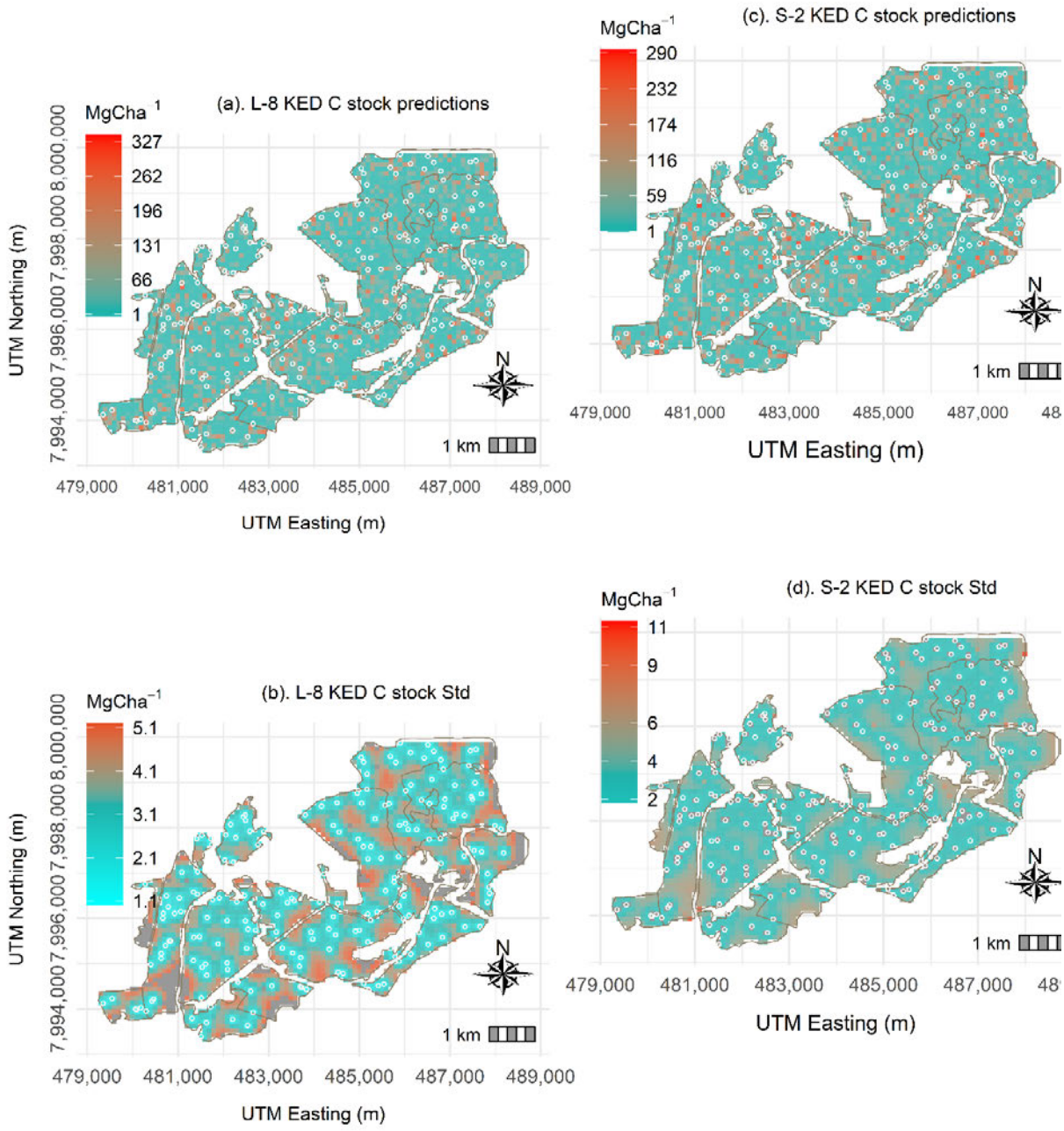


Figure 5-6: Frequentist Landsat- 8 (L-8) and Sentinel-2 (S-2) based C stock predictions and 95 % confidence intervals. Points represent locations of sampled C stock.

5.8.3.4. Frequentist geostatistical predictive model evaluation

We evaluated the frequentist-based C stock predictive models using leave one out-cross validation statistics and model residual diagnostics. As illustrated in **Table 5-5**, the Sentinel-2 based C stock predictive model has the lowest RMSE compared to its Landsat-8 based C stock model. Validation statistics shown in **Table 5-5** therefore suggest the Sentinel-2 based KED C stock model to be the most ideal under the frequentist geostatistical approach as it has the best characteristics. This is evident from the predictive model's RMSE (1.19 $MgCha^{-1}$) and PRESS (6.06).

Table 5-5: Frequentist geostatistical model diagnostics test statistics

Modelling approach	Test statistic	*p-value*	Modelling technique
Frequentist approach	0.097	0.264	Landsat-8
Frequentist approach	0.132	0.136	Landsat-8
Bayesian approach	0.975	0.367	Sentinel-2
Bayesian approach	0.773	0.278	Sentinel-2

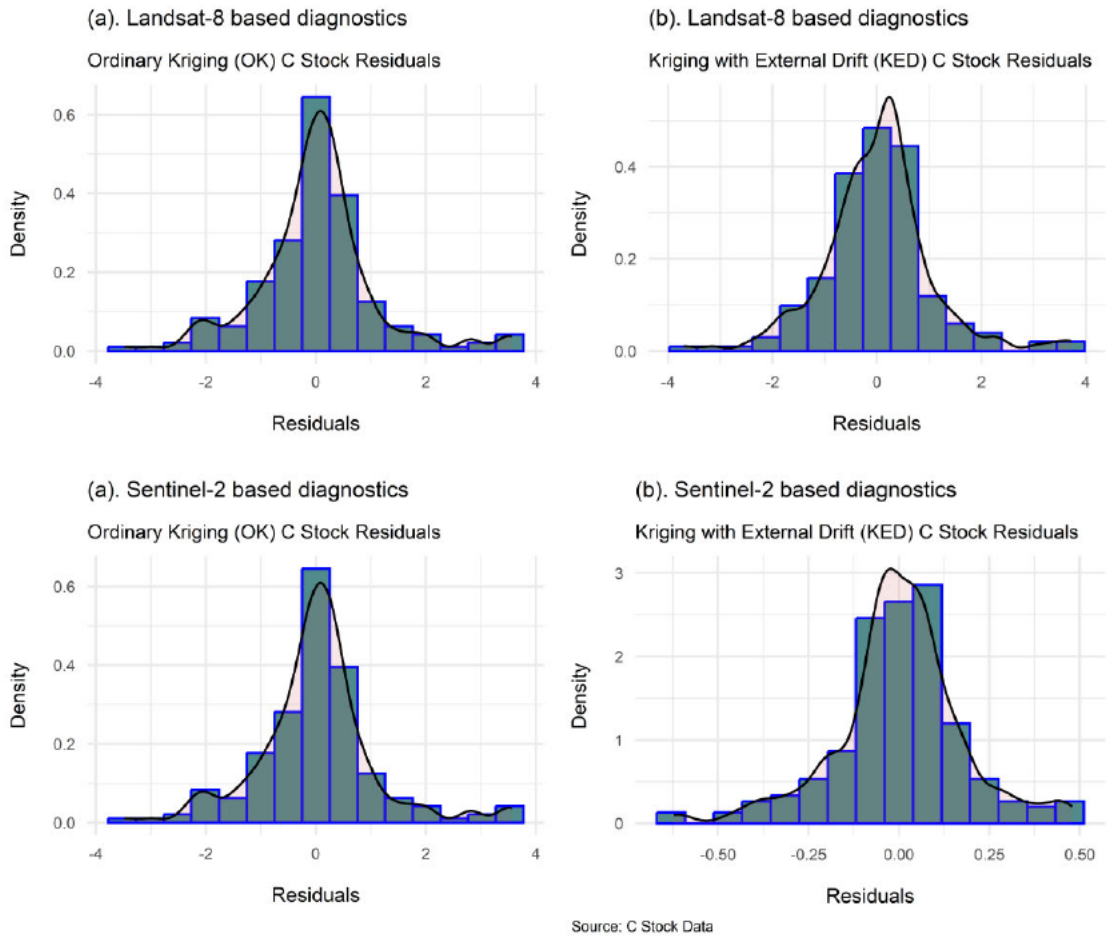


Figure 5-7: Frequentist based Landsat-8 and Sentinel-2 C stock KED residual diagnostics

Table 5-6: Frequentist geostatistical C stock prediction validation statistics

Predictors	Landsat-8 based C stock predictions		Sentinel-2 based C stock predictions	
	Ordinary Kriging (OK)	Kriging with External Drift (KED)	Ordinary Kriging (OK)	Kriging with External Drift (KED)
ME	1.01	1.00	1.01	1.00
MSE	1.01	1.00	1.01	1.00
RMSE	2.94	2.84	2.91	1.19
PRESS	222.34	208.42	222.34	6.06

The supremacy of the Sentinel-2 based C stock predictive model is also corroborated by the model diagnostics illustrated in **Figure 5-7**, showing robust model residuals that are evidently insensitive to outlying observations. Furthermore, the test statistics regarding the normality of the model residuals for the two kriging variants (**Table 5-5**) show the *p*-values more than the 0.05 test statistic, leading to the rejection of the null hypothesis of non-normal residual errors. Rejection of the Shapiro normality null hypothesis holds for both Lnadsat-8 and Sentinel-2 based C stock predictive models (**Table 5-5**). The best linear model of feature space making use of Landsat-8 and Sentinel-2 derived vegetation indices gives Sentinel-2 based C stock predictive model with a more symmetrical error distribution than its Landsat-8 based counterpart (**Figure 5-7**). This fact is confirmed and reinforced in the residuals of the Sentinel-2 C stock predictive model.

5.9. Bayesian and frequentist-based C stock predictive model summaries

A summarised overview of the performance of the tested inferential approaches, one using the Bayesian framework and the other using the frequentist approach, are illustrated in **Table 5-6**. We deduced from the summary statistics depicted in **Table 5-6** that all the Landsat-8 based C stock predictive models from either the Bayesian or the frequentist paradigms tend to overpredict the modelled C stock values whilst the Sentinel-2 based C stock based predictive models gave just within range C stock values. The Sentinel-2 based C stock predictive model from the Bayesian statistical paradigm offers the best qualities of C stock prediction model in terms of the quality of predictions and ability to predict true values outside the sampled data (**Table 5-6**).

5.10. Discussion

This section discusses the results of the study with regards to the performance of the Bayesian and the frequentist approach in predicting C stock.

5.10.1. Bayesian geostatistical approach and C stock predictions

The hierarchical design of stochastic models is intrinsically related to Bayesian inference where the probability distribution on observed C stock (in the present case) are specified in a hierarchical fashion (Sahu, 2022).

Table 5-7: Summaries of the Bayesian and the frequentist geostatistical approaches. *CIWs* = Confidence Interval Widths, ME = Mean Error, RMSE = Root Mean Squared Error

Validation criterion	Bayesian geostatistical approach		Frequentist geostatistical approach	
	Landsat-8 based C stock model	Sentinel-2 based C stock model	Landsat-8 based C stock model	Sentinel-2 based C stock model
RMSE	0.97	0.17	2.84	1.19
ME	0.57	0.13	1.01	1.00
Error/ CIWs	$2 \leq MgC \leq 4.7$	$0.4 \leq MgC \leq 1.8$	$1.1 \leq MgC \leq 5.1$	$2 \leq MgC \leq 11$
Prediction range	$1 \leq MgC \leq 497$	$1 \leq MgC \leq 285$	$1 \leq MgC \leq 327$	$1 \leq MgC \leq 290$
Conclusion	Overprediction	Perfect	Overprediction	Perfect

From the three tested predictors of $NDVI$, $SAVI$ and EVI for C stock prediction, we established $NDVI$ as the only significant covariate of the modelled C stock variable for both Landsat-8 and Sentinel-2 derived C stock predictive models. Because of the higher spatial resolution of Sentinel-2, $NDVI$ emerged as a stronger predictor of C stock for the Sentinel-2 based model than the Landsat-8 based C stock predictive model (Dube and Mutanga, 2015).

The importance of $NDVI$ as a vegetation index correlated with the biophysical properties of vegetation such as Leaf Area Index (LAI) is well well-known (Baloloy *et al.*, 2018). This shows the relative importance of employing Sentinel-2 for C stock prediction over Landsat-8 derived vegetation indices. Consequently, Sentinel-2 derived C stock predictions have better credibility than their Landsat-8 derived equivalent. Landsat-8 based C stock predictions have higher uncertainty compared to Sentinel-2 based C stock predictions. The utilisation of new generation remote sensing derived vegetation indices as predictors of C stock for a managed plantation forest ecosystem under a Bayesian framework is unique. Previous studies including Babcock et al. (Babcock *et al.*, 2015) and Babcock et al. (Babcock *et al.*, 2016) have used LiDAR on its own as predictors of forest biomass rather than a comparative approach as we employed in this research. The tendency of underprediction in Sentinel-2 derived C stock models can partly be attributed to the finer spatial resolutions of the sensor within the visible and near-infrared segments of the electromagnetic spectrum (EMS) (Sovdat *et al.*, 2019; Wang *et al.*, 2020; T.S. Chinembiri, Mutanga and Dube, 2023). As highlighted in (Dube and Mutanga, 2015), there are notable improvements in the radiometric resolution of the Landsat-8 OLI sensor to 12 bits, up from 8 bits.

Enhancements in the spatial resolution of Sentinel-2 confirms the much shorter 95% Credible Interval Widths (*CIWs*) displayed by the Sentinel-2 based C stock predictive model. Predicted values in both new generation remote sensing based models look more attractive compared to the ones reported in previous research including Jiang et al. (Jiang *et al.*, 2021) who established a mean AGB RMSE of 40.9 $MgCha^{-1}$ in northeast China and Dang et al. (Dang *et al.*, 2019) who determined an AGB RMSE of 36.67 $MgCha^{-1}$ in Vietnam using the Random Forest (RF) algorithm. Furthermore, Takagi et al. (Takagi *et al.*, 2015) employed LiDAR for the prediction of forest biomass in Hokkaido, Japan, and determined a RMSE biomass prediction of 19.1 $MgCha^{-1}$. Differences between the prediction accuracy results reported in literature and our study can also be justified by the differences in forest density since the erstwhile studies were carried out in subtropical rainforest biomes (Dang *et al.*, 2019; Wang *et al.*, 2020; Jiang *et al.*, 2021). For instance, Japan is regarded as the most forested country in the world, with approximately 70 % of its land forested compared to location of the present study, Zimbabwe, where about 40 % or 15 624 000 ha of the land is forested (FAO, 2005).

Machine learning methods for the mapping of AGB premised on Landsat-8 imagery were compared by Wu et al. (2016) and Xiong and Wang (2022) and Random forest method with an RMSE of 26.43 tons/ha was found to be superior to the other methods like the stochastic gradient boosting, k-nearest neighbour and support vector regression. The current study establishes AGB accuracies using the Bayesian and the frequentist geostatistical approach with prediction accuracies of the former outweighing that of the later. Bayesian methods utilised in this study are both superior to both the frequentist approach and methods utilised in literature for the mapping of AGB. However, Bayesian methods are not easy to implement and adopt by ordinary forest practitioner due to their complexity (Gelfand, 2012). It is this complexity and the lack of simple software packages that greatly hinder their adoption and operationalization for forestry monitoring.

5.10.2. Frequentist geostatistical approach and C stock predictions

The geostatistical kriging variant, KED, provides an optimal geostatistical technique under the frequentist paradigm that is employed in the description of spatial patterns and predicting values of a variable at unsampled locations and, consequently, evaluate the uncertainty associated with the predicted values (Wheeler and Waller, 2009). We employed and subjected the C stock model to the same new generation remote sensing derived predictors of vegetation indices as we did under the Bayesian approach and established some notable differences and similarities. Similar to the Bayesian C stock model for Landsat-8, $NDVI$ also came out to be

the only significant predictor in the frequentist Landsat-8 based C stock predictive. However, both $SAVI$ and $NDVI$ were significant predictors for the frequentist Sentinel-2 based C stock model. We employed KED for both the Landsat-8 and the Sentinel-2 based C stock models. Finer spatial scale of the Sentinel-2 sensor, coupled with the wider spectrum of its NIR band (760-900 nm) compared to the narrow spectrum of Landsat-8 NIR band (850-880 nm) resulted in $SAVI$ and $NDVI$ being incorporated in the C stock predictive model (Dube and Mutanga, 2015; Gupta, Kamble and Machiwal, 2017). The NIR band is widely known to be vital to the biophysical factors of vegetation assessment and monitoring. Consequently, the spectral response profile of Landsat-8 and Sentinel-2 display some minor differences in the NIR and visible regions of the electromagnetic spectrum, which could explain differences in the prediction performance as established in the current study (Babcock *et al.*, 2016; Pascual, Tupinambá-Simões and de Conto, 2022).

KED has been widely applied for estimation of AGB using other factors as predictors of the forest biomass (Hudson and Wackernagel, 1994; Tveito, Wegehenkel and Wel, 2008). For instance, Sales *et al.* (2007) made an average AGB prediction of 32 $Mgha^{-1}$ in the Brazilin Amazon. KED performed better in this study than the pure-based approach of the ordinary kriging algorithm as $NDVI$ and $SAVI$ formed the best linear model of feature space in the Sentinel-2 C stock predictive model. Most studies comparing the prediction performance of Landsat-8 and Sentinel-2, including Korhonen et al. (2017) and Meyer et al. (2019) have not established significant or systematic differences in the predictive performance of the aforementioned sensors. In the current study, both Bayesian and frequentist Landsat-8 based C models tend to overpredict the modelled and estimated C stock values across the sampled domain as their range fall outside the range of observed data. This implies that forest practitioners can feasibly exploit the scale and band spectrum characteristics in the Sentinel-2 for representative and accurate C reporting needed for monitoring and verification in climate change mitigation actions.

5.10.3.Comparative Bayesian and frequentist C stock predictive model evaluation

Sentinel-2 based prediction of C stock using the frequentist based KED illustrates predictions within the range of measured C stock values, but the variances are high compared to the same predictions made using Sentinel-2 within a Bayesian inferential framework. On the contrary, both frequentist and Bayesian C stock predictive models constructed using Landsat-8 overpredict the sampled C stock as the range of predicted values fell outside the observed C stock values. This observation is further bolstered by the results of the diagnostic residuals of

the models from the frequentist and the Bayesian techniques, where the plots of the observed versus the predicted values of C stock are perfectly predicted in the Bayesian based Sentinel-2 models. As reported in the study, Sentinel-2 based C stock predictive models are more accurate than their Landsat-8 equivalents for both the Bayesian and the frequentist inferential approaches. As we established in the present study, the Bayesian based Sentinel-2 C stock predictive model is more accurate than its frequentist-based Sentinel-2 C stock equivalent. Previous studies mapping AGB accuracies using either the frequentist or the Bayesian approach, coupled with satellite imagery data of Landsat-8 and Sentinel-2 report lower prediction accuracies (Jiang *et al.*, 2021; Wai, Su and Li, 2022). As hypothesised, this confirms the superiority of the Bayesian geostatistical approach in handling the uncertainty and improving the accuracy for predicting C stock compared to the frequentist geostatistical approach.

Important research in the space of biomass estimation using the Bayesian techniques include (Finley and Banerjee, 2008; T.S. Chinembiri, Mutanga and Dube, 2023). Recent studies assessing AGB distribution in temperate Eurepean ecosystems established accuracies of 17.52 Mg/ha Babcock et al. (2015) and 1.16 $MgCha^{-1}$ and 2.69 $MgCha^{-1}$ for Sentinel-2 based and Landsat-8 based C stock predictive models, respectively (Chinembiri, Mutanga and Dube, 2023). Other remote sensing and machine learning-based efforts in the estimation and prediction of AGB in rcent times include Do et al. (2022) who found that mangrove AGB predictions in Vietnam range from 6.51 to 368 $Mgha^{-1}$ and from 13.70 to 320.1 $Mgha^{-1}$ for remote sensing and Artificial Neural Networks, respectively. Because the present study utilised data from improved remote sensing platforms for predicting C stock, our reported accuracies surpass those reported in literature utilising different methodologies with the same satellite sensors with lower quality.

Past studies utilising the frequentist geostatistical approach separately from the Bayesian technique for C stock estimation are also well documented in literature. ales *et al.* (2007) mapped AGB in Brazilian Amazon using Kriging with External Drift and established prediction accuracies for different sample sizes ranging from 0 to 110 km for distances within 300 km radii from the prediction locations. The lowest RMSE for the estimated AGB for a 110-sample size was 32.8 $Mgha^{-1}$ whilst the lowest accuracy for the lowest sample size of $n > 0$ was 48.06 $Mgha^{-1}$. Furthermore, Jiang et al. (2022) predicted AGB in Wangyedia forest farm in China using Landsat-8 and the newly launched Landsat-9 and reported RMSEs of 16.83 tha^{-1} and 17.91 tha^{-1}, respectively. Wai, Su and Li (2022) coupled remotely sensed derived Sentinel-2 explanatory variables with geostatistics and machine learning algorithms in

Mayanmar for predicting above ground biomass and established accuracies of 24.91 $Mgha^{-1}$ and 34.72 $Mgha^{-1}$ for the Random Forest based ordinary kriging and the Random Forest based co-kriging, respectively.

As demonstrated by the results of Jiang *et al.* (2022), Landsat-8 built AGB estimation models can still be superior than AGB models built from the successor Landsat-9 sensor, despite the relative spectral and radiometric improvements in the latest Landsat-9 mission. This is also justified by the results of the present study as the results from both Landsat-8 and Sentinel-2 based Bayesian and frequentist approaches are not significantly different from each other. From a practitioner's point of view, the recommendation for the use and adoption of a particular sensor and statistical approach largely depends on the policy problem being addressed. Furthermore, the costs involved and ease of use and application of the sensors and methodology also play a bigger role. Despite its lack of simplicity and difficulties in implementation, the Bayesian approach is more appealing and pragmatic for natural resources monitoring and reporting as different studies of a particular region would eventually constitute a database where subsequent studies on carbon assessment would rely on, for updating of priors and posteriors in the prediction process.

5.11. Conclusions

This study set out to compare the prediction performance of two inferential geostatistical frameworks, one making use of a Bayesian geostatistical hierarchical approach and the other utilizing a frequentist geostatistical approach in C stock prediction in a managed plantation forest ecosystem in the eastern highlands of Zimbabwe. Broadband spectral indices from two multispectral remote sensing platforms of Landsat-8 and Sentinle-2 were employed as prediction aids for the geostatistical methodologies. We established notable differences between the two methods in the prediction of C stock in a managed plantation forest ecosystem, with the Sentine-2 based Bayesian hierarchical geostatistical approach yielding lower ($0.4 \leq MgCha^{-1} \leq 1.8$) prediction uncertainty than its frequentist geostatistical KED ($2 \leq MgCha^{-1} \leq 11$) modelling counterpart. In both geostatistical methods, the Sentinel-2 driven C stock prediction models outmanoeuvred their Landsat-8 driven C stock prediction models.

The bigger policy problem pertaining to climate change mitigation and climate change action being addressed in this study requires accurate and sustainable carbon accounting and verification tools. In that regard, despite the ease of application and use of the frequentist inferential methodology, we conclude that the Bayesian based technique, coupled with quality remote sensing information, is a better method for predicting C stock. This is a critical stepping

stone in carbon accounting for climate change adaptation and mitigation under the United Nations Framework Convention on Climate Change (UNFCCC). The Bayesian based Sentinel-2 C stock predictive model is preferable over its Sentinel-2 based frequentist counterpart for assisting natural resources managers and other forest practitioners in providing advice to governments on decision for afforestation and reforestation.

Using the Bayesian concept that "today's posterior is tomorrow's prior", we are able to construct a meaningful database of forest parameters, specifically C stock parameters, that would aid future estimation and prediction, even with limited datasets thereby eliminating the need for costly sampling campaigns. The sustainability of such databases is further made handy by the availability of freely available satellite data that is continuously being improved in terms of quality and scale of coverage. Conversely, the frequentist geostatistical approach cannot work optimally under limited sample sizes and therefore, makes it a costly alternative in the long run for updating C stock databases kept for C stock monitoring and accounting. We therefore conclude and recommend that the Bayesian based C stock prediction method, coupled with high quality remote sensing information like Sentinel-2, is a useful inferential statistical approach for presenting carbon stock in disturbed plantation forest ecosystems.

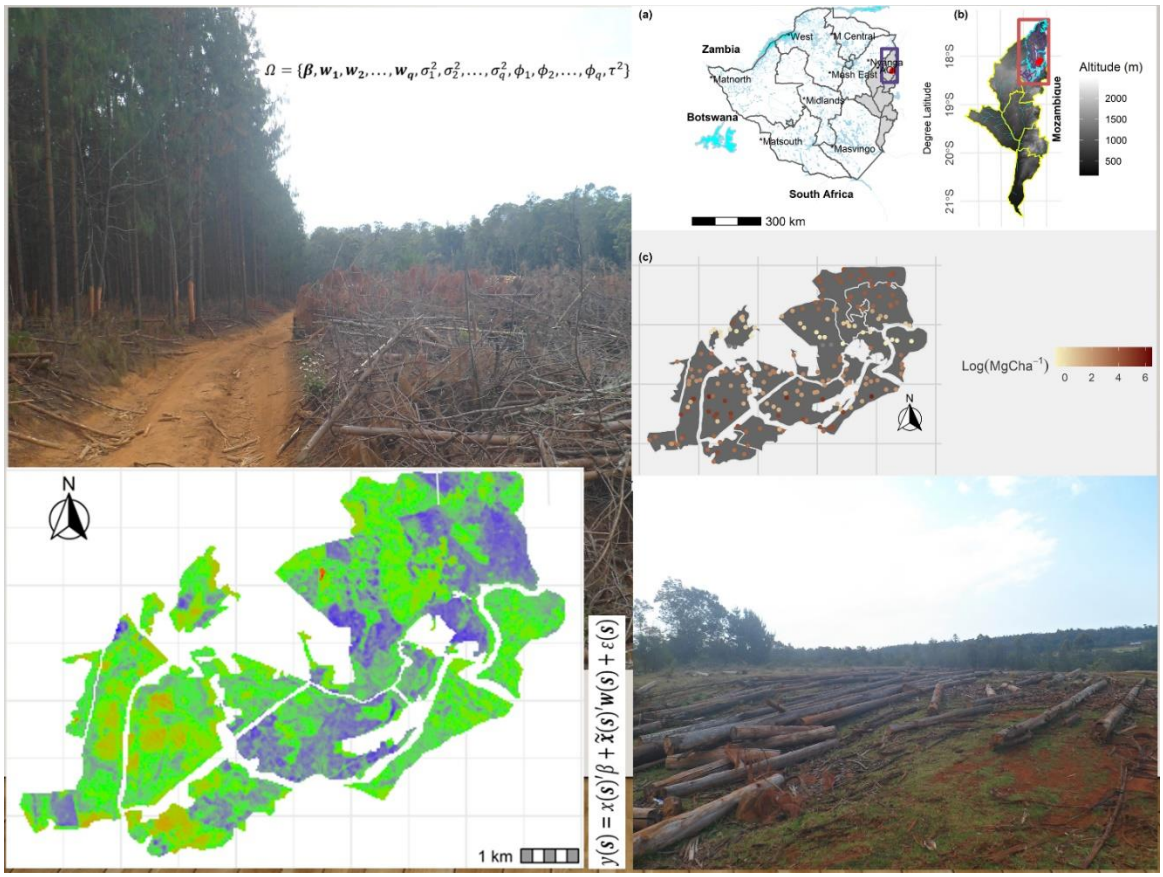
5.12. Summary

Easy to use analytical methods for predicting forest AGB together with cheap and readily available data sources like the Landsat mission and Sentinel-2, provide convenience for governments and other climate change-based organisations for C reporting needed under the UNFCCC. This chapter assessed how the prediction performance of the less commonly utilised Bayesian hierarchical approach with that of the frequentist geostatistical methodology. Despite difficulties in implementation, the Sentinel-2 driven Bayesian hierarchical approach prevails over its Landsat-8 based Bayesian hierarchical geostatistical counterpart and, over both the Sentinel-2 and the Landsat-8 based frequentist predictive models. As such, the Sentinel-2 based Bayesian C stock predictive model have tighter 95% credible interval widths than any of the Landsat-8 and Sentinel-2 based frequentist C stock models. Ultimately, cheap and readily available ancillary data from the Sentine-2 multispectral remote sensing mission, provides the most accurate C stock predictions when coupled with a Bayesian geostatistical hierarchical approach. This is despite the fact that the Bayesian approach is computationally expensive and conceptually more complicated to apply. The present chapter therefore establishes the Sentinel-2 based Bayesian hierarchical C stock predictive model as better than both its Sentinel-2 and Landsat-8 based frequentist geostatistical counterparts. As a way of applying and deriving

value from the best C stock predictive model, the next chapter evaluates how C stock distribution and density of the studied plantation forest ecosystem would be impacted under a warming climate in the projected future (2075) RCP8.5 based CO_2 emission scenario. The future RCP8.5 scenario reflects an 8.5 W m^{-2} radiative forcing change at the end of the 21st century as greenhouse gas emissions and population growth continue to increase and rise respectively.

CHAPTER SIX

Multispectral remote sensing-based hierarchical modelling of forest C stock under projected climate warming in Zimbabwe



This chapter is based on

Tsikai Solomon Chinembiri*, Onisimo Mutanga, Timothy Dube, "Multispectral remote sensing-based hierarchical modelling of forest C stock under projected climate warming in Zimbabwe: Under Review

Abstract

The threat of a warming and extreme future climate triggered by increased and unabated CO_2 emissions is likely to modify forest ecosystems and subsequently impact vulnerable communities relying on primary resources for livelihoods by the year 2100. The year 2000 land redistribution programme reduced buffer zones between ecologically sensitive forest ecosystems and the new crop of land reform beneficiaries in one of Zimbabwe's biggest reservoirs of forest biomass carbon. Given the threats and exigencies around plantation forest ecosystems, we sought to characterise how an ecologically and economically strategic plantation forest ecosystem in the eastern highlands of Zimbabwe would appear under a warming climate pathway. We modelled and predicted the distribution and density of forest carbon (C) stock using the revised Coupled Model Inter-comparison Project Phase 6 (CMIP6) of the Intergovernmental Panel on Climate Change (IPCC) using the current (1970-2000) and the business-as-usual SSP5-8.5 climate scenarios. We employed a hierarchical Bayesian geostatistical approach using the 1970-2000 as the baseline scenario and the RCP8.5 2075 future climate pathway to predict forest C stock distribution and density in the studied region. The mean of the credible C stock values for the baseline scenario with a 95% Highest Density Interval (HDI) are from 92.6 $MgCha^{-1}$ to 96.7 $MgCha^{-1}$ whilst the 95% HDI for the future RCP8.5 (2075) climate pathway is from 54.8 $MgCha^{-1}$ to 58.3 $MgCha^{-1}$. We deduce the posterior predictive C stock range under the current (1970-2000) ($1 \leq MgCha^{-1} \leq 133$) to be much higher than the C stock predictive range under the future 2075 ($2 \leq MgCha^{-1} \leq 82$) climate warming pathway. As the difference in posterior mean C stock ($\boldsymbol{\mu_1 - \mu_2}$), 38.3 $MgCha^{-1}$ is well above zero, we deduce that the posterior mean C stock distribution of the future RCP8.5 2075 climate projection is indeed credibly different from the current baseline (1970-2000) scenario. There is more than 90% probability that the forest plantations of the modelled C stock would be adversely affected by the RCP8.5 business-as-usual future climate warming projection. The studied forest plantation is projected to substantially shrink under the worst-case climate future scenario. The findings of this study are significant and appropriate for coming up with pragmatic climate change mitigation strategies, specifically, reforestation programs and meticulous choice of plantation tree species selection.

Key words

Climate warming, Highest Density Interval, hierarchical geostatistics, shared socio-economic pathway

6.1. Introduction

There is undeniable evidence of global climate warming in the current millennium and the warming is expected to continue and will partly be characterised by drastic changes in precipitation regimes (Van Amstel, 2006). At the time that the global temperature has been warming by approximately 0.6^0C ($\pm 0.2^0C$) throughout the 20th century (IPCC, 2006), higher temperature increases have impacted land masses over the same period. Some locations including the Alps exhibited a remarkably high warming trajectory where temperature increases reached as high as 1.7^0C in some regions (Rebetez *et al.*, 2006; Rebetez and Reinhard, 2008). It is therefore a daunting task to forecast how the climate might appear over the next 40-100 years, a span that is appropriate for forest management. Scientists and climatologists have been utilising an array of models that can generate probable climate futures and each model and each simulation run is regarded a prototype of how the climate progression throughout the 21st century would potentially look like (Gibbs *et al.*, 2007). As such, most planning will be based on forecasted trends together with their uncertainties. A series of the IPCC's assessment reports dating back to its First Assessment Report (FAR) in 1990 has been providing and updating evidence of the world's projected warming, together with possible mitigation and action plans for specific regions of the world (IPCC, 2014).

A follow up to the Fifth Assessment Report (FAR) by the Sixth AR of the IPCC shows stronger evidence of global warming in the last century and a more comprehensive sign of the impact of anthropogenic activities on climate warming (IPCC, 2013). Global climate warming poses a considerable impact on forest ecosystems that can result in human survival being threatened and pose permanent and serious challenges to the existence and development of human kind (Allen *et al.*, 2010). The forest ecosystem plays an important role on the terrestrial biosphere by promoting universal ecological adaptation and community succession (Fang *et al.*, 2001; Acharya, Maraseni and Cockfield, 2019). Forest biomass as the fundamental building block of plantation forest ecosystems can save as an indicator for the reciprocate relationship between forest energy flows and material circulation with the environment (Brown, 1997; Pan *et al.*, 2011). Consequently, any changes on forest biomass can signal the status and quality of the forest ecosystem together with the effects of anthropogenic activities, natural perturbation and climate change (Myneni *et al.*, 2001). Hence, local and regional carbon (C) stock estimates under a warming climate form the basis for research on global climate change and terrestrial ecosystem carbon cycle and offers a deeper understanding regarding the possible responses of ecosystems to greenhouse gas emissions (Dixon *et al.*, 1994). This is not only important for providing an understanding of global climate change dynamics at various spatial scales, but

also critical for managing forests in a sustainable manner and is key for reasonable use of forest resources needed for making the ecological environment better.

Sustainable and accurate prediction of forest C stock on a local and regional scale has historically been an important issue in climate change and climate adaptation studies. C stock estimation and prediction approaches can broadly be classified as field measurement and application of remote sensing in combination with plot-based field data (Lu *et al.*, 2016; Li and Li, 2019). Field based measurement data is appropriate for small-scale forest stands as it cannot be applied for C stock modelling at the regional scale due to the prohibitive costs of survey (Crosby *et al.*, 2017). Thus, field measured forest biomass data is more useful and informative when combined with remote sensing data. The contemporary remote sensing technology does not only provide multi-date and multi-scale data spanning large tracts of areas, but also makes it possible to make near real time monitoring of forest parameters for investigating and evaluating spatiotemporal trends of forest C stock (Lu, 2006). Many and various studies including Ali *et al.* (2015), Cao *et al.* (2016) and Li *et al.* (2020), have demonstrated that AGB is highly correlated with remote sensing derived data. As such, locally and regionally constructed C stock models can greatly be improved when combined with earth observation derived metrics. There are relatively cheap sources of remote sensing data available for use in forestry studies from the range of remote sensing technologies that include LiDAR, radar and optical remote sensing. The European Space Agency (ESA) based and freely available Sentinel-2 remote sensing data has proven useful and more accurate than its Landsat mission counterparts in recent AGB based studies (Meyer *et al.*, 2019; Chinembiri, Mutanga and Dube, 2023).

A climate projection study by Kumar, Sharma and Joshi (2016) shows important distribution patterns of important forest communities in the Himalayas using the maximum entropy model. Models involving species distribution modelling have largely been utilised for current (1950-2000), 2050 (approximate climate for the 2041-2060 projection) and the 2070s (average climate for 2061-2080) (Vieilledent *et al.*, 2016). Results have shown that by the year 2050, the habitat of approximately 17% of angiosperm endemic species would have shrunken or completely lost, while the percentage goes up to 20% by the year 2070 (Kumar, Sharma and Joshi, 2016). In another study by Zomer *et al.* (2014), projected climate change effects on the spatial distribution of ecoregions and bioclimatic strata was done using a transboundary approach in Nepal, India and China. Results from the climate change model demonstrated a substantial drift in climatic conditions with 54% drift to a different bioclimatic zone, 35% shift

to a different ecoregion and a 75% drift to a different stratum by the year 2050 (Zomer *et al.*, 2014).

The eastern highlands of Zimbabwe, specifically, Nyanga district, is habitat to some of the ecologically and economically important plantation forest species that are bordered by farmlands (Forestry-Commission, 2021). However, the same region of economically important timber plantations has not been spared by the year 2000 Fast Track Land Reform (FTLR) that redistributed most of the formerly owned commercial timber estates to indigenous farmers. Thus, some significant portions of formerly managed timber plantation pieces of land have been transformed and converted into agricultural enterprises by the indigenous farmers. In other cases, the indigenous farmers encroach into the commercial plantations as they fetch resources for grazing, tobacco curing and commercial timber for thatching (European-Commission, 2017). Such activities from the neighbouring communities have raised concern about the sustainability of forest development that the government is eagerly trying to resuscitate and seek investment as most forest plantations remain under threat of forest degradation. Unfortunately, some of the most successful small-scale indigenous tobacco farmers who benefited from the year 2000 land reform have been relying on firewood for tobacco curing illegally harvested from some of the established plantation forests (Business Times, 2023). Despite the European Union (EU) being the country's biggest market for the top foreign currency earning tobacco crop, the block has threatened to ban the importation of Zimbabwe grown tobacco into the EU market if the country does not transform its tobacco curing technologies to become environmentally sustainable within the next two years (Business Times, 2023). As a signatory to the United Nations Framework Convention on Climate Change (UNFCCC), Zimbabwe is also concerned about climate change and its subsequent impacts on vulnerable communities relying on primary resources for survival.

Given this background, it will be interesting to model and evaluate how one of the country's biggest reservoirs of forest biomass carbon looks like in the future under the revised Shared Socio-Economic Pathways (SSP5-8.5) under the Coupled Model Intercomparison Project Phase 6 (CMIP6) climate projections, presumed as the worst-case climate change scenario by 2100. This study therefore aims to enhance and update existing knowledge regarding C stock density and distribution patterns in managed plantation forest ecosystems of Zimbabwe under a possible and extreme future climate warming. The Governor's Office of Planning and Research (OPR) recommends the use of RCP8.5 for studies and analysis looking at impacts through 2050 since there are insignificant differences between probable future scenarios of greenhouse gas emissions throughout the first half of the century (Https://www.opr.ca.gov/,

2021). We therefore model and characterise the probable distribution and density of one of Zimbabwe's biggest reservoirs of forest biomass carbon in the eastern highlands. We achieve this using the current (1970-2000) and the RCP8.5 (2075) future scenarios by employing a Bayesian hierarchical geostatistical approach. Most studies modelling Above Ground Biomass (AGB) and other forest parameters under climate change including Deng *et al.* (2014), Li and Li (2019) and Nunes *et al.* (2020) have mainly relied on classical regression and machine learning techniques for characterising forest biomass distribution under climate warming. The current study therefore capitalises on the power of the hierarchical Bayesian approach in quantifying and providing an associated measure of uncertainty on how current (1970-2000) and future (2075) C stock predictions will differ. The future RCP8.5 scenario reflects an 8.5 W m^{-2} radiative forcing change at the end of the 21^{st} century as greenhouse gas emissions and population growth continue to increase and rise respectively. We therefore hypothesise a significant reduction in C stock density in the sampled region under a projected and warming climate scenario than the current (1970-2000) climate scenario.

6.2. Methodology

We sampled C stock data at Lot 75 A of Nyanga Downs in Nyanga district in Manicaland province of Zimbabwe. The sampled area has *Eucalyptus camaldulensis, Pinus patula* and *Eucalyptus grandis* as the most predominant plantation with its borders characterised by portions of agriculture enterprises and grazing activities (Forestry-Commission, 2021). The sampled area is found between latitude 32º40ˈ E and 32º54ˈ E and 18º10ˈ12ˈˈ S and 18º25ˈ4ˈˈ S longitude and is approximately 2 767 ha as shown in **Figure 6-1**. Grazing and gold panning are some of the activities that started developing after the FTLRP of 2000 and seem to be modifying the status quo of surrounding forest plantations (Zvobgo and Tsoka, 2021). Rainfall amounts of the area range from 740 mm to 2,996 mm and a mean annual precipitation of 1151 mm. The region experiences humid weather distinguished by extensive veld fires closer to high-altitude grassland areas from October to November. Annual mean temperature range between 10^0 C and 13^0 C and 24^0 C and 27^0 C are experienced throughout the winter and summer seasons, respectively (Initiative For Climate Action Transparency (ICAT), 2022).

6.3. Remote sensing data

We made use of the Sentinel-2 and WorldClim (Fick and Hijmans, 2017) data to derive predictors for C stock in a managed plantation forest ecosystem in Manicaland province of Zimbabwe. We acquired the Sentinel-2 imagery covering the entire study area (lot 75A Nyanga

Downs in the eastern highlands of Zimbabwe) on the 20th of September 2020. Sentinel-2 data was obtained as level-1C 12-bit pre-set TOA reflectance values.

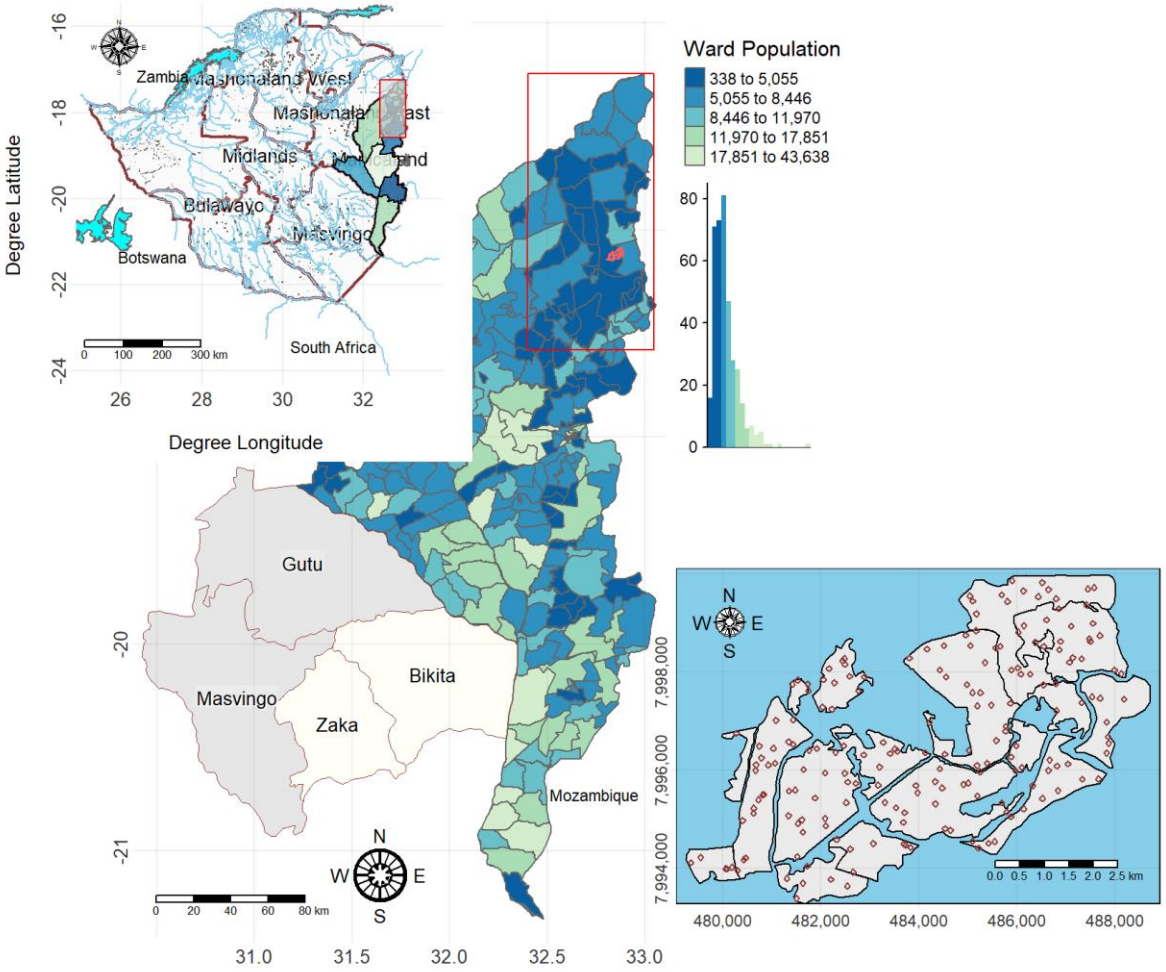


Figure 6-1: Map of the study area indicating (a) the province where samples were derived, (b) study area location within the sampled province and (c) the spatial distribution of the sampled regionalised variable in the lower panel.

The Sentinel-2 1-C product was pre-processed and orthorectified in the *R Statistical Environment* using the *sen2r* library (Ranghetti *et al.*, 2020). The Sentinel-2 satellite platform with a geolocation error of 12.5 m, profiles data in 13 spectral bands with four in the visible range with 10 m spatial resolution, six in the near-infrared region with 20 m spatial resolution and three bands in the short-wave infrared region with a 60 m spatial resolution (Gerald *et al.*, 2017). The same scene can therefore be recorded by a Sentinel-2 satellite sensor at varying spectral bandwidth, spatial resolutions and spectral response function (Zhang et al., 2018). All the orthorectified visible and infrared bands of Sentinel-2 were stacked, mosaicked and clipped in order to come up with a subset raster of the sampled area.

6.3.1. Climate data

Mean annual temperature (MAT), temperature seasonality (TS) and mean annual precipitation (MAP) from the Model Inter-comparison Project Sixth Assessment Report (CMIP6) of the Inter-governmental Panel on Climate Change (IPCC) were acquired from the WorldClim database (Fick and Hijmans, 2017). These variables were selected as they are biologically relevant and therefore correlated with climatic adaptation and spatial distribution of flora (Rodima-Taylor, Olwig and Chhetri, 2012). Again, past studies including MALHI *et al.* (2006) and Liu *et al.* (2014) have demonstrated a significant relationship between MAP and MAT to AGB distribution. Raster data for the climate variables were obtained at 30 arc seconds (~ 1 km) spatial resolution (Wickham *et al.*, 2015). We acquired the bioclimatic variables from the WorldClim data repository (https://www.worldclim.org/) We resampled all the datasets to a common 1 km^2 spatial resolution using the nearest neighbour interpolation technique (Christman and Rogan, 2012).

6.3.2. Vegetation indices

The Soil Adjusted Vegetation Index ($SAVI$), the Enhanced Vegetation Index (EVI) and the Normalised Diference Vegetation Index ($NDVI$) were derived from the Sentinel-2 multispectral remote sensing platform (Ahmed, Atzberger and Zewdie, 2022). Robust comparison of the potential of vegetation indices derived from the new generation remote sensing platforms of Landsat-8 and the Sentinel-2 have shown Sentinel-2 outperforming the Landsat-8 derived sensor (Semela, Ramoelo and Adelabu, 2020; Chinembiri, Mutanga and Dube, 2023). The widely known relationships and correlations between AGB and vegetation indices influenced our decision to incorporate them into the C stock modelling equation (Li and Li, 2019; Bordoloi *et al.*, 2022).

6.3.3. Anthropogenic variables

The distance to settlements ($DIST$) in the sampled study area was employed as a proxy variable indicating accessibility to forest resources by villagers living within and around plantation forests. All settlements were directly digitized from the revised 1:50 000 topographic maps by the department of the Surveyor General, Harare (Surveyor General, 2018) Nearest distances to settlements were then calculated from each of the sampled and georeferenced C stock plots for the Sentinel-2 satellite raster data in a GIS environment. The necessity for modelling AGB with distance to settlements is premised on the principle that ease of access to forest resources puts pressure on the forest health and condition. This then increases the probability of direct and indirect contact with established plantations (Mon *et al.*, 2012; Chinembiri *et al.*, 2013).

6.4. Climate change-based forecasting of C stock distribution

The IPCC through its Fifth Assessment Report (AR5) shows how the average temperature of the earth's surface has been increasing over the past century (1880-2012), where most of the temperature rise has happened in the last sixty years (1950-2010) (Stocker, 2013). Four RCPs are discussed under the IPCC's AR5, namely, RCP2.6, RCP4.5, RCP6.0 and RCP8.5 where the assigned digits denote the amount of radiative forcing in watts per square meter (wm^{-2}) in the year 2100 relative to the 1750 value (Vieilledent *et al.*, 2016). We elected to use the most typical current (1970-2000) and the highly concentrated RCP8.5 future trajectory pathway under the revised SSP5-8.5 of the CMIP6 for the 2075. The RCP8.5 representative concentration projection represents the worst case climate change scenario where the population and CO_2 gas emissions will continue on an increasing trajectory throughout the 21st century (van Vuuren and Carter, 2014). Further, RCP8.5 represents elevated levels of greenhouse gas emissions and energy demand, leading to an 8.5 Wm^{-2} radiative forcing by the year 2100 (Moss *et al.*, 2010). We did not utilise the other Shared Socio-Economic Pathways (SSPs) as trends for the studied region are not expected to diverge dramatically until the end of the century (IPCC, 2013). SSPs are the third generation of the climate change narrative and therefore, more preferable to Special Report on Emission Scenarios (SRES) to allow better flexibility and low costs in the modelling process (van Vuuren and Carter, 2014).

6.5. Above ground biomass measurement and derivation of C stock

We sampled the forest plantation study area through measurement of all the trees with at least 10 cm diameter at breast height (DBH) (~ 1.3 m) using circular 500 m^2 supports from the 19th of September 2021 to the 24th of October 2021. Gibbs et al. (2007) recommend measurement of at least 10 cm DBH trees as those with lower than this DBH store insignificant amount of C stock. We utilised sampling locations by employing uniform coverage sampling using the *spcosa* R library for statistical computing where 200 sample locations were obtained (Brus, de Gruijter and van Groenigen, 2006). We pre-uploaded the 200 samples into a 72 H Garmin handheld GPS with a 5-metre accuracy before starting the field data collection programme. The actual field data collection programme resulted in 191 sample observations being collected as the other nine samples fell outside the defined sampling area.

There are three dominant plantation forest species found in the studied region which are; (i) *Pinus patula*; (ii) *Eucalyptus grandis* and (iii) *Eucalyptus camaldulensis*. Biomass of each of the three tree species types was calculated using specific allometric equations. We applied the allometric equations used by Brown (1997) for calculating the biomass of the *Pinus patula*

species whilst we derived the tree biomass for the *Eucalyptus* species using Zunguze (2012) allometric equations. Similar allometric equations were also used for the same tree species in the derivation biomass for the *Pinus* and the *Eucalyptus* tree species in Manica Province in Mozambique. Manica province of Mozambique strongly resembles the climatic conditions of the sampled area as they have a common border in the eastern highlands of Zimbabwe. AGB was then derived for each and every individual tree species to C stock using the IPCC (2006) conversion factor. Each per plot value estimate was then standardised to a $MgCha^{-1}$ measurement unit.

6.6. C stock prediction modelling

We made use of the hierarchical geostatistical model fitted using predictors derived from vegetation indices and anthropogenic variables derived from the Sentinel-2 multispectral remote sensing data and bioclimatic variables derived from the WorldClim (http://www.worldclim.org/). We made climate projections for the following bioclimatic variables: $bio1$=mean annual temperature, $bio4$=temperature seasonality and $bio12$=annual precipitation. Since changes due to climatic factors were our area of interest, we assumed frozen land use change between the baseline scenario and the RCP8.5 (2070s) climate projection. We also made assumptions of no changes in the values of the anthropogenic variables and vegetation indices in the projected future.

We therefore accomplished C stock predictions through updating of the MAP, temperature seasonality and the MAT climate vectors for the 2075 (RCP8.5) scenario while controlling for the underlying spatial processes and holding the non-climatic vectors constant (Vieilledent *et al.*, 2016). Our C stock prediction model therefore only showed the marginal effects of climate change on C stock distribution through changes coming as a result of the climatic variables. We assumed the sampled C stock variable to be normally distributed with a data model that can be expressed as:

$\boldsymbol{Z}|\boldsymbol{y},\boldsymbol{\Sigma} \sim N(\boldsymbol{K}_{\boldsymbol{y}},\boldsymbol{\Sigma})$, ---------- (6-1)

Where;

$\boldsymbol{Z}$ denotes measurements from an underlying process, $\boldsymbol{y}$ with an associated measurement error process

$\boldsymbol{K}$ denotes a matrix mapping the observed data to process locations

The general kriging model for the observed response variable adopts the general kriging equation of the form:

$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\eta}$ where $\boldsymbol{\eta} \sim N(\boldsymbol{0},\boldsymbol{\Sigma})$ ---------- (6-2)

where;

$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \end{bmatrix}$, $\boldsymbol{y_1} = (y_{s_1}, \ldots, y_{s_n})'$ represents the spatial process at latitude-longitude locations with data

$\boldsymbol{y_2}$ denotes the process at unsampled locations (e.g., s_0) or location

$\boldsymbol{X\beta}(=\boldsymbol{\mu})$ denotes the mean of the spatial process explained by predictors and spatially correlated error η with covariance;

$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$

Given our prediction problem, kriging involves finding the best linear predictors of $\boldsymbol{y}_2$ given $\boldsymbol{y}_1$:

$\boldsymbol{y}_2|\boldsymbol{y}_1 \sim N(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}{\boldsymbol{\Sigma}_{11}}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}, {\boldsymbol{\Sigma}_{11}}^{-1}\boldsymbol{\Sigma}_{12})$,

Where;

$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{bmatrix}$, $E(\boldsymbol{y}_1) = \boldsymbol{\mu}_1$ and $E(\boldsymbol{y}_2) = \boldsymbol{\mu}_2$

From the universal kriging model shown in Equation. 6-2, we then formulated the hierarchical representation of this model with parameters defined as before as follows:

$\boldsymbol{Z}|\boldsymbol{y}, \boldsymbol{\sigma}^2 \sim N(\boldsymbol{K}y, \boldsymbol{\sigma}^2\boldsymbol{I})$ ---------- (6-3)

$\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}_y \sim N(\boldsymbol{X\beta}, \boldsymbol{\Sigma}_y)$ ---------- (6-4)

$\boldsymbol{\beta}|\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_\beta \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_\beta)$ ---------- (6-5)

$\boldsymbol{\sigma}^2, \boldsymbol{\theta}_y, \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_\beta \sim [\boldsymbol{\sigma}^2, \boldsymbol{\theta}_y, \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_y, \boldsymbol{\Sigma}_\beta]$ ---------- (6-6)

Under the hierarchical specification shown in Equation. 6-3 to Equation. 6-6, the full Bayesian kriging for the predicted C stock gets the predictive distribution of $Y(\boldsymbol{s}_0)$ given $\boldsymbol{x}(\boldsymbol{s}_0)$ along with $\boldsymbol{Y}$ and X. That is,

$p(Y(\boldsymbol{s}_0)|\boldsymbol{x}(\boldsymbol{s}_0), \boldsymbol{Y}, X) = \int p(Y(\boldsymbol{s}_0)|\boldsymbol{x}(\boldsymbol{s}_0), \boldsymbol{Y}, \boldsymbol{\theta}) p(\theta|\boldsymbol{Y}, X) d\boldsymbol{\theta}$ ------- (6-7)

where;

$p(Y(\boldsymbol{s}_0)|\boldsymbol{x}(\boldsymbol{s}_0), \boldsymbol{Y}, \boldsymbol{\theta}) p(\theta|\boldsymbol{Y}, X)$ is a normally distributed process arising from the joint multivariate normal distribution of $Y(s_0)$ and the original data Y. We therefore specified priors for the modelled C stock variable using the spatial model, the independent error model and the spatial intercept model as:

$p(\boldsymbol{\theta}) = \boldsymbol{Unif}(\phi|0.39,0.0013)x\boldsymbol{IG}(\sigma^2|0.053,0.0029)x\boldsymbol{IG}(\tau^2|0.1,1.53)x\boldsymbol{MVN}(\beta|\boldsymbol{0}, \Sigma_\beta)$.

We used a k −fold validation algorithm in order to assess the predictive performance of each of the three different types of models fit to the data through random splitting of the 191 measured C stock data into ten equally sized segments (Duchene et al., 2016). We calculated Deviance Information Criteria (DIC), the Root Mean Square Error ($RMSE$) and the Mean Absolute Error (MAE) as the major metrics for evaluating the models' ability to fit the

measured data (Spiegelhalter et al., 2002). Better fitting models were judged using the DIC, $RMSE$ and MAE (Jackman, 2000; Green, Finley and Strawderman, 2020).

6.6.1. Bayesian paired t-test

We compared the posterior mean C stock predictions under the current (1970-2000) baseline scenario and the posterior mean C stock predictions for the future RCP8.5 projected climate scenario using the Bayesian paired t-test (Jeffreys, 1961). By placing prior probabilities, π_0 and π_1 on hypothesis H_0 and H_1, we updated their values through the Bayes' theorem in order to obtain posterior probabilities as follows:

$$P(H_j|data) = \frac{\pi_j P(data|H_j)}{\pi_0 P(data|H_0) + \pi_1 P(data|H_1)} j = 0,1 \text{ ---------- (6-8)}$$

Where;

$\pi_0 + \pi_1 = 1$

$(data|H_j)$ indicates the marginal density of the sampled C stock data under hypothesis *j*

Because posterior probabilities are sensitive priors π_0 and π_1, we employed the Bayes Factor (BF) as follows:

$BF = \frac{P(data|H_0)}{P(data|H_1)}$ and in cases where the $BF > 1$, the data presents evidence for H_0 and where $BF < 1$, the data presents evidence for H_1 (and against H_0). As suggested by Jeffreys (1961), $BF < 0.1$ presents 'strong', evidence against H_0 whilst $BF < 0.01$ presents 'decisive' evidence. Since the posterior probability of the modelled C stock variable is related to the BF, we calculated it as:

$P(H_0|data) = \left[1 + \frac{\pi_1}{\pi_0}\frac{1}{BF}\right]$.

6.7. Results

This section provides specific information on the results of the study.

6.7.1. Posterior predictions of mean C stock under the current and the future climate scenarios

The posterior means of the two climate projections, that is, the current (1970-2000) and the future (2075) are credibly different. The mean and standard deviation of the current climate scenario are credibly larger than the mean and standard deviation of the future SSP5-8.5 (RCP8.5) climate pathway. On the other hand, variability within individuals found in the studied ecosystem could also affect other species in a favorable manner. The mean of the credible C stock values for the current climate scenario (**Table 6-1**) with a 95% Highest Density Interval (HDI) is from 92.6 $MgCha^{-1}$ to 96.7 $MgCha^{-1}$ whilst the 95% HDI for the

future climate projection is from 54.8 $MgCha^{-1}$ to 58.3 $MgCha^{-1}$ (**Table 6-1**). The mean C stock of the future projection climate scenario Markov Chain Monte Carlo (MCMC) is therefore 58.1 $MgCha^{-1}$ with a 95% HDI from 54.8 $MhCha^{-1}$ to 58.3 $MgCha^{-1}$. The effective sample sizes for the parameters derived from the climatic, vegetation indices and anthropogenic variables present stable estimates of the 95% credible intervals as they are well over 10000. In the same vein, the *Rhat* values (**Table 6-1**) of the modeled variable indicate strong convergence of the posterior predictive distribution as most of the values are 1.

Table 6-1: Posterior distributions of the derived parameters for the current (1970-2000) and the RCP (2075) climate scenarios. HDIIo and HDIIup are the lower and upper limits of a 95% Highest Density Interval (HDI). Rhat is the scale reduction factor

Parameter	Mean	Median	HDI%	HDIIo	HDIIup	Rhat	Effective sample size
$\mu_1(1970-2000)$ $MgCha^{-1}$	96.4	96.1	80	92.6	96.7	1	41871
$\mu_2(Future, 2075)$ $MgCha^{-1}$	58.1	56.2	80	54.8	58.3	1	43370
$(\mu_1 - \mu_2)$ $MgCha^{-1}$	38.3	39.9	80	37.8	38.4	-	-
$\sigma_1(1970-2000)$ $MgCha^{-1}$	8.2	8.2	80	7.9	8.5	1	34004
$\sigma_2(Future, 2075)$ $MgCha^{-1}$	3.5	3.5	80	3.6	3.6	1	32834
$(\sigma_1 - \sigma_2)$ $MgCha^{-1}$	4.7	4.7	80	4.2	4.2	-	-

6.7.2. Fixed effects and latent spatial processes

Fitting of the Bayesian hierarchical C stock model with bioclimatic and vegetation indices derived variables for both the current and the future climate scenarios produced fixed and random effects model parameters illustrated in **Table 6-2**. Fixed effects from the vegetation indices and *DIST* covariables are significant predictors of the modelled C stock model as evidenced by their 95% credible intervals, excluding zero **Table 6-2**). Despite their influence being kept constant in the future RCP8.5 climate scenario modelling, the influence of the fixed effects is relatively weak in the posterior predictions for the 2075.

All the three bioclimatic variables, that is, the MAP, the MAT and temperature seasonality have no significant influence on C stock under the baseline climate scenario (1970-2000). However, their effect is significant in the projected RCP8.5 climate projections as all their 95% posterior credible intervals exclude zero. Both temperature seasonality and the MAP are predicted to have a positive effect on C stock distribution under the RCP8.5 radiative forcing climate scenario. Thus, forest plantations are expected to benefit from increased photosynthetic

activity due to higher rainfall amounts and more greenhouse gas (CO_2) emitted under the projected climate warming.

Table 6-2: Relationship between C stock, current and future climate projection derived fixed and random effects

Parameter	Current (1970-2000)				Projected RCP8.5 (2070s)			
	Mean	**s.d**	**2.5%**	**97.5%**	**Mean**	**s.d**	**2.5%**	**97.5%**
Intercept	-2.22	0.44	-3.09	-1.37	-3.08	1.44	-4.09	-1.39
NDVI	4.90	0.25	4.40	5.38	3.97	0.28	4.10	4.68
SAVI	0.99	0.38	0.22	1.74	0.86	0.68	0.32	1.94
DIST	0.85	0.14	0.58	1.12	0.75	0.16	0.48	1.32
Bio1	-0.0007	0.001	-0.0027	0.0013	-0.0003	0.009	-0.0023	-0.0014
Bio4	0.0001	0.0001	-0.0002	0.0001	0.031	0.007	-0.021	0.036
Bio12	0.0001	0.0001	-0.0001	0.0003	0.0007	0.004	0.004	0.009
σ_w^2	0.0610	0.0102	0.0417	0.0808	0.0310	0.0302	0.0217	0.0508
σ_ε^2	0.0054	0.0023	0.0019	0.0099	0.0064	0.0043	0.0029	0.0190
$\phi(m)$	2000	1200	1500	2500	1700	1600	1300	2000

The spatial structure of predicted C stock is expected to weaken under the RCP8.5 climate pathway as evidenced from the results illustrated in **Table 6-2**. The spatial dependence of the modelled variable will be reduced from its current mean 95% posterior credible interval of 2000 m to about 1700 m whilst spatially structured variance is expected to weaken from a $0.042 \leq \boldsymbol{\sigma_w^2} \leq 0.081$ 95% credible interval to a $0.022 \leq \boldsymbol{\sigma_w^2} \leq 0.051$ 95% posterior credible interval (**Table 6-2**).

6.7.3. C stock predictions under a warming climate scenario

Projected changes are particularly striking in the southern and northern parts of the studied region in the eastern highlands of Zimbabwe (**Figure 6-2**). Significant reduction of C stock concentration is noticeable within these regions under the higher emission scenario 8.5-5SSP. The C stock posterior predictive range under the current (1970-2000) ($1 \leq \text{MgCha}^{-1} \leq 133$) is much higher than the posterior C stock predictive range under the future ($2 \leq M\text{gCha}^{-1} \leq 82$) climate warming scenario. Thus, future C stock predictions varied between ($2 \leq \text{MgCha}^{-1} \leq 82$), corresponding to 50% and 38%. Hence, extrapolating predictions to the year 2075 under the RCP8.5 climate scenario shows that forest C stock distribution will shrink significantly, with more than 40.2% of C stock lost due to climate extremes (**Figure 6-2**). The overall spatial distribution trends of AGB for climate scenarios in the baseline and 2075 were similar to the current distribution of AGB, that is, high AGB was predominantly distributed in areas of high altitudes and less human activity, whereas low AGB was predominantly distributed in the valleys and plains with low altitude and frequent human activities in the central areas of the study area.

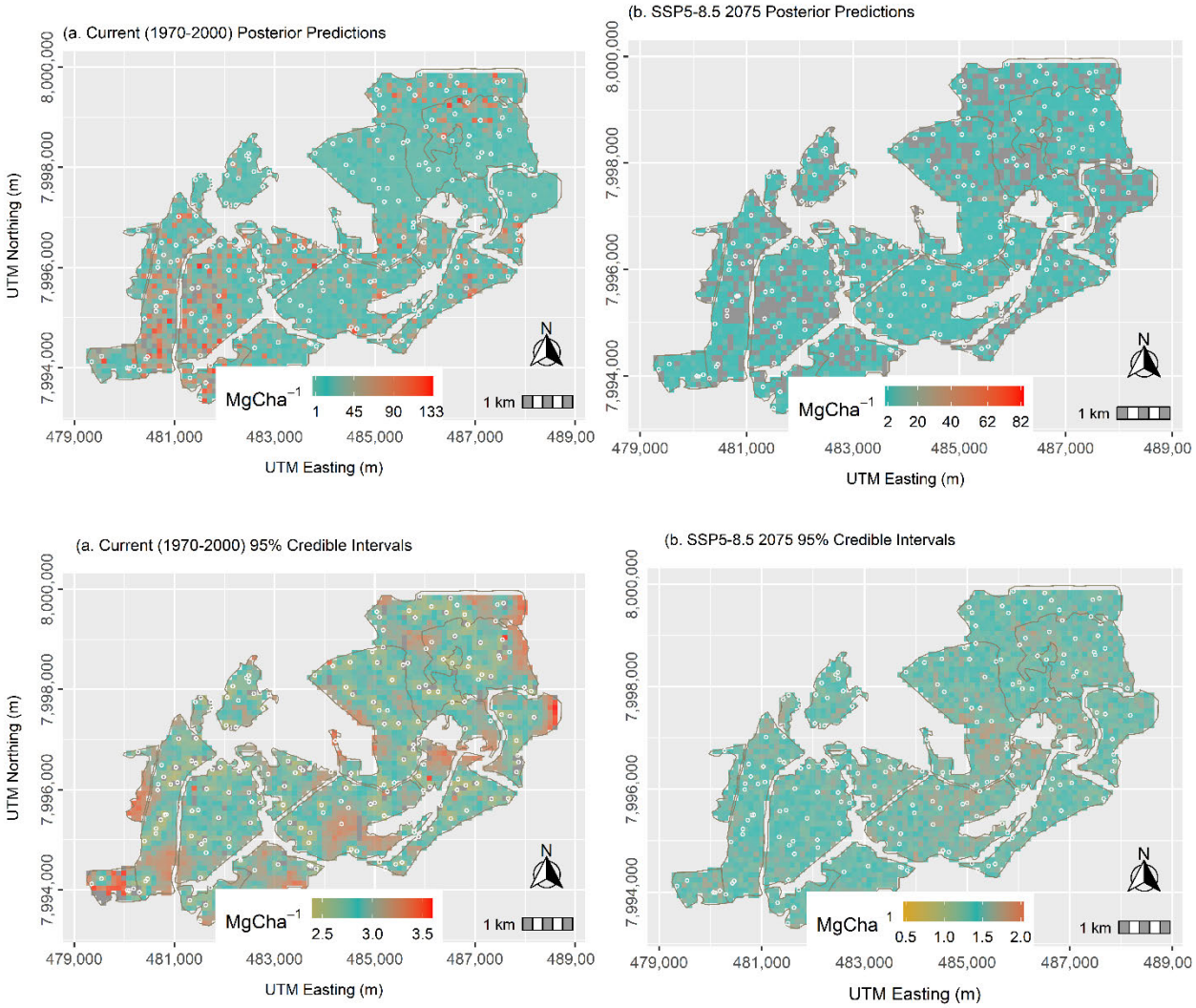


Figure 6-2: Climate projection scenarios for the modelled C stock. The upper panel indicates posterior predictions of the current and the projected climate warming under the RCP8.5 scenario. The lower panel maps indicate the posterior standard deviation (95% CI) for the predicted maps in the upper panel. Points represent locations of sampled C stock.

The general spatial distribution trends of the predicted C stock for the current climate scenario are thus quite different to the spatial distribution of predicted C stock in the future, 2075 climate scenario. The northern and southern extremes of the studied region have relatively high C stock than in the projected future climate pathway. Predicted C stock under the projected RCP5.8 climate change pathway has higher uncertainty ($0.5 \leq MgCha^{-1} \leq 2.0$) than predicted C stock under the current (1970-2000) ($2.5 \leq MgCha^{-1} \leq 3.5$) baseline scenario (**Figure 6-2**). However, patches of modelled C stock uncertainty are widespread in the projected future climate scenario with substantially higher CO_2 greenhouse gas emissions than in the baseline period.

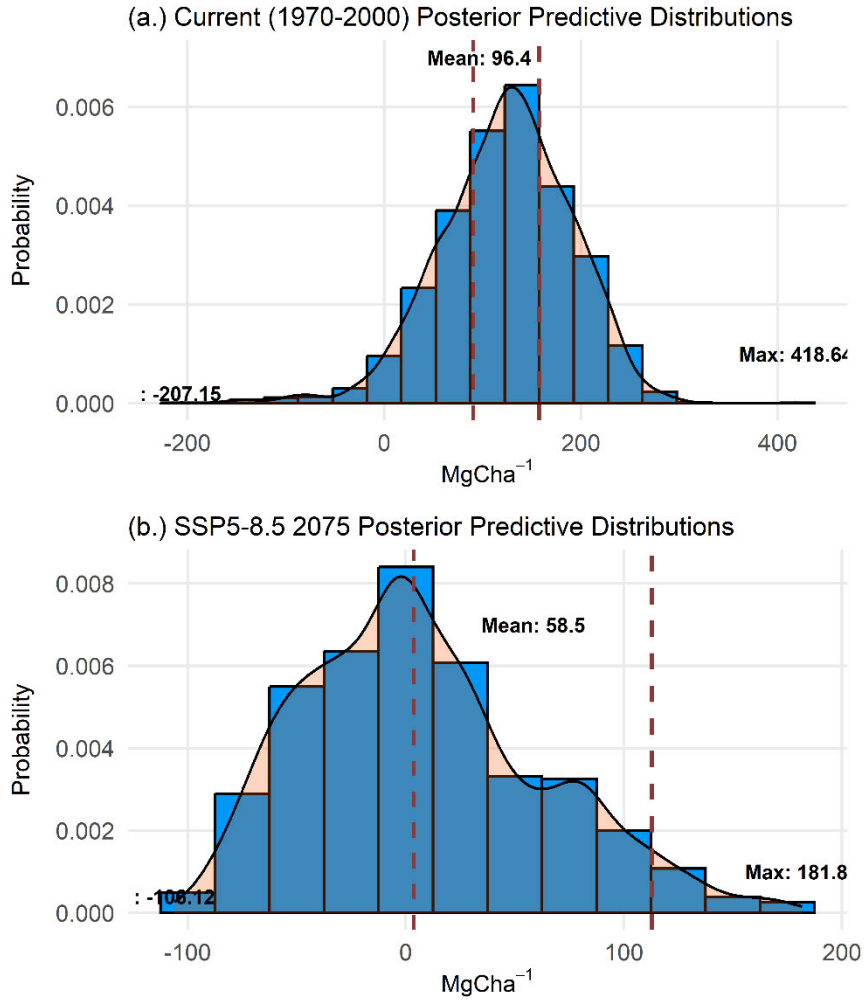


Figure 6-3: C stock posterior predictive distributions under (a). the current (1970-2000) scenario and (b). the projected (RCP8.5) climate pathway

6.7.4. Posterior mean comparison between the current and the projected climate scenarios

A posterior distribution of the difference in means of the current (1970-2000) and the projected RCP8.5 histogram is illustrated in **Figure 6-3**. The C stock predictions for the current climate scenario is evidently negatively skewed with a posterior mean C stock estimate of 96 MgCha^{-1} whilst its projected RCP8.5 for the future climate scenario is positively skewed with a posterior C stock mean estimate of 58.1 MgCha^{-1}. A Bayesian comparison of the current climate scenario posterior mean and the future climate under the RCP8.5 2075 scenario points to a significant difference in the predicted C stock for the two scenarios. There is therefore a 90% probability that the difference between the projected scenario (2075) and the current baseline (1970-2000) is less than zero (**Figure 6-3**).

Hence, the difference ($\boldsymbol{\mu_1 - \mu_2}$), 38.3 MgCha^{-1}, is well above zero, we deduce that the mean C stock distribution of the two climate projection scenarios are indeed credibly different.

Again, not only the mean of the current climate scenario (1970-2000) is credibly larger than the mean of the future climate projections but the posterior standard of the current scenario is also credibly larger than the posterior standard deviation of the projected RCP8.5 future scenario (**Table 6-1**).

6.7.5. Model validation under climate change scenarios

Table 3 presents validation statistics for the three hierarchical Bayesian models utilised in the prediction of C stock using the current and the future RCP8.5 projected climate pathways. We selected the spatial model for both the current and the RCP8.5 projected climate scenarios as it has the lowest RMSE and widest coverage (CVG) compared to the independent error model and the spatial intercept only models.

Table 6-3: Current and RCP8.5 (2075) projected C stock prediction validation

Model Validation criterion	Current (1970-2000) Scenario			Future (RCP8.5) Scenario		
	Independent Error Model	**Spatial Intercept only Model**	**Spatial Model**	**Independent Error Model**	**Spatial Intercept only Model**	**Spatial Model**
RMSE ($MgCha^{-1}$)	3.06	2.45	2.35	2.94	3.79	2.68
MAE ($MgCha^{-1}$)	2.36	2.43	2.55	2.22	2.97	2.13
CRPS ($MgCha^{-1}$)	2.01	2.49	2.52	2.20	2.31	2.06
CVG (%)	92.7	86.4	95.4	94.3	88.5	90.8
DIC	211.2	57.9	44.1	-245.4	38.8	-555.7

However, the spatial model for the future RCP8.5 climate pathway shows marginally wider variability in performance coupled with low coverage (90.8%) compared to the current (1970-2000) baseline scenario (95.4%) (**Table 6-3**). Hence, the predictive C stock spatial model generated from the current scenario ($RMSE = 2.35$) is more accurate than the predictive model generated by the future climate scenario ($RMSE = 2.68$).

6.8. Discussion

This section discusses the results of the study with regards to the performance of bioclimatic variables in influencing C stock prediction under a projected climate warming scenario.

6.8.1. The influence of bioclimatic vectors on climate change driven C stock prediction

The IPCC proposed climate scenarios for the current and future period encompass a future climate in which there will be more greenhouse gases in the atmosphere due to rising energy demand and population growth by the year 2100. This study evaluated the effect of the worst-

case climate scenario under the RCP8.5 climate scenario on the distribution and density of C stock in a managed forest ecosystem in the eastern highlands of Zimbabwe. Temperature and precipitation climate vectors are expected to exert the biggest influence on the distribution of flora and fauna under the projected climate warming scenarios

Vegetation indices ($NDVI$ and $SAVI$) employed in the Bayesian modelling under the current and the projected climate change warming have a positive effect on C stock prediction. Slope based vegetation indices are commonly correlated with AGB and the results of Li *et al.* (2020) and Ahmed, Atzberger and Zewdie (2022) corroborate with our findings. However, the DIST independent variable has a negative effect on predicted C stock (**Table 6-2**). This is an indication that easily accessible forest plantation forest areas are likely to be deforested more than forests that are inaccessible to settlements. Accessibility also relates to the availability of easy roads to use for transportation of logged timber by communities in the neighbourhood of plantation forest and the terrain of the forest areas. The current study is located in a gently sloping landscape, making it more accessible and navigable by closer communicates, either for firewood or for timber harvesting for commercial use. Similar to the present findings, studies establishing a correlation between anthropogenic variables and AGB, including Chinembiri et. al (2013) have also been undertaken in sub-tropical biomes.

All the three bioclimatic vectors (MAT, MAP and temperature seasonality) utilised in the predictive model are not statistically significant in C stock prediction under the current (1970-2000) baseline climate scenario (**Table 6-2**). However, MAT and MAP have a statistically significant effect on C stock prediction under the future RCP8.5 climate pathway. The RCP8.5 climate change scenario represents a high climate change future where CO_2 emissions and population will continue to increase throughout the 21st century (van Vuuren *et al.*, 2011). An inverse relationship exists between mean annual temperature, mean annual precipitation and C stock in the projected 2075 probably because of a projected increase in precipitation. However, the increase in precipitation is predicted to have a negative feedback effect on vegetation growth as vegetation is likely to reach maturity and mortality faster than before. Again, a warming climate could invite other disturbances for plantation forest like disease due to warmer and wetter winters. The work of Chave et al. (2014) have demonstrated that temperature seasonality (a proxy for the growing season length) was the most important bioclimatic variable in explaining tree height and diameter relationship and subsequently, tree biomass. The present study established no effect between temperature seasonality and C stock, thus differing with literature results obtained by Chave et al. (2014).

In a broader study using data from 271 observation plots from tropical and temperate forests in South, Central and North America, Stegen et al. (2011) established weak relationships between climate and forest biomass. There were no significant relationships established between mean annual temperature and forest biomass across forest types and a weak, although significant, positive relationship was established between annual precipitation and forest biomass. In another study utilising biomass data from 137 plots from the world's highly carbon-dense prime forests, Keith, Mackey & Lindenmayer (2014) reached a conclusion that the maximum biomass carbon density is found in moderately cool and wet climates in temperate moist forest ecosystems. Results obtained by Keith, Mackey & Lindenmayer (2009) and Stegen et al. (2011) are then congruent to the results obtained by Chave et al. (2014) pertaining to the relationship between climate and tree biomass in sub-tropical forests. However, these results are at variance with the findings of our study. Hence, more work is still needed in order to have a better and precise understanding of the effect of climate on forest carbon storage across forest types (Keith, Mackey & Lindenmayer 2009; Stegen et al. 2011; Koven, Riley and Stern, 2013).

Because of the difficulty of interpreting the spatial decay parameter directly, we present the effective spatial range of the modelled C stock variable in meters in the results illustrated in **Table 6-2**. The effective range for the predicted C stock variable for the current climate scenario (~ 2000 m) is higher than the effective range in the projected future climate scenario (~ 1700 m). As such, there is reduced spatial dependence in the future RCP8.5 driven C stock predictions with effective spatial short-range dependence of ~ 1700 (with a 95% credible interval from 1300 to 2000 m). This implies that the effect of MAP and MAT in the 2070s will be effective in influencing both the predictions and distributions of C stock in the studied region. Effective spatial short-range dependence in the projected climate pathway is caused by effective bioclimatic MAP and MAT climate vectors. Increased greenhouse gas emissions under the RCP8.5 climate scenario would impact plantation forest negatively, making them sparse and highly unpredictable. The fact that certain plantation forest species will be adversely affected in the projected future (2070s) under the worst case 8.5RCP climate pathway and its subsequent effects on C stock distribution variability has practical precedents (Lazarus and Eriksen, 1952). The estimation of effective spatial range aid in understanding how far-reaching spatially structured dependence is in the modelled C stock after accounting for influential bioclimatic vectors under a climate warming scenario (Babcock *et al.*, 2018).

6.8.2. Current and future RCP8.5 C stock predictions

A number of studies have shown the current and projected impacts of climate on forest density, plant phenology forest productivity, biodiversity and the function of ecosystems as CO_2 sinks (Seo and Song, 2019; Mourguiart *et al.*, 2023). Coordinated and integrated energy and material exchanges occur between forests and the climate, which restrict and affect each through negative feedback mechanisms (BOISVENUE and RUNNING, 2006; Nunes *et al.*, 2020). Consequently, forest ecosystems as the biggest carbon source and sink, will respond to changes in climate and demonstrate feedbacks and regulation effects with regard to climate change (Richardson *et al.*, 2013; Mitchard, 2018). Forest types and environmental factors are prime in driving the magnitude and direction of the impacts of climate change on forest ecosystems (Medlyn, Duursma and Zeppel, 2011; Yao, Piao and Wang, 2018). The productivity of forest ecosystems typically increases as a result of extended growth periods emanating from increasing temperatures and rapid fertilization caused by increasing CO_2 concentrations. However, some studies have revealed contrasting results, showing climate change to have a negative effect on forest ecosystem productivity (Alig, Adams and McCarl, 2002; Mitchard, 2018).

The mean and standard deviation of the current climate scenario are credibly larger than the mean and standard deviation of the future SSP5-8.5 (RCP8.5) climate pathway. This implies that the future climate through the 8.5 Wm^{-2} radiative forcing of elevated greenhouse gas emissions and energy demand will increase variability in C stock distribution across the forest plantations. This could signal that some forest species may be adversely affected by the radiative forcing high greenhouse gas emissions and high energy demands. On the other hand, variability within individuals found in the studied ecosystem could also affect other species in a favorable manner. The Bayesian geostatistical model in this study estimates an average loss of C stock of approximately 40% due to the projected climate warming scenario in the 2070s. Precipitation and temperature will have far-reaching impacts on C stock distribution in the plantation forest ecosystems of Zimbabwe as precipitation is generally anticipated to increase under the projected climate change scenarios. This would result in more plantation forest species being negatively impacted due to modified vegetation phenology triggered by increased precipitation and CO_2 concentrations in the atmosphere.

Human activities have been recognised as one of the major drivers of forest ecosystem change with a rapid and drastic impact to forest health (Rudel *et al.*, 2005; Titeux *et al.*, 2016). In fact, the IPCC has demonstrated how climate change is closely related to anthropogenic activities

by highlighting the exceptional rate and magnitude of the impact of human activities on environmental change over the last century (IPCC, 2013). Human activities can result in a variety of forest management actions that can either slow down or worsen climate change impacts (Karl and Trenberth, 2003; Wang, Chen and Dong, 2006). The formulation of targeted forest ecosystem management plans is also dependent on the quantification and prediction of C stock distribution under anthropogenic influence and a warning projected climate. The present study predicted C stock density and distribution under the current and the future RCP8.5 climate scenarios. We did not take the impact of human activities on projected C stock prediction in the future into consideration. The present GCMs do not have a meaningful index for predicting or quantifying the impact of anthropogenic activities as the variable is too complex. It will be beneficial if future research could focus on how forest ecosystems respond to climate change using a stratification by forest communities, tree species and forest landscapes together with forest management planning targeting climate change adaptation.

Global change entails that climate change will be predominantly manifested by changing patterns of precipitation, increasing temperature and widespread climate extremes (Y. Li *et al.*, 2020). Thus, there is significant decrease of stomatal conductance between the canopy and the leaves, resulting in subdued vegetation growth caused by elevated temperatures and high CO_2 concentrations (Surano *et al.*, 1986; Kimball *et al.*, 1993). High rates of temperature increase during the night than during the day due to the greenhouse effect leads to the dark respiration of vegetation cancelling out forest organic matter accumulated during the day (Díaz *et al.*, 1993). At the same time, the greenhouse effect results in increased soil temperature, stimulating the activity and growth of edaphon and increasing soil organic matter mineralization. However, the increased soil organic carbon would also result in increased carbon nitrogen ration (C: N) which inhibits normal plant growth by slowing down litter decomposition rate (Díaz *et al.*, 1993; Rogers, Runion and Krupa, 1994).

Our C stock modelling approach under an RCP8.5 warming climate scenario adopts a similar hypothesis as when bioclimatic envelope models are utilised in the prediction of climate change effects on species distribution (Pearson and Dawson, 2003). For the future RCP8.5 C stock predictions, we assumed that forest species will be able to cope up with a changing climate (full dispersal hypothesis). Again, we assumed no variation in the values of the Sentinel-2 derived vegetation indices in the projected RCP8.5 although climate change should also impact them. We therefore make a conservative conclusion regarding the findings of our study predicting a significant decrease in C stock distribution and density under a warming climate scenario in the studied plantation forest ecosystem in Zimbabwe. It is highly unlikely that the

values of vegetation indices will increase with climate warming. Static and lower values of vegetation indices caused by an increase in the abundance of small and underdeveloped plantation forest species and individuals are anticipated. In light of the aforementioned hypothesis, our model has likely underestimated plantation forest loss under the projected RCP8.5 climate pathway.

Coverage and accuracy ($RMSE$) of future Bayesian prediction models could be improved by increasing the sample size since the validation metrics calculated in the present study are based on 43 observations. A small validation sample size could explain the large variability in the projected 2075 model compared to the current (1970-2000) model. Notwithstanding the fact that this study only focused on a single future climate scenario (RCP8.5), it ushers robust and far-reaching projections using a comprehensive Bayesian hierarchical model that handles prediction uncertainty much better than other approaches. Furthermore, the revised SSPs in the CMIP6 encompass a much comprehensive projected future where the business as-usual future climate scenario has substantially higher CO_2 emissions than the predecessor RCP8.5 (Yao, Piao and Wang, 2018). Characterisation of both current and future climates in a localised plantation forest in Zimbabwe advances an understanding of optimal management conditions for devising adaptation strategies to climate change impacts. Evidence from the outlook of forest plantation health points to a significant shrinkage of the area under plantation forests.

6.8.3. Conclusions

Forest C stock prediction modelling under climate warming scenarios presents unique opportunities for natural resources managers and policy makers for evaluating and assessing probable habitats at risk of adverse climate change impacts and those favorable to the same. This study demonstrated that extrapolation of C stock predictions under future SSP5-8.5 2075 climate scenario will substantially shrink the area under plantation forest at the turn of the year 2100. The application of novel and robust Bayesian hierarchical geostatistical approach captures prediction uncertainty under climate warming scenarios for the modelled C stock variable with high accuracy. The choice of the baseline data for climate warning modelling for the present study was largely governed by availability of relevant climate data. However, it would be interesting in future studies to see how C stock SSP5-8.5 future projections influence C stock predictions when the baseline data and period are changed. Such modelling variations would assist in designing more pragmatic and realistic climate change mitigation policies like the introduction of more adaptive plantation forest species. We therefore conclude that even with the best Bayesian hierarchical C stock predictive model, there is more than 90%

probability that the SSP5-8.5 business-as-usual climate scenario would have adverse impacts on the current health of C stock in the modelled plantation forest ecosystem in Zimbabwe. This would see substantial shrinkage of the present C stock density and distribution in Zimbabwe's managed plantation forest ecosystems.

CHAPTER SEVEN

A Bayesian Geo-statistical Approach for Plantation Forest Productivity Assessment after the Fast-Track Land Reform in Zimbabwe

7.1. Introduction

Climate change related international councils like the Kyoto Protocol of the United Nations Framework Convention on Climate Change (UNFCCC) have recognised and stressed the importance of C sinks and the subsequent need for monitoring, preserving and expanding global C sinks (Millington, A.; Townsend, 1989). As such, quantification of uncertainties associated with C stock prediction and estimation can greatly be improved through the incorporation of pre-experimental or historical data known for influencing forest biomass dynamics in specific forest ecosystems (González-Vélez *et al.*, 2021; Do *et al.*, 2022). Uncertainty quantification in disturbed plantation forest ecosystems is a cause for concern as reliable prediction of the resource variable of interest cannot be consistently reported as the environments are constantly under stress, partly from natural disturbances and largely from anthropogenic perturbation. Coupling the contemporary Bayesian hierarchical geostatistical inferential framework with the much-improved and freely available multispectral remote sensing ancillary data is a great opportunity to improve prediction uncertainty of forest biomass prediction in these environments.

To this day, various methods that are used to estimate and predict AGB ranging from allometric equations, remote sensing and machine learning techniques have widely been tested under the influence of remote sensing ancillary data (Nwobi and Williams, 2021; Jiang *et al.*, 2022). Various approaches of AGB estimation including optical and active remote sensing have their limitations that include saturation and high operation costs, respectively (Tonolli *et al.*, 2011a; Fararoda *et al.*, 2021). However, the opportunities presented by the Bayesian inferential technique coupled with freely available new generation remote sensing derived ancillary data from Landsat-8 and Sentinel-2 present immense opportunities for mapping resources in disturbed environments. The Bayesian approach, from its theoretical foundations, seem to be a better alternative to these methods as it has access to the entire posterior predictive distribution (PPD) (Goulard and Voltz, 1992). As Beloconi and Vounatsou (2020) postulates, full access to the PPD enables subsequent analysis that can inform management and ecological objectives whilst accounting for prediction uncertainty. Coupled with cheap and freely available

multispectral remote sensing datasets and open source programming languages, the Bayesian approach can be automated and reduce the need for extensive surveys in future sampling campaigns. The need to maintain sustainable and dependable databases of forest attributes in plantation forest monitoring and management cannot be overemphasised. This study therefore aimed at assessing how multispectral remote sensing informed Bayesian hierarchical geostatistical models handle uncertainty in C stock prediction in a disturbed plantation forest ecosystem in Zimbabwe. The study also endeavoured to evaluate how the best multispectral remote sensing driven Bayesian hierarchical C stock predictive model predicts and characterise Zimbabwe's managed plantation forest ecosystems under a warming future (RCP8.5) projected climate pathway.

7.2. Conclusion

This study sets out to assess how the application of different sources of multispectral and new generation remote sensing information using Bayesian methods influence C stock prediction uncertainty in a plantation forest ecosystem in the eastern highlands of Zimbabwe. Evidence of the application of the Bayesian hierarchical geostatistical approach with Sentinel-2 driven ancillary data outperforming other methods in the modelling of C stock is conspicuous. In respect of the findings coming out of the different studied objectives of the study, the following conclusions were made:

1. Sentinel-2 derived predictors in the form of vegetation indices and the distance to nearest settlements, $DIST$, significantly reduce prediction uncertainty in C stock modelling under a Bayesian hierarchical geostatistical framework than with Landsat-8 derived independent variables. The freely available and easy to use Sentinel-2 is a good candidate dataset for use in monitoring plantation forest ecosystems with histories of disturbance.

2. Although the Sentinel-2 based spatially varying coefficient (SVC) model applied within a Bayesian hierarchical geostatistical framework is preferable for predicting C stock over the Landsat-8 based C stock SVC model, it does not improve C stock prediction uncertainty over the conventional Bayesian hierarchical geostatistical model. As such, a Bayesian hierarchical SVC geostatistical model derived from the best multispectral new generation remote sensing satellite ancillary data (Sentinel-2) does not improve C stock prediction uncertainty in managed plantation forest ecosystems in Zimbabwe.
3. Despite the comparability of the Bayesian hierarchical geostatistical approach with the frequentist geostatistical methodology, prediction of C stock needed for climate change

and climate change mitigation strategies does not give the same prediction outcomes. When utilised with multispectral remote sensing, Sentinel-2 derived ancillary data, the Bayesian geostatistical hierarchical method provides C stock predictions with lower prediction uncertainty compared to C stock predictions from the Sentinel-2-based frequentist geostatistical inferential framework. Regardless of the ease of application and use of the frequentist geostatistical inferential methodology, the Bayesian driven inferential technique, coupled with high quality multispectral remote sensing data is a better method for predicting C stock in managed plantation forest ecosystems.

4. A Sentinel-2 driven multi-source data approach at a local scale does not improve C stock prediction uncertainty in managed forest ecosystems in Zimbabwe. A Bayesian geostatistical hierarchical model using vegetation related multi-source climatic, topographic and vegetation indices for C stock prediction established $NDVI$ and $SAVI$ as significant predictors of plantation forest ecosystems.

5. C stock density and distribution would substantially shrink under the 8.5 $W\,m^{-2}$ radiative forcing future climate scenario at the end of the 21st century as greenhouse gas emissions and population growth continue to increase and rise respectively.
6. Since the EVI was developed to better respond to vegetation signal in regions with dense AGB, its exclusion in the C stock prediction in the sampled region in Nyanga district implies that the sparse vegetation in the area respond better to soil influence that atmospheric factors. The sampled region is a disturbed environment where the distribution of plantation forest is patchy due to encroachment from neighbouring farming activities.

7.3. Recommendations for future work

Natural or managed forest plantations in the developing world usually occur in environments suffering from disturbance and distress due to encroachment from illegal human activities and pressure from expanding population, threatening land use and land cover changes. C stock prediction models built for such plantation ecosystems should therefore be sensitive to the history and dynamics of disturbance by accommodating the right basket of predictors. Although data sets from multispectral remote sensing like the Landsat-series and Sentinel-2 are a cheap source of prediction aids, consideration of anthropogenic factors is a critical step towards production of adaptive and error free C stock predictions. It is also known and acknowledged in the literature that climate and topographic variables play a part in the

distribution of C stock. Given this background, the study makes the following recommendations:

1. There is need to expand the scope of the study by considering all the possible sources of C stock spatial variability and upgrade the study to a regional scale and allow the high-quality Bayesian hierarchical geostatistical models to make a natural selection of C stock determinants from a broad basket of predictors. This is important as practitioners and policy makers would make decisions on models with low uncertainty.

2. When inventory plots are comprised of bigger sample sizes, the matrix operations of immense dimensions are required for computing model parameter estimates of higher magnitude and may not be achievable through ordinary PCs. Future work should therefore be aimed at exploring algorithms for resolving the dimensionality problem when fitting SVC models as accommodation of more climatic and topographic predictors in such models can result in the computers running slow. Resolution of the dimensionality curse is important as forest C stock is normally modelled with independent variables from multiple data sources and spatial heterogeneity is inevitable for forest resources occurring in environments with history of disturbance and different management practices.

3. Future studies should assess how a change in baseline (current) climate scenarios influences projected future climate resource predictions using contemporary Bayesian hierarchical approaches with the capability of accommodating a wide range of sources of uncertainty.

4. Future studies should also consider the utility of hyperspectral and microwave remote sensing as sources of prediction information in the modelling of C stock in the sampled region. This would create an opportunity to evaluate how microwave and hyperspectral remote sensing compare with the results of Bayesian C stock prediction models constructed from optical remotes sensing and recommend the most appropriate sensors for C stock prediction.

8. REFERENCES

Acharya, R. P., Maraseni, T. and Cockfield, G. (2019) 'Global trend of forest ecosystem services valuation – An analysis of publications', *Ecosystem Services*, 39, p. 100979. doi: https://doi.org/10.1016/j.ecoser.2019.100979.

Agarwal, D. K. *et al.* (2005) 'Tropical deforestation in Madagascar: analysis using hierarchical, spatially explicit, Bayesian regression models', *Ecological Modelling*, 185(1), pp. 105–131. doi: https://doi.org/10.1016/j.ecolmodel.2004.11.023.

Ahmed, De Marsily, G. (1987) 'Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity', *Water Resources Research*, 23(9), pp. 1717–1737.

Ahmed, N., Atzberger, C. and 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, p. 101545. doi: https://doi.org/10.1016/j.ecoinf.2021.101545.

Ali, A. *et al.* (2019) 'Climate and soils determine aboveground biomass indirectly via species diversity and stand structural complexity in tropical forests', *Forest Ecology and Management*, 432, pp. 823–831. doi: https://doi.org/10.1016/j.foreco.2018.10.024.

Ali, I. *et al.* (2015) 'Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data', *Remote Sensing*, pp. 16398–16421. doi: 10.3390/rs71215841.

Alig, R. J., Adams, D. M. and McCarl, B. A. (2002) 'Projecting impacts of global climate change on the US forest and agriculture sectors and carbon budgets', *Forest Ecology and Management*, 169(1), pp. 3–14. doi: https://doi.org/10.1016/S0378-1127(02)00290-6.

Allen, C. D. *et al.* (2010) 'A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests', *Forest Ecology and Management*, 259(4), pp. 660–684. doi: https://doi.org/10.1016/j.foreco.2009.09.001.

Álvarez-Dávila, E. *et al.* (2017) 'Forest biomass density across large climate gradients in northern South America is related to water availability but not with temperature', *PloS one*. Public Library of Science San Francisco, CA USA, 12(3), p. e0171072.

Van Amstel, A. (2006) *IPCC 2006 Guidelines for National Greenhouse Gas Inventories*.

Araki, S., Yamamoto, K. and Kondo, A. (2015) 'Application of Regression Kriging to Air Pollutant Concentrations in Japan with High Spatial Resolution', *Aerosol and Air Quality Research*, 15, pp. 234–241. doi: 10.4209/aaqr.2014.01.0011.

Astola, H. *et al.* (2019) 'Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region', *Remote Sensing of Environment*, 223, pp. 257–273. doi: https://doi.org/10.1016/j.rse.2019.01.019.

Azevedo, L. (2021) 'Model reduction in geostatistical seismic inversion with functional data analysis', *GEOPHYSICS*. Society of Exploration Geophysicists, 87(1), pp. M1–M11. doi: 10.1190/geo2021-0096.1.

Babcock, C. *et al.* (2012) 'Multivariate spatial regression models for predicting individual tree

structure variables using LiDAR data', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. IEEE, 6(1), pp. 6–14.
Babcock, C. *et al.* (2015) 'LiDAR based prediction of forest biomass using hierarchical models with spatially varying coefficients', *Remote Sensing of Environment*, 169, pp. 113–127. doi: https://doi.org/10.1016/j.rse.2015.07.028.
Babcock, C. *et al.* (2016) 'Modeling forest biomass and growth: Coupling long-term inventory and LiDAR data', *Remote Sensing of Environment*, 182, pp. 1–12. doi: https://doi.org/10.1016/j.rse.2016.04.014.
Babcock, C. *et al.* (2018) 'Geostatistical estimation of forest biomass in interior Alaska combining Landsat-derived tree cover, sampled airborne lidar and field observations', *Remote Sensing of Environment*, 212, pp. 212–230. doi: https://doi.org/10.1016/j.rse.2018.04.044.
Baccini, A. *et al.* (2004) 'Forest biomass estimation over regional scales using multisource data', *Geophysical Research Letters*. John Wiley & Sons, Ltd, 31(10). doi: https://doi.org/10.1029/2004GL019782.
Baloloy, A. B. *et al.* (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 Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* Copernicus Publications, IV–3, pp. 29–36. doi: 10.5194/isprs-annals-IV-3-29-2018.
Banerjee, S. *et al.* (2010) 'Hierarchical spatial process models for multiple traits in large genetic trials', *Journal of the American Statistical Association*. Taylor & Francis, 105(490), pp. 506–521.
Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2003) *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.
Baugh, W. M. and Groeneveld, D. P. (2006) 'Broadband vegetation index performance evaluated for a low-cover environment', *International Journal of Remote Sensing*. Taylor & Francis, 27(21), pp. 4715–4730. doi: 10.1080/01431160600758543.
Beloconi, A. and Vounatsou, P. (2020) 'Bayesian geostatistical modelling of high-resolution NO2 exposure in Europe combining data from monitors, satellites and chemical transport models', *Environment International*, 138, p. 105578. doi: https://doi.org/10.1016/j.envint.2020.105578.
BOISVENUE, C. and RUNNING, S. W. (2006) 'Impacts of climate change on natural forest productivity – evidence since the middle of the 20th century', *Global Change Biology*. John Wiley & Sons, Ltd, 12(5), pp. 862–882. doi: https://doi.org/10.1111/j.1365-2486.2006.01134.x.
Bordoloi, R. *et al.* (2022) 'Satellite based integrated approaches to modelling spatial carbon stock and carbon sequestration potential of different land uses of Northeast India', *Environmental and Sustainability Indicators*, 13, p. 100166. doi: https://doi.org/10.1016/j.indic.2021.100166.
Box, G. E. P. and Cox, D. R. (1964) 'An analysis of transformations (with discussion)', *Journal of the Royal Statistical Society*, 26, pp. 211–252.
Box, G. E. P. and Cox, D. R. (1982) 'An analysis of transformations revisited, rebutted. , 77, 209–

210', *Journal of the American Statistical Association*, 77, pp. 209–210.
Brown, S. (1997) 'Estimating Biomass and Biomass Change of Tropical Forests: A Primer', *FAO Forestry Paper*, 134.
Brus, D. J., de Gruijter, J. J. and van Groenigen, J. W. (2006) 'Chapter 14 Designing Spatial Coverage Samples Using the k-means Clustering Algorithm', *Developments in soil science*, 31, pp. 183–192.
Business Times (2023) *Zim faces new tobacco ban threat*. Available at: https://businesstimes.co.zw/zim-faces-new-tobacco-ban-threat/.
Cameletti, M. and Biondi, F. (2019) 'Hierarchical modeling of space-time dendroclimatic fields: Comparing a frequentist and a Bayesian approach', *Arctic, Antarctic, and Alpine Research*. Taylor & Francis, 51(1), pp. 115–127.
Cao, L. *et al.* (2016) 'Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data', *Remote Sensing of Environment*, 178, pp. 158–171. doi: https://doi.org/10.1016/j.rse.2016.03.012.
Chave, J. *et al.* (2014) 'Improved allometric models to estimate the aboveground biomass of tropical trees', *Global change biology*. Wiley Online Library, 20(10), pp. 3177–3190.
Chen, G. *et al.* (2012) 'The influence of sampling density on geographically weighted regression: a case study using forest canopy height and optical data', *International Journal of Remote Sensing*. Taylor & Francis, 33(9), pp. 2909–2924. doi: 10.1080/01431161.2011.624130.
Chiles, J.-P. and Delfiner, P. (2009) *Geostatistics: modeling spatial uncertainty*. John Wiley & Sons.
Chinembiri, T. S. *et al.* (2013) 'The Precision of C Stock Estimation in the Ludhikola Watershed Using Model-Based and Design-Based Approaches', *Natural Resources Research*, 22(4), pp. 297–309. doi: DOI 10.1007/s11053-013-9216-6.
Chinembiri, T. S., Mutanga, O. and Dube, T. (2023) 'Carbon Stock Prediction in Managed Forest Ecosystems Using Bayesian and Frequentist Geostatistical Techniques and New Generation Remote Sensing Metrics', *Remote Sensing*, 15(1649). doi: https://doi.org/10.3390/rs15061649.
Chinembiri, T.S., Mutanga, O. and Dube, T. (2023) 'Hierarchical Bayesian geostatistics for C stock prediction in disturbed plantation forest in Zimbabwe', *Ecological Informatics*, 73, p. 101934.
Christman, Z. J. and Rogan, J. (2012) 'Error Propagation in Raster Data Integration: Impacts on Landscape Composition and Configuration', *Photogrammetric Engineering & Remote Sensing*, 78(6), pp. 617–624. doi: DOI: 10.14358/PERS.78.6.617.
Chrysafis, I. *et al.* (2017) 'Assessing the relationships between growing stock volume and Sentinel-2 imagery in a Mediterranean forest ecosystem', *Remote Sensing Letters*. Taylor & Francis, 8(6), pp. 508–517. doi: 10.1080/2150704X.2017.1295479.
Clerici, N. *et al.* (2016) 'Estimating Aboveground Biomass and Carbon Stocks in Periurban Andean Secondary Forests Using Very High Resolution Imagery', *Forests*. doi: 10.3390/f7070138.
CMS (2014) *NASA carbon monitoring system, http://carbon.nasa.gov*. Available at: http://carbon.nasa.gov (Accessed: 1 June 2022).
Cressie, N. (1993) *Statistics for spatial data. Wiley series in probability and mathematical statistics:*

Applied probability and statistics. John Wiley & Sons.
Crosby, M. K. *et al.* (2017) 'Consequences of Landsat Image Strata Classification Errors on Bias and Variance of Inventory Estimates: A Forest Inventory Case Study', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(1), pp. 243–251. doi: 10.1109/JSTARS.2016.2597762.
Dang, A. T. N. *et al.* (2019) 'Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam', *Ecological Informatics*, 50, pp. 24–32. doi: https://doi.org/10.1016/j.ecoinf.2018.12.010.
Datta, A. *et al.* (2016) 'Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets', *Journal of the American Statistical Association*. Taylor & Francis, 111(514), pp. 800–812. doi: 10.1080/01621459.2015.1044091.
Demirhan, H. and Kalaylioglu, Z. (2015) 'Joint prior distributions for variance parameters in Bayesian analysis of normal hierarchical models', *Journal of Multivariate Analysis*, 135, pp. 163–174. doi: https://doi.org/10.1016/j.jmva.2014.12.013.
Deng, S. *et al.* (2014) 'Estimating Forest Aboveground Biomass by Combining ALOS PALSAR and WorldView-2 Data: A Case Study at Purple Mountain National Park, Nanjing, China', *Remote Sensing*, pp. 7878–7910. doi: 10.3390/rs6097878.
Díaz, S. *et al.* (1993) 'Evidence of a feedback mechanism limiting plant response to elevated carbon dioxide', *Nature*, 364(6438), pp. 616–617. doi: 10.1038/364616a0.
Diggle, P. and Ribeiro Jr, P. (2007) 'Model-Based Geostatistics', in *J. R. Stat. Soc. C*. doi: 10.1007/978-0-387-48536-2.
Dixon, R. K. *et al.* (1994) 'Carbon Pools and Flux of Global Forest Ecosystems', *Science*. American Association for the Advancement of Science, 263(5144), pp. 185–190. doi: 10.1126/science.263.5144.185.
Do, A. N. T. *et al.* (2022) 'Monitoring landscape fragmentation and aboveground biomass estimation in Can Gio Mangrove Biosphere Reserve over the past 20 years', *Ecological Informatics*, 70, p. 101743. doi: https://doi.org/10.1016/j.ecoinf.2022.101743.
Drusch, M. *et al.* (2012) 'Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services', *Remote Sensing of Environment*, 120, pp. 25–36. doi: https://doi.org/10.1016/j.rse.2011.11.026.
Dube, T. and 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, pp. 36–46. doi: https://doi.org/10.1016/j.isprsjprs.2014.11.001.
Duchene, S. *et al.* (2016) 'Cross-validation to select Bayesian hierarchical models in phylogenetics', *BMC Evolutionary Biology*, 16. doi: 10.1186/s12862-016-0688-y.
Eamus, D. (2003) 'How does ecosystem water balance affect net primary productivity of woody ecosystems?', *Functional Plant Biology*, 30(2), pp. 187–205. Available at:

https://doi.org/10.1071/FP02084.
El-Askary, H. *et al.* (2014) 'Change detection of coral reef habitat using Landsat-5 TM, Landsat 7 ETM+ and Landsat 8 OLI data in the Red Sea (Hurghada, Egypt)', *International Journal of Remote Sensing*. Taylor & Francis, 35(6), pp. 2327–2346. doi: 10.1080/01431161.2014.894656.
European-Commission (2017) *Timber Trade Flows within, to and from Eastern and Southern African Countries*. Brussels.
Fan, M., Wang, X. and Yang, G. (2022) 'Spatial characteristics of vegetation habitat suitability and mountainous settlements and their quantitative relationships in upstream of Min River, southwestern of China', *Ecological Informatics*, 68, p. 101541. doi: https://doi.org/10.1016/j.ecoinf.2021.101541.
Fang, J. *et al.* (2001) 'Changes in Forest Biomass Carbon Storage in China Between 1949 and 1998', *Science*. American Association for the Advancement of Science, 292(5525), pp. 2320–2322. doi: 10.1126/science.1058629.
FAO (2003) *Forestry Outlook Study for Africa. Regional Report - Opportunities and challenges towards 2020*. Rome, Italy.
FAO (2005) *Global Forest Resources Assessment Country Report, Zimbabwe*.
FAO (2016) *Global forest resources assessment. Howare the world's forests changing?* Rome, Italy.
Fararoda, R. *et al.* (2021) 'Improving forest above ground biomass estimates over Indian forests using multi source data sets with machine learning algorithm', *Ecological Informatics*, 65, p. 101392. doi: https://doi.org/10.1016/j.ecoinf.2021.101392.
Farr, T. G. *et al.* (2007) 'The Shuttle Radar Topography Mission', *Reviews of Geophysics*. John Wiley & Sons, Ltd, 45(2). doi: https://doi.org/10.1029/2005RG000183.
Fassnacht, F. E. *et al.* (2014) 'Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass', *Remote Sensing of Environment*, 154, pp. 102–114. doi: https://doi.org/10.1016/j.rse.2014.07.028.
Fick, S. E. and Hijmans, R. J. (2017) 'WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas', *International Journal of Climatology*. John Wiley & Sons, Ltd, 37(12), pp. 4302–4315. doi: https://doi.org/10.1002/joc.5086.
Finley, A. O. and Banerjee, S. (2008) 'Bayesian Spatial Regression for Multi-source Predictive Mapping BT - Encyclopedia of GIS', in Shekhar, S. and Xiong, H. (eds). Boston, MA: Springer US, pp. 45–52. doi: 10.1007/978-0-387-35973-1_97.
Finley, A. O., Banerjee, S. and MacFarlane, D. W. (2011) 'A Hierarchical Model for Quantifying Forest Variables Over Large Heterogeneous Landscapes With Uncertain Forest Areas', *Journal of the American Statistical Association*. Taylor & Francis, 106(493), pp. 31–48. doi: 10.1198/jasa.2011.ap09653.
Finley, A. O., Banerjee, S. and McRoberts, R. E. (2008) 'A Bayesian approach to multi-source forest area estimation', *Environ Ecol Stat*, 15, pp. 241–258.
Finley, A. O., Banerjee, S. and McRoberts, R. E. (2009) 'Hierarchical spatial models for predicting tree species assemblages across large domains', *The Annals of Applied Statistics*, 3(3), pp. 1052–

1079. doi: 10.1214/09-AOAS250.
Finley, A., Sudipto, B. and Carlin, B. (2007) 'spBayes: An R Package for Univariate and Multivariate Hierarchical Point-referenced Spatial Models', *Journal of Statistical Software*, 19. doi: 10.18637/jss.v019.i04.
Forestry-Commission (2021) *Zimbabwe land and vegetation cover area estimates*. Harare.
Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2003) *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.
Frampton, W. J. *et al.* (2013) 'Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation', *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, pp. 83–92. doi: https://doi.org/10.1016/j.isprsjprs.2013.04.007.
Fuller, A. *et al.* (2010) 'Physiological mechanisms in coping with climate change', *Physiological and Biochemical Zoology*. The University of Chicago Press, 83(5), pp. 713–720.
Gelfand, A. *et al.* (2004) 'Nonstationary multivariate process modeling through spatially varying coregionalization', *Test*, 13, pp. 263–312.
Gelfand, A. E. *et al.* (2003) 'Spatial Modeling With Spatially Varying Coefficient Processes', *Journal of the American Statistical Association*. Taylor & Francis, 98(462), pp. 387–396. doi: 10.1198/016214503000170.
Gelfand, A. E. *et al.* (2004) 'Nonstationary multivariate process modeling through spatially varying coregionalization', *Test*, 13(2), pp. 263–312. doi: 10.1007/BF02595775.
Gelfand, A. E. (2012) 'Hierarchical modeling for spatial data problems', *Spatial Statistics*, 1, pp. 30–39. doi: https://doi.org/10.1016/j.spasta.2012.02.005.
Gelman, A. *et al.* (1995) *Bayesian data analysis*. Chapman and Hall/CRC.
Gelman, A. *et al.* (2004) *Bayesian Data Analysis*. 2nd edn. London: Chapman and Hall/CRC.
Gelman, A. (2006) 'Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)', *Bayesian Analysis*, 1(3), pp. 515–534. doi: 10.1214/06-BA117A.
Gelman, A. and Rubin, D. (1992) 'Inference from iterative simulation using multiple sequences', *Statistical Science*, (7), pp. 457–511.
Gerald, F. *et al.* (2017) 'lansat-8 vs. sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso', *GIScience & Remote Sensing*, pp. 1–26.
Ghasemi, A. R. (2015) 'Changes and trends in maximum, minimum and mean temperature series in Iran', *Atmospheric Science Letters*. John Wiley & Sons, Ltd, 16(3), pp. 366–372. doi: https://doi.org/10.1002/asl2.569.
Ghosh, G. and Carriazo, F. (2007) 'Bayesian and Frequentist Approaches to Hedonic Modeling in a Geo-Statistical Framework'.
Gibbs, H. K. *et al.* (2007) 'Monitoring and estimating tropical forest carbon stocks: making REDD a reality', *Environmental Research Letters*. IOP Publishing, 2(4), p. 45023. doi: 10.1088/1748-9326/2/4/045023.
González-Vélez, J. C. *et al.* (2021) 'An artificial intelligent framework for prediction of wildlife

vehicle collision hotspots based on geographic information systems and multispectral imagery', *Ecological Informatics*, 63, p. 101291. doi: https://doi.org/10.1016/j.ecoinf.2021.101291.
Gordon, C. E. *et al.* (2018) 'Aboveground carbon sequestration in dry temperate forests varies with climate not fire regime', *Global Change Biology*. John Wiley & Sons, Ltd, 24(9), pp. 4280–4292. doi: https://doi.org/10.1111/gcb.14308.
Goulard, M. and Voltz, M. (1992) 'Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix', *Mathematical Geology*, 24(3), pp. 269–286. doi: 10.1007/BF00893750.
Government of Zimbabwe, Z. (2021) *Zimbabwe Revised Nationally Determined Contribution, 2021*. Available at: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjD9ZrW75v8AhULUcAKHUfJDiIQFnoECAwQAw&url=https%3A%2F%2Funfccc.int%2Fsites%2Fdefault%2Ffiles%2FNDC%2F2022-06%2FZimbabwe%2520Revised%2520Nationally%2520Determined%2520C.
Green, E. J., Finley, A. O. and Strawderman, W. E. (2020) *Introduction to Bayesian Methods in Ecology and Natural Resources*. Springer, Cham. doi: https://doi.org/10.1007/978-3-030-60750-0.
Guhaniyogi, R. and Banerjee, S. (2019) 'Multivariate spatial meta kriging', *Statistics & Probability Letters*, 144, pp. 3–8. doi: https://doi.org/10.1016/j.spl.2018.04.017.
Gupta, A., Kamble, T. and Machiwal, D. (2017) 'Comparison of ordinary and Bayesian kriging techniques in depicting rainfall variability in arid and semi-arid regions of north-west India', *Environmental Earth Sciences*, 76(15), p. 512. doi: 10.1007/s12665-017-6814-3.
Hazra, A. (2017) 'Using the confidence interval confidently', *Journal of Thoracic Disease; Vol 9, No 10 (October 31, 2017): Journal of Thoracic Disease*. Available at: https://jtd.amegroups.com/article/view/16406.
He, Q. *et al.* (2022) 'Investigating and predicting spatiotemporal variations in vegetation cover in transitional climate zone: a case study of Gansu (China)', *Theoretical and Applied Climatology*. Springer, 150(1–2), pp. 283–307.
Hengl, T. *et al.* (2004) 'A double continuous approach to visualization and analysis of categorical maps', *International Journal of Geographical Information Science*. Taylor & Francis, 18(2), pp. 183–202. doi: 10.1080/13658810310001620924.
Hoeting, J. A. (2009) 'The importance of accounting for spatial and temporal correlation in analyses of ecological data', *Ecological Applications*. John Wiley & Sons, Ltd, 19(3), pp. 574–577. doi: https://doi.org/10.1890/08-0836.1.
Https://www.opr.ca.gov/ (2021) *Governor's Office of Planning and Research*. Available at: https://www.opr.ca.gov/.
Hu, T. *et al.* (2016) 'Mapping Global Forest Aboveground Biomass with Spaceborne LiDAR, Optical Imagery, and Forest Inventory Data', *Remote Sensing*. doi: 10.3390/rs8070565.
Hu, T. *et al.* (2020) 'Mapping the global mangrove forest aboveground biomass using multisource remote sensing data', *Remote sensing*. MDPI, 12(10), p. 1690.

Hudak, A. T. *et al.* (2002) 'Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height', *Remote Sensing of Environment*, 82(2), pp. 397–416. doi: https://doi.org/10.1016/S0034-4257(02)00056-1.
Hudson, G. and Wackernagel, H. (1994) 'Mapping temperature using kriging with external drift: Theory and an example from scotland', *International Journal of Climatology*. John Wiley & Sons, Ltd, 14(1), pp. 77–91. doi: https://doi.org/10.1002/joc.3370140107.
Initiative For Climate Action Transparency (ICAT) (2022) *Zimbabwe on track to better climate action transparency*. Available at: https://climateactiontransparency.org/zimbabwe-on-track-to-better-climate-action-transparency/ (Accessed: 28 December 2022).
IPCC (2006) *Good Practice Guidance for Land Use, Land-Use Change and Forestry*. Kanagawa, Japan: Institute for Global Environmental Strategies.
IPCC (2014) *The IPCC's Fifth Assessment Report: Whats in it for Africa*.
IPCC, C. C. (2013) 'The Physical Science Basis, An overview of the Working Group 1 contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)'. Cambridge University Press.
Isaaks, E. H. and Srivastava, R. M. (1989) *Applied Geostatistics*. Oxford University Press. Available at: https://books.google.co.zw/books?id=vC2dcXFLI3YC.
Jackman, S. (2000) 'Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo', *American Journal of Political Science*. [Midwest Political Science Association, Wiley], 44(2), pp. 375–404. doi: 10.2307/2669318.
Jeffreys, H. (1961) *Theory of probability*. 3rd ed. New York, NY: Oxford University Press.
Jha, N. *et al.* (2021) 'The real potential of current passive satellite data to map aboveground biomass in tropical forests', *Remote Sensing in Ecology and Conservation*, 7. doi: 10.1002/rse2.203.
Jia, K. *et al.* (2014) 'Land cover classification using Landsat 8 operational land imager data in Beijing, China', *Geocarto international*. Taylor & Francis, 29(8), pp. 941–951.
Jiang, F. *et al.* (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, p. 147335. doi: https://doi.org/10.1016/j.scitotenv.2021.147335.
Jiang, F. *et al.* (2022) 'Above-Ground Biomass Estimation for Coniferous Forests in Northern China Using Regression Kriging and Landsat 9 Images', *Remote Sensing*. doi: 10.3390/rs14225734.
Johnson, K. D. *et al.* (2014) 'Integrating forest inventory and analysis data into a LIDAR-based carbon monitoring system', *Carbon Balance and Management*, 9(1), p. 3. doi: 10.1186/1750-0680-9-3.
Jucker, T. *et al.* (2018) 'Canopy structure and topography jointly constrain the microclimate of human-modified tropical landscapes', *Global change biology*. Wiley Online Library, 24(11), pp. 5243–5258.
Karl, T. R. and Trenberth, K. E. (2003) 'Modern Global Climate Change', *Science*. American Association for the Advancement of Science, 302(5651), pp. 1719–1723. doi:

10.1126/science.1090228.
Keith, H. *et al.* (2014) 'Accounting for biomass carbon stock change due to wildfire in temperate forest landscapes in Australia', *PLoS One*. Public Library of Science San Francisco, USA, 9(9), p. e107126.
Keith, H., Mackey, B. G. and Lindenmayer, D. B. (2009) 'Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 106(28), pp. 11635–11640. doi: 10.1073/pnas.0901970106.
Kimball, B. A. *et al.* (1993) 'Effects of increasing atmospheric CO2 on vegetation', *Vegetatio*, 104(1), pp. 65–75. doi: 10.1007/BF00048145.
Korhonen, L. *et al.* (2017) 'Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index', *Remote Sensing of Environment*, 195, pp. 259–274. doi: https://doi.org/10.1016/j.rse.2017.03.021.
Koven, C. D., Riley, W. J. and Stern, A. (2013) 'Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 Earth System Models', *Journal of Climate*, 26(6), pp. 1877–1900.
Kramer, C. Y. (1956) 'Extension of multiple range tests to group means with unequal numbers of replications.', *Biometrics*, 12, pp. 307–310. doi: 10.2307/3001469.
Kumar, A., Sharma, P. and Joshi, S. (2016) 'Assessing the impacts of climate change on land productivity in Indian crop agriculture: An evidence from panel data analysis', *Journal of Agricultural Science and Technology*. Journal of Agricultural Science and Technology, 18(1), pp. 1–13.
Kupfersberger, H., Deutsch, C. V and Journel, A. G. (1998) 'Deriving constraints on small-scale variograms due to variograms of large-scale data', *Mathematical geology*. Springer, 30(7), pp. 837–852.
Larjavaara, M. and Muller-Landau, H. C. (2012) 'Still rethinking the value of high wood density', *American Journal of Botany*, 99(1), pp. 165–168. doi: https://doi.org/10.3732.
Lazarus, R. S. and Eriksen, C. W. (1952) 'Effects of failure stress upon skilled performance.', *Journal of Experimental Psychology*, 43(2), pp. 100–105. doi: https://doi.org/10.1037/h0056614.
Lefsky, M. A. (2010) 'A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System', *Geophysical Research Letters*. Wiley Online Library, 37(15).
Li, C. and Li, X. (2019) 'Hazard rate and reversed hazard rate orders on extremes of heterogeneous and dependent random variables', *Statistics & Probability Letters*, 146, pp. 104–111. doi: https://doi.org/10.1016/j.spl.2018.11.005.
Li, L. *et al.* (2020) 'Estimating Urban Vegetation Biomass from Sentinel-2A Image Data', *Forests*. doi: 10.3390/f11020125.
Li, S. *et al.* (2017) 'Sentinel-2 MSI Radiometric Characterization and Cross-Calibration with Landsat-

8 OLI', *Advances in Remote Sensing*, 6(2), pp. 147–159. doi: 10.4236/ars.2017.62011.
Li, Y. *et al.* (2020) 'Optimized Maxent Model Predictions of Climate Change Impacts on the Suitable Distribution of Cunninghamia lanceolata in China', *Forests*. doi: 10.3390/f11030302.
Liang, S. and Wang, J. B. T.-A. R. S. (Second E. (eds) (2020) 'Chapter 14 - Aboveground biomass', in. Academic Press, pp. 543–580. doi: https://doi.org/10.1016/B978-0-12-815826-5.00014-3.
Lisboa, S. *et al.* (2018) 'Biomass allometric equation and expansion factor for a mountain moist evergreen forest in Mozambique', *Carbon Balance and Management*, 13. doi: 10.1186/s13021-018-0111-7.
Liu, Y. *et al.* (2014) 'Carbon carry capacity and carbon sequestration potential in China based on an integrated analysis of mature forest biomass', *Science China Life Sciences*, 57(12), pp. 1218–1229. doi: 10.1007/s11427-014-4776-1.
Liu, Y. *et al.* (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*. Elsevier, 151, pp. 277–289.
Lloyd, J. and Farquhar, G. D. (1996) 'The CO 2 dependence of photosynthesis, plant growth responses to elevated atmospheric CO 2 concentrations and their interaction with soil nutrient status. I. General principles and forest ecosystems', *Functional Ecology*. JSTOR, pp. 4–32.
López-Serrano, P. M. *et al.* (2016) 'A Comparison of Machine Learning Techniques Applied to Landsat-5 TM Spectral Data for Biomass Estimation', *Canadian Journal of Remote Sensing*. Taylor & Francis, 42(6), pp. 690–705. doi: 10.1080/07038992.2016.1217485.
Lu, D. (2006) 'The Potential and Challenge of Remote Sensing–Based Biomass Estimation', *International Journal of Remote Sensing*, 27. doi: 10.1080/01431160500486732.
Lu, D. *et al.* (2016) 'A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems', *International Journal of Digital Earth*. Taylor & Francis, 9(1), pp. 63–105. doi: 10.1080/17538947.2014.990526.
MALHI, Y. *et al.* (2006) 'The regional variation of aboveground live biomass in old-growth Amazonian forests', *Global Change Biology*. John Wiley & Sons, Ltd, 12(7), pp. 1107–1138. doi: https://doi.org/10.1111/j.1365-2486.2006.01120.x.
Mansfield, E. R. and Helms, B. P. (1982) 'Detecting Multicollinearity', *The American Statistician*. Taylor & Francis, 36(3a), pp. 158–160. doi: 10.1080/00031305.1982.10482818.
Matose, F. (2008) 'Institutional configurations around forest reserves in Zimbabwe: the challenge of nested institutions for resource management', *Local Environment*. Routledge, 13(5), pp. 393–404. doi: 10.1080/13549830701809627.
Medlyn, B. E., Duursma, R. A. and Zeppel, M. J. B. (2011) 'Forest productivity under climate change: a checklist for evaluating model studies', *WIREs Climate Change*. John Wiley & Sons, Ltd, 2(3), pp. 332–355. doi: https://doi.org/10.1002/wcc.108.
Meyer, H. L. *et al.* (2019) 'Comparison of Landsat-8 and Sentinel-2 Data for Estimation of Leaf Area Index in Temperate Forests', *remote sensing*, 11(1160), pp. 1–16.

Millington, A.; Townsend, J. (1989) *Biomass assessment. Woody biomass in the SADC region*. London: Earthscan Publications Ltd.

Mitchard, E. T. A. (2018) 'The tropical forest carbon cycle and climate change', *Nature*, 559(7715), pp. 527–534. doi: 10.1038/s41586-018-0300-2.

Mon, M. S. *et al.* (2012) 'Factors affecting deforestation andforest degradation in selectively logged productionforest: a case study in Myanmar', *Forest Ecologyand Management*, 267, pp. 190–198.

Moss, R. H. *et al.* (2010) 'The next generation of scenarios for climate change research and assessment', *Nature*. Nature Publishing Group UK London, 463(7282), pp. 747–756.

Mourguiart, B. *et al.* (2023) 'A new method to explicitly estimate the shift of optimum along gradients in multispecies studies', *Journal of Biogeography*. John Wiley & Sons, Ltd, 50(5), pp. 1000–1011. doi: https://doi.org/10.1111/jbi.14570.

Murphy, K. P. (2012) 'A probabilistic perspective', *Text book*.

Mutanga, O., Dube, T. and Ahmed, F. (2016) 'Progress in remote sensing: vegetation monitoring in South Africa', *South African Geographical Journal*. Routledge, 98(3), pp. 461–471. doi: 10.1080/03736245.2016.1208586.

Myneni, R. B. *et al.* (2001) 'A large carbon sink in the woody biomass of Northern forests', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 98(26), pp. 14784–14789. doi: 10.1073/pnas.261555198.

Newsday (2017) *Forestry Commision decentralise issuance of timber movement*. Available at: https://www.newsday.co.zw/2017/07/forestry-commision-decentralise-issuance-timber-movement/.

Neyman, J. (1937) 'Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability', *Philosophical Transactions of the Royal Society A 236*, pp. 333–380. doi: https://doi.org/10.1098/rsta.1937.0005.

Nunes, L. J. R. *et al.* (2020) 'Forest Contribution to Climate Change Mitigation: Management Oriented to Carbon Capture and Storage', *Climate*. doi: 10.3390/cli8020021.

Nwobi, C. J. and Williams, M. (2021) 'Natural and Anthropogenic Variation of Stand Structure and Aboveground Biomass in Niger Delta Mangrove Forests ', *Frontiers in Forests and Global Change* . Available at: https://www.frontiersin.org/article/10.3389/ffgc.2021.746671.

Pan, Y. *et al.* (2011) 'A Large and Persistent Carbon Sink in the World's Forests', *Science*. American Association for the Advancement of Science, 333(6045), pp. 988–993. doi: 10.1126/science.1201609.

Pan, Y. *et al.* (2013) 'The Structure, Distribution, and Biomass of the World's Forests', *Annual Review of Ecology, Evolution, and Systematics*. Annual Reviews, 44(1), pp. 593–622. doi: 10.1146/annurev-ecolsys-110512-135914.

Pascual, A., Tupinambá-Simões, F. and de Conto, T. (2022) 'Using multi-temporal tree inventory data in eucalypt forestry to benchmark global high-resolution canopy height models. A showcase in Mato Grosso, Brazil', *Ecological Informatics*, 70, p. 101748. doi: https://doi.org/10.1016/j.ecoinf.2022.101748.

Pearson, R. G. and Dawson, T. P. (2003) 'Predicting the impacts of climate change on the distribution

of species: are bioclimate envelope models useful?', *Global ecology and biogeography*. Wiley Online Library, 12(5), pp. 361–371.
Pesaresi, M. *et al.* (2016) 'Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas', *Remote Sensing* . doi: 10.3390/rs8040299.
Poorter, L. *et al.* (2016) 'Biomass resilience of Neotropical secondary forests', *Nature*, 530(7589), pp. 211–214. doi: 10.1038/nature16512.
Popescu, S. C. *et al.* (2011) 'Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level', *Remote Sensing of Environment*, 115(11), pp. 2786–2797. doi: https://doi.org/10.1016/j.rse.2011.01.026.
Puliti, S. *et al.* (2020) 'Modelling above-ground biomass stock over Norway using national forest inventory data with ArcticDEM and Sentinel-2 data', *Remote Sensing of Environment*. Elsevier, 236, p. 111501.
R Core Development Team (2008) 'A language and Environment for Statistical Computing'. Vienna: R Foundation for Statistical Computing.
Rabus, B. *et al.* (2003) 'The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar', *ISPRS Journal of Photogrammetry and Remote Sensing*, 57(4), pp. 241–262. doi: https://doi.org/10.1016/S0924-2716(02)00124-7.
Ranghetti, L. *et al.* (2020) '"sen2r": An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data', *Computers & Geosciences*, 139, p. 104473. doi: https://doi.org/10.1016/j.cageo.2020.104473.
Ravindranath, N. H. and Ostwald, M. (2008) 'Carbon Inventory Methods Handbook for Greenhouse Gas Inventory, Carbon Mitigation and Roundwood Production Projects', *Advances in Global Change Research Volume*, 29. doi: 10.1007/978-1-4020-6547-7_1.
Rebetez, M. *et al.* (2006) 'Heat and drought 2003 in Europe: a climate synthesis', *Annals of Forest Science*. EDP Sciences, 63(6), pp. 569–577.
Rebetez, M. and Reinhard, M. (2008) 'Monthly air temperature trends in Switzerland 1901–2000 and 1975–2004', *Theoretical and Applied Climatology*. Springer, 91, pp. 27–34.
Richardson, A. D. *et al.* (2013) 'Climate change, phenology, and phenological control of vegetation feedbacks to the climate system', *Agricultural and Forest Meteorology*, 169, pp. 156–173. doi: https://doi.org/10.1016/j.agrformet.2012.09.012.
Robinson, W. S. (1950) 'Ecological correlations and the behavior of individuals', *American Sociological Review*, 15(3), pp. 351–357.
Rodima-Taylor, D., Olwig, M. F. and Chhetri, N. (2012) 'Adaptation as innovation, innovation as adaptation: An institutional approach to climate change', *Applied Geography*, 33(0), pp. 107–111.
Rogers, H. H., Runion, G. B. and Krupa, S. V (1994) 'Plant responses to atmospheric CO2 enrichment with emphasis on roots and the rhizosphere', *Environmental Pollution*, 83(1), pp. 155–189. doi: https://doi.org/10.1016/0269-7491(94)90034-5.
Rudel, T. K. *et al.* (2005) 'Forest transitions: towards a global understanding of land use change',

Global Environmental Change, 15(1), pp. 23–31. doi: https://doi.org/10.1016/j.gloenvcha.2004.11.001.
SAATCHI, S. S. *et al.* (2007) 'Distribution of aboveground live biomass in the Amazon basin', *Global Change Biology*. John Wiley & Sons, Ltd, 13(4), pp. 816–837. doi: https://doi.org/10.1111/j.1365-2486.2007.01323.x.
Sahu, S. K. (2022) *Bayesian Modeling of Spatio Temporal Data with R*. 1st ed. Highfield, Southampton, UK: Chapman and Hall/CRC. doi: https://doi.org/10.1201/9780429318443.
Sales, M. H. *et al.* (2007) 'Improving spatial distribution estimation of forest biomass with geostatistics: A case study for Rondônia, Brazil', *Ecological Modelling*, 205(1), pp. 221–230. doi: https://doi.org/10.1016/j.ecolmodel.2007.02.033.
Schabenberger, O. and Gotway, C. A. (2004) *Statistical Methods for Spatial Data Analysis*. Taylor & Francis (Chapman & Hall/CRC Texts in Statistical Science). Available at: https://books.google.co.zw/books?id=iVJuVLArmZcC.
Semela, M., Ramoelo, A. and Adelabu, S. (2020) 'Testing and Comparing the Applicability of Sentinel-2 and Landsat 8 Reflectance Data in Estimating Mountainous Herbaceous Biomass Before and After Fire Using Random Forest Modelling', in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 4493–4496. doi: 10.1109/IGARSS39084.2020.9323446.
Seo, J.-I. and Song, J. J. (2019) 'A Bayesian nonparametric model for upper record data', *Applied Mathematical Modelling*. Elsevier, 71, pp. 363–374.
Serrano, P. M. L., Corral-Rivas, J. J. and López-Sánchez, C. A. (2019) 'Modeling of Aboveground Biomass with Landsat 8 OLI and Machine Learning in Temperate Forests', *Remote Sensing*, 11(1). doi: 0.3390/f11010011.
Shi, L. and Liu, S. (2017) 'Methods of Estimating Forest Biomass: A Review', in Tumuluru, J. S. (ed.). Rijeka: IntechOpen, p. Ch. 2. doi: 10.5772/65733.
Shoko, C. and Gara, T. (2021) 'Remote sensing of aboveground grass biomass between protected and non-protected areas in savannah rangelands', *African Journal of Ecology*, 59. doi: 10.1111/aje.12861.
Sibanda, M. *et al.* (2017) 'Estimating Biomass of Native Grass Grown under Complex Management Treatments Using WorldView-3 Spectral Derivatives', *Remote Sensing* . doi: 10.3390/rs9010055.
Simard, M. *et al.* (2011) 'Mapping forest canopy height globally with spaceborne lidar', *Journal of Geophysical Research: Biogeosciences*. Wiley Online Library, 116(G4).
Simard, M. *et al.* (2019) 'Mangrove canopy height globally related to precipitation, temperature and cyclone frequency', *Nature Geoscience*, 12(1), pp. 40–45. doi: 10.1038/s41561-018-0279-1.
Slik, J. W. F. *et al.* (2010) 'Environmental correlates of tree biomass, basal area, wood specific gravity and stem density gradients in Borneo's tropical forests', *Global Ecology and Biogeography*. John Wiley & Sons, Ltd, 19(1), pp. 50–60. doi: https://doi.org/10.1111/j.1466-8238.2009.00489.x.
Somvanshi, S. S. and Kumari, M. (2020) 'Comparative analysis of different vegetation indices with respect to atmospheric particulate pollution using sentinel data', *Applied Computing and Geosciences*,

7, p. 100032. doi: https://doi.org/10.1016/j.acags.2020.100032.
Song, C. *et al.* (2010) 'Estimating average tree crown size using spatial information from Ikonos and QuickBird images: Across-sensor and across-site comparisons', *Remote Sensing of Environment*, 114, pp. 1099–1107. doi: 10.1016/j.rse.2009.12.022.
Sovdat, B. *et al.* (2019) 'Natural color representation of Sentinel-2 data', *Remote Sensing of Environment*, 225, pp. 392–402. doi: https://doi.org/10.1016/j.rse.2019.01.036.
Spiegelhalter, D. J. *et al.* (2002) 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), pp. 583–639. doi: https://doi.org/10.1111/1467-9868.00353.
Stegen, J. C. *et al.* (2011) 'Variation in above-ground forest biomass across broad climatic gradients', *Global Ecology and Biogeography*. Wiley Online Library, 20(5), pp. 744–754.
Stocker, T. F. (2013) 'The closing door of climate targets', *Science*. American Association for the Advancement of Science, 339(6117), pp. 280–282.
Stoyan, D. (2014) 'Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, 3rd edition. P. J. Diggle (2013). Boca Raton: Chapman & Hall/CRC Monographs on Statistics and Applied Probability. 267 pages, ISBN: 978-1-4665-6023-9.', *Biometrical Journal*. John Wiley & Sons, Ltd, 56(4), p. 720. doi: https://doi.org/10.1002/bimj.201400024.
Su, Y. *et al.* (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, pp. 187–199. doi: https://doi.org/10.1016/j.rse.2015.12.002.
Surano, K. A. *et al.* (1986) 'Growth and physiological responses of Pinus ponderosa Dougl ex P. Laws. to long-term elevated CO2 concentrations', *Tree Physiology*, 2(1-2–3), pp. 243–259. doi: 10.1093/treephys/2.1-2-3.243.
Surveyor General (2018) *Revised topo maps in Zimbabwe*. Harare.
Takagi, K. *et al.* (2015) 'Forest biomass and volume estimation using airborne LiDAR in a cool-temperate forest of northern Hokkaido, Japan', *Ecological Informatics*, 26, pp. 54–60. doi: https://doi.org/10.1016/j.ecoinf.2015.01.005.
Timmons, D. S., Buchholz, T. and Veeneman, C. H. (2016) 'Forest biomass energy: assessing atmospheric carbon impacts by discounting future carbon flows', *GCB Bioenergy*. John Wiley & Sons, Ltd, 8(3), pp. 631–643. doi: https://doi.org/10.1111/gcbb.12276.
Titeux, N. *et al.* (2016) 'Biodiversity scenarios neglect future land-use changes', *Global Change Biology*. John Wiley & Sons, Ltd, 22(7), pp. 2505–2515. doi: https://doi.org/10.1111/gcb.13272.
Tonolli, S. *et al.* (2011a) 'Fusion of airborne LiDAR and satellite multispectral data for the estimation of timber volume in the Southern Alps', *Remote Sensing of Environment*. Elsevier Inc., 115(10), pp. 2486–2498. doi: 10.1016/j.rse.2011.05.009.
Tonolli, S. *et al.* (2011b) 'Fusion of airborne LiDAR and satellite multispectral data for the estimation of timber volume in the Southern Alps', *Remote Sensing of Environment*, 115(10), pp. 2486–2498. doi: https://doi.org/10.1016/j.rse.2011.05.009.

Tran, T. V, Reef, R. and Zhu, X. (2022) 'A Review of Spectral Indices for Mangrove Remote Sensing', *Remote Sensing*. doi: 10.3390/rs14194868.
Traore, M. and Tieguhong, J. (2018) *How Forestry contributes to the African Development Bank's 'High 5s' Priorities for Africa: challenges and opportunities*.
Tveito, O. E., Wegehenkel, M. and Wel, F. van der (2008) 'The use of geographic information systems in climatology and meteorology'. Luxembourg: Publications Office of the European Union.
UN-REDD (2009) *The UN-REDD programme, http://www.un-redd.org*. Available at: http://www.un-redd.org (Accessed: 1 June 2022).
UNFCCC (2020) *Yearbook of Global Climate Action 2020*. London: Marrakech Partnership for Global Climate Action. Available at: https://unfccc.int/sites/default/files/resource/2020_Yearbook_final_0.pdf.
Vieilledent, G. *et al.* (2016) 'Bioclimatic envelope models predict a decrease in tropical forest carbon stocks with climate change in Madagascar', *Journal of Ecology*. John Wiley & Sons, Ltd, 104(3), pp. 703–715. doi: https://doi.org/10.1111/1365-2745.12548.
van Vuuren, D. P. *et al.* (2011) 'The representative concentration pathways: an overview', *Climatic Change*, 109(1), p. 5. doi: 10.1007/s10584-011-0148-z.
van Vuuren, D. P. and Carter, T. R. (2014) 'Climate and socio-economic scenarios for climate change research and assessment: reconciling the new with the old', *Climatic change*. Springer, 122, pp. 415–429.
Wackernagel, H. (2003) *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media.
Wai, P., Su, H. and Li, M. (2022) 'Estimating Aboveground Biomass of Two Different Forest Types in Myanmar from Sentinel-2 Data with Machine Learning and Geostatistical Algorithms', *Remote Sensing*. doi: 10.3390/rs14092146.
Walvoort, D. J. J., Brus, D. J. and de Gruijter, J. J. (2010) 'An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means', *Computers & Geosciences*, 36(10), pp. 1261–1267. doi: https://doi.org/10.1016/j.cageo.2010.04.005.
Wang, Q. *et al.* (2020) 'Comparative Analysis of Landsat-8, Sentinel-2, and GF-1 Data for Retrieving Soil Moisture over Wheat Farmlands', *Remote Sensing* . doi: 10.3390/rs12172708.
Wang, X., Chen, F. and Dong, Z. (2006) 'The relative role of climatic and human factors in desertification in semiarid China', *Global Environmental Change*, 16(1), pp. 48–57. doi: https://doi.org/10.1016/j.gloenvcha.2005.06.006.
Wang, Y. *et al.* (2022) 'Daytime and nighttime temperatures exert different effects on vegetation net primary productivity of marshes in the western Songnen Plain', *Ecological Indicators*. Elsevier, 137, p. 108789.
Webster, R. *et al.* (2006) 'Estimating the spatial scales of regionalized variables by nested sampling, hierarchical analysis of variance and residual maximum likelihood', *Computers & Geosciences*, 32(9), pp. 1320–1333. doi: https://doi.org/10.1016/j.cageo.2005.12.002.

Wheeler, D. C. and Waller, L. A. (2009) 'Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests', *Journal of Geographical Systems*, 11(1), pp. 1–22. doi: 10.1007/s10109-008-0073-5.
Whitlow, T. (1998) *Land degradation in Zimbabwe. A geographical study*. Harare.
Wickham, J. *et al.* (2015) 'Combining NLCD and MODIS to create a land cover-albedo database for the continental United States', *Remote Sensing of Environment*. Elsevier, 170, pp. 143–152.
Wu, C. *et al.* (2016) 'Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery', *Journal of Applied Remote Sensing*, 10(3), p. 35010. doi: 10.1117/1.JRS.10.035010.
Xiong, Y. and Wang, H. (2022) 'Spatial relationships between NDVI and topographic factors at multiple scales in a watershed of the Minjiang River, China', *Ecological Informatics*, 69, p. 101617. doi: https://doi.org/10.1016/j.ecoinf.2022.101617.
Yao, Y., Piao, S. and Wang, T. (2018) 'Future biomass carbon sequestration capacity of Chinese forests', *Science Bulletin*, 63(17), pp. 1108–1117. doi: https://doi.org/10.1016/j.scib.2018.07.015.
Zhang, R. P. *et al.* (2023) 'Inversion models of aboveground grassland biomass in Xinjiang based on multisource data ', *Frontiers in Plant Science* . Available at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1152432.
Zomer, R. J. *et al.* (2014) 'Projected climate change impacts on spatial distribution of bioclimatic zones and ecoregions within the Kailash Sacred Landscape of China, India, Nepal', *Climatic change*. Springer, 125, pp. 445–460.
Zunguze, A. X. (2012) 'Quantificação de carbono sequestrado em povoamentos de eucalyptus spp na floresta de Inhamacari-Manica'. Universidade Eduardo Mondlane.
Zvobgo, L. and Tsoka, J. (2021) 'Deforestation rate and causes in Upper Manyame Sub-Catchment, Zimbabwe: Implications on achieving national climate change mitigation targets', *Trees, Forests and People*, 5, p. 100090. doi: https://doi.org/10.1016/j.tfp.2021.100090.