Retrieving Plant Secondary Metabolites Across Biomes using remotely sensed data and Radiative Transfer Models

Bongokuhle S'phesihle Sibiya

215035638

A dissertation submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy in Environmental Sciences, in the School of Agriculture and Science, University of KwaZulu-Natal

Supervisor: Professor Onisimo Mutanga

Co-supervisor: Professor John Odindi

Co-supervisor: Professor Moses Azong Cho

Co-supervisor: Doctor Cecilia Masemola

Pietermaritzburg, South Africa

NOVEMBER 2025

Abstract

Secondary metabolites are diverse defence organic chemical compounds produced by plants against biotic and abiotic environmental stresses. Examples of secondary metabolites include terpenes, phenolic and nitrogenous compounds. Because these metabolites absorb electromagnetic energy in specific wavelengths within the 400 to 2500 nm spectral range, they can be detected non-destructively using remote sensing. Methods for quantifying plant secondary metabolites are generally divided into empirical and physical approaches. Empirical methods rely on finding the statistical relationship between plant attributes and remote sensing. Although they have been praised for producing desirable results over the past decades, they are trained and tested on a specific geography or dataset, sensor-specific, hence tend to lack generalization and transferability. On the other hand, physical models, such as radiative transfer models (RTMs) simulate the interactions of light with plant canopies, making them more robust across different seasons, locations, and sensor types. Despite their potential, most RTM applications focus on retrieving more common plant attributes like Leaf Area Index (LAI) and chlorophyll, particularly in agricultural, grassland, and forest landscapes. There is a notable gap in the literature regarding the use of RTMs to directly estimate secondary metabolites such as flavonoids and carotenoids, especially in ecosystems rich in phenolic compounds like the fynbos biome. Hence, this study aimed to address this gap by exploring how remotely sensed data and RTMs can be used to retrieve plant secondary metabolites across various biomes.

The first objective examined the application of RTMs on different terrestrial ecosystems for estimating both biophysical and biochemical properties. The review highlights that most studies have employed 1D models such as PROSPECT and PROSAIL, while for more complex 3D models they have used DART and FLIGHT. Furthermore, most of the existing research has utilized sensors like MODIS and Landsat on parameters like LAI and chlorophyll through inversion techniques such as look-up tables (LUT). However, research efforts are concentrated mainly in the Global North, with limited studies conducted in regions like Africa. The review reveals that PROSPECT and PROSAIL have remained the dominant RTMs over the past twenty years.

Building upon this, the second objective used the PROSPECT-5 model to estimate secondary metabolites such as carotenoid content using different wavelengths at the leaf level across three datasets (i.e., savanna, tropical forest, and a combination). Results showed that the green spectral region (510–580 nm) provided the best performance for estimating carotenoids, with R^2 values up to 0.90 and RMSE of 0.71 in the tropical forest dataset. Wavelengths selected by

the Random Forest algorithm, particularly in the green, yellow, and red-edge regions, improved estimation accuracy, especially in the savanna dataset, with R^2 values exceeding 0.80. The visible spectrum (400–700 nm) also played a crucial role in the detection, emphasizing targeted wavelength selection for better model performance. These findings highlight the importance of focusing on specific wavelengths rather than the full spectrum.

The third objective, also at the leaf level, restricted the wavelengths in the form of vegetation indices to study the relationship between fire and secondary metabolites like phenolics and flavonoids in the phenolic-rich fynbos biome. The data revealed that areas burned in2015, 2016, and 2021 had higher phenolic levels, compared to recently burned areas, indicating fire's role in stimulating secondary metabolite production and ecosystem regeneration. Spectral techniques such as continuum removal identified significant signals around 1660 nm for phenols and 415 nm for flavonoids. Vegetation indices like FI415,710 and FI420,710 showed correlations ($R^2 \approx 0.55$) with phenolic content, highlighting the importance of specific wavelengths.

The fourth objective tested whether including secondary metabolites in the PROSAIL model improves canopy-level estimates of plant cover in the phenolic-rich fynbos region. The results demonstrated that the PROSAIL-D model, which accounts for secondary metabolites, outperformed the traditional PROSAIL-4 model, achieving R^2 values around 0.72–0.78 and lower NRMSE errors (approximately 3.86–3.99%) when using AVIRIS and field spectral data dataset. This emphasizes the importance of secondary metabolites in the PROSAIL-D model to effectively simulate areas rich in phenolic compounds.

Finally, the fifth objective used the PROSAIL-D model to estimate flavonoid content using AVIRIS hyperspectral data. The vegetation index $FI_{417,693}$ yielded the highest accuracy, with an R^2 of 0.63 and RMSE of 0.24 mg CAE/g. This demonstrated the potential of using narrow spectral bands and targeted indices for reliable retrieval of flavonoids, even within complex canopies. The study underscores that focusing on specific spectral features enhances the capacity of remote sensing to effectively quantify plant biochemical traits.

In conclusion, this thesis demonstrates that RTMs, especially when incorporating secondary metabolites such as carotenoids and anthocyanins, are valuable tools for monitoring plant biochemical and biophysical properties across diverse ecosystems. Targeted spectral band selection and advanced modelling approaches can significantly improve the accuracy of remote

sensing estimates. While most studies are concentrated in developed regions, expanding research into underrepresented areas like Africa is essential for global ecology.

Keywords: Radiative Transfer Model (RTM), PROSPECT, PROSAIL, secondary metabolites, hyperspectral image, and biomes.

Preface

This study was conducted in the School of Agriculture and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa, under the supervision of Professor Onisimo Mutanga, Professor John Odindi, Professor Moses Azong Cho, and Doctor Cecilia Masemola from January 2022 to November 2025. I declare that the current work represents my own ideas and has never been submitted to any other academic institutions. Acknowledgment has been duly made for statements originating from other authors.

Bongokuhle S'phesihle Sibiya [redacted] Signed Date 03/11/2025

1. Professor Onisimo Mutanga (Supervisor) Signed [redacted] Date: 27/10/2025
2. Professor John Odindi (Co-Supervisor) Signed..... [redacted] Date: 10/11/2025
3. Professor Moses Cho (Co-Supervisor) Signed..... [redacted] Date…10/11/2025……..
4. Doctor Cecilia Masemola (Co-Supervisor) Signed [redacted] Date: 10/11/2025

Declaration 1: Plagiarism I

I Bongokuhle S'phesihle Sibiya declare that:

1. The research reported in this dissertation is my original work unless otherwise indicated.
2. This dissertation has not been submitted for the attainment of a degree or examination purposes at another university.
3. This dissertation does not contain any data, graphics, and other information from other persons unless duly acknowledged.
4. This dissertation does not contain other persons' writings unless duly acknowledged as such. In cases where written sources have been cited;
   a. Their words have been paraphrased and general information attributed to them has been referenced.
   b. Where exact words have been used, they were placed inside quotation marks and referenced.
5. This dissertation does not contain text, graphics, and or tables directly copied and pasted from the internet unless otherwise sources were duly acknowledged within the content of this dissertation.

Signed…… Date…03/11/2025…….

Declaration 2: List of Publications and Manuscripts

1. Sibiya, B.S., Odindi, J., Mutanga, O., Cho, M.A. and Masemola, C., 2025. The utility of radiative transfer models (RTM) on remotely sensed data in retrieving biophysical and biochemical properties of terrestrial biomes: A systematic review. *Advances in Space Research,* 75 (10), 7424-7444.
2. Sibiya, B., Cho, M., Mutanga, O., Odindi, J., Masemola, C., and Bonnet, W., 2025. Determining the value of different wavelength ranges of non-imaging hyperspectral reflectance to estimate carotenoid content using the PROSPECT-5 model. *International Journal of Remote Sensing*, *46*(4), pp.1696-1719.
3. Sibiya, B., Cho, M., Mutanga, O., Odindi, J., Masemola, C., Van Staden, J. and Vambe, M., 2025. Leaf spectroscopy reveals post-fire phenolic chrono sequence in the Fynbos Biome. *International Journal of Remote Sensing*, Under Review, Manuscript ID: TRES-PAP-2025-1611 (252838703)
4. Sibiya, B., Cho, M., Mutanga, O., Odindi, J., Masemola, C., and Wessel, B., 2025. Inclusion of secondary metabolites in PROSAIL RTM improves remote sensing of PAI in a phenolic-rich landscape. *Journal of Photogrammetry, Remote Sensing and Geoinformation Science,* Under Review, Manuscript ID: ef090ca2-acdb-439d-b8b7-dd5f87f25029
5. Sibiya, B., Cho, M., Mutanga, O., Odindi, J., Masemola, C. and Van Staden, J., 2025. Estimating flavonoids using radiative transfer model inversion on imaging spectroscopy data. *Frontiers in Remote Sensing*, Under Review, Manuscript ID: 1701367

Dedication

I dedicate this thesis to God almighty, the giver of life and strength, and to my amazing parents and loved ones.

Acknowledgment

You make all things work together for my good! You, who once appeared to Moses and declared, "*I am Jehovah; nothing is too hard for me*," have been my source of strength, guidance, and providence. First and foremost, I would like to give all Honor and Glory to You, **Lord Jesus**! Without Your blessings, divine direction, and unconditional love, none of this would have been possible.

I would also like to express my deepest gratitude to the School of Agriculture and Science and the Discipline of Geography at the University of KwaZulu-Natal. I am sincerely thankful for the opportunity to pursue my PhD and for your support throughout this journey.

My heartfelt appreciation goes to my supervisors: **Professor Onisimo Mutanga**, **Professor Moses Cho, Professor John Odindi**, and **Dr. Cecilia Masemola**. Your unwavering dedication, tireless commitment, and expert guidance have been instrumental in ensuring the successful completion of this project. I am grateful for your constructive scientific criticism, patience, and professional expertise, which have greatly enriched this work. Thank you for your continuous support, valuable advice, and passionate mentorship since the beginning of this journey, your encouragement has been vital from fund acquisition and training at the CSIR, which has significantly enhanced my remote sensing knowledge.

I also wish to acknowledge the invaluable contributions of the entire BioScape team, supported by NASA grant 80NSSC21K0086, the South African government (NRF/SAEON), and the United Nations Educational, Scientific and Cultural Organization (UNESCO). My gratitude extends to the University of the Western Cape (UWC) for granting access to their facilities, especially to Professor Timothy Dube, and Mmasechaba Lebogang Moropane, Odirile Kgothatso Mamba, and Sibuyisele Pakati for their assistance with data collection. I am also thankful to Dr. Kerry Cawse-Nicholson and Dr. Luciana F. Alves for their support during data collection.

Furthermore, I appreciate the University of Cape Town (UCT) for allowing me to use their oven for sample drying, with special thanks to Dr. Jasper Slingsby and Cherie Forbes. The collaborative efforts and support from the entire BioScape team have been invaluable. This research was made possible through the support of NASA, the South African government via NRF/SAEON, and UNESCO. We are grateful for the AVIRIS-NG imagery provided by NASA, and for the support and access to the Greater Cape Floristic Region (GCFR) from Sandberg Nature Reserve and Smith Farm SANPARKS. Our sincere thanks also go to the CSIR for their

warm hospitality during my visit as guest scientists, and to team members—Dr. Sabelo Madonsela, Dr. Nobuhle Majozi, Yonwaba Atyosi, Yoliswa Mkhize, Adri Alberts—and Wessel Bonnet, whose deep knowledge of RTMs was instrumental.

I also wish to acknowledge the support of the National Research Foundation (NRF) through the NEO Frontiers innovation grant (Grant no. NEO960912 SIBIYA), Extension Support (PMDS240612226925) and the SARChI Grant no. 84157. Special thanks go to Professor Johannes Van Staden and Dr. McMaster Vambe for graciously allowing us to utilize their laboratory facilities for the wet analysis of secondary metabolites.

To my dearest friends and colleagues, your words and unwavering support kept me motivated and resilient during challenging times. I am particularly grateful to Wandile Khumalo, Sibongakonke Hadebe, Sisanda Mkhize, Mpho Letseka, Amanda Nyawose, the UKZN PMB boxing team, Celuxolo Dlamini, Bheka Mlambo, Ntuthuko Mncwabe, Sifundo Mthiyane, Luthando Ngwane, Mbali Manyaka, Siphiwokuhle Buthelezi, Wendy Khumalo, Snokuthemba Mkhize, Anita Masenyama, Dr. Trylee Matongera, Sandiswa Gwala, Slindile Qwabe, Ananda Mbanjwa, Nosipho Ntshuntshe, Phindile Mthembu, Dr. Odebiri Omosalewa, Dr. Mthembeni Mngadi, and Mr. Brice Gijsbertsen. Special thanks to Mr. Afika Singano for working alongside me in data collection, analysis, and lab work, as well as providing constructive criticism. I also extend my gratitude to Scottsville Presbyterian Church for their hospitality, especially Fiona and Mairi, our gracious hosts.

Lastly, my heartfelt gratitude goes to my family—Mr. and Mrs. Sibiya, Mncedisi, Bongeka, Phumelele, Ayabonga, Aphelele, Sizwe, and Kwanele—for their unwavering love. God bless you all.

Table of Contents

Abstract ii
Preface v
Declaration 1: Plagiarism I vi
Declaration 2: List of Publications and Manuscripts vii
Dedication viii
Acknowledgment ix
Table of Contents xi
List of Tables xv
List of Figures xvi
Acronyms xix
Chapter One: 1
1.1. Introduction 1
1.2. Aims and objectives 6
1.3. Research questions 6
1.4. Study site description 6
1.5. Thesis structure 9
Chapter Two: 12
2.1. Introduction 13
2.2. Methodology 15
2.2.1. Phase 1: Literature search 15
2.2.2. Phase 2: Data extraction 17
2.2.3. Phase 3: Analysis 18
2.3. Results 19
2.3.1. Analysis of keywords frequency, growth, and co-occurrence 19
2.3.2. Global trend and distribution of RTM publications 20
2.3.3. Sensor datasets used 21
2.3.4. Vegetation types and biophysical and biochemical variables 22
2.3.5. Applications of RTM 24
2.3.6. The hybrid approaches 26
2.4. Discussion 27

2.4.1. Global distribution publication trends 27
2.4.2. Earth observation sensors 28
2.4.3. Biophysical and biochemical properties 31
2.4.4. Application of Radiative Transfer Model (RTM) 33
2.4.4.1. One-dimensional (1D) models 33
2.4.4.2. Three-dimensional (3D) models 34
2.4.5. Hybrid inversion 37
2.4.5.1. Look-up table (LUT) 37
2.4.5.2. Machine learning-based 37
2.4.5.3. Vegetation Indices 38
2.4.6. Findings, challenges, and prospects 39
2.5. Conclusion 41
2.6. Chapter summary 42
Chapter Three: 43
3.1. Introduction 44
3.2. Methods 47
3.2.1. Field survey 47
3.2.2. Spectral measurement and wet chemical analysis 48
3.2.3. Statistical Analysis 49
3.2.3.1. Random Forest Regression (RFR) 49
3.2.3.2. XGBoost 50
3.2.3.3. Partial Least Squares Regression (PLSR) 50
3.2.3.4. Artificial Neural Network (ANN) 50
3.2.4. Simulated datasets using the PROSPECT-5 model 51
3.3. Results 52
3.3.1. Algorithms prediction performance 52
3.3.2. Variable importance selection using the RFR 53
3.3.3. Results of PROSPECT-5 model for Savanna, tropical forest, and Combined datasets 54
3.3.3.1. Savanna 56
3.3.3.2. Tropical Forest 58
3.3.3.3. Combined 60

3.4. Discussion 62
3.5. Conclusion 66
3.6. Chapter summary 66
Chapter Four: 67
4.1. Introduction 68
4.2. Materials and methods 70
4.2.1. Field sampling 70
4.2.2. Leaf reflectance measurements 71
4.2.3. Biochemical analysis 71
4.2.4. Spectral data processing 72
4.2.5. Calculation of vegetation indices 73
4.2.6. Model validation and accuracy assessment 74
4.3. Results 74
4.3.1. Changes in phenolic content post-fire occurrence 74
4.3.2. Evaluation of leaf spectra 76
4.3.2.1. Comparison of the spectral characteristics 76
4.3.2.2. Influence of Leaf Area Index and Soil Moisture on Phenolic Compounds in Plants 77
4.3.3. Phenolic absorption features 78
4.3.4. Relationships between flavonoid and vegetation Indices 80
4.4. Discussion 82
4.4.1. Changes in the concentration of phenolic compounds 82
4.4.2. Absorption features of phenolics 83
4.4.3. Flavonoids Indices 84
4.4.4. Challenges and prospects 85
4.5. Conclusions 87
4.6. Chapter summary 87
Chapter Five: 88
5.1. Introduction 89
5.2. Methodology 91
5.2.1. Sampling 91
5.2.2. Hyperspectral images 93

5.2.3. PROSAIL calibration 94
5.3. Result 97
5.3.1. Statistical Analysis 97
5.3.2. Sensitivity analysis of PROSAIL-D parameters 98
5.3.3. Results of the PROSAIL models for field spectral data data and AVIRIS-NG image 99
5.3.4. Results of the PROSAIL model 101
5.4. Discussion 103
5.5. Conclusion 107
5.6. Chapter summary 108
Chapter Six: 109
6.1. Introduction 110
6.2. Materials and Methods 111
6.2.1. Data collection and chemical analysis 111
6.2.2. AVIRIS data and Flavonoid Indices 112
6.2.3. PROSAIL-D model 114
6.3. Results 117
6.3.1. Result of PROSAIL-D 117
6.3.2. Performance of Flavonoids indices 118
6.4. Discussion 120
6.4.1. Impact of AVIRIS-NG Image bandwidth on flavonoid estimation 120
6.4.2. The utility of the PROSAIL-D model 121
6.4.3. Assessment of flavonoid indices 121
6.4.4. Factors affecting the distribution of flavonoids in the GCFR 122
6.4.5. Limitations and potentials 124
6.5. Conclusion 125
6.6. Chapter summary 125
Chapter Seven: 126
7.1. Introduction 126
7.2. Challenges and limitations 128
7.3. Recommendation and prospects 129
References 131

List of Tables

Table 2.1: Key searched terms 17
Table 3.1: Leaf datasets 48
Table 3.2: Input parameters of PROSPECT 52
Table 4.1: Chemical Assay Procedures for Quantifying Total Phenol and Flavonoid..... 72
Table 4.2: Utilized Vegetation Indices..... 73
Table 4.3: Accuracy of various flavonoids spectral indices. 81
Table 5.1: Ranges for input parameters used for generating look-up table (LUT) using the PROSAIL model. The model parameter LAI represents PAI..... 94
Table 5.2: The descriptive statistics showing the Maximum (Max), Minimum (Min), Standard Deviation (Std), Range, and Coefficient of Variation (CV). 97
Table 6.1: Flavonoids-related indices. 113
Table 6.2: Forward mode ranges used in PROSAIL-D..... 115
Table 6.3: Relationships between simulated LAI and measured Flavonoids derived from Flavonoids indices. 119

List of Figures

Figure 1.1. Map of South Africa showing the relative locations of the three study sites (Kruger, Dukuduku, and Greater Cape Floristic Region (GCFR)). 8
Figure 2.1. Studies selected for review. 18
Figure 2.2. Topical concepts of RTM and biophysical and biochemical property retrieval from terrestrial biomes derived using data from titles, abstracts, and keywords. 20
Figure 2.3. Frequency of studies from year 2000 to 2023. 20
Figure 2.4. Global distribution of the selected published articles on the use of RTMs to retrieve plant attributes in terrestrial biomes. 21
Figure 2.5. The frequency of use for sensor type (sensors used only once are classified as other). 22
Figure 2.6. Vegetation types analyzed by the evaluated studies. The category 'natural' includes forest, grassland, Savanna, Tundra/boreal, shrubland, Taiga, and Chaparral biomes. 23
Figure 2.7. Biophysical and biochemical variables predicted using the RTMs ((leaf and canopy chlorophyll content (Cab, LCC and CCC), carotenoid content (Car), dry matter content (Cm or LMA), Equivalent water thickness or canopy/leaf water content (EWT, CWC, CAW, and Cw), specific leaf area (SLA) and fuel moisture content (FMC), above ground biomass (AGB), fraction of vegetation cover (fcover), fraction of absorbed photosynthetically active radiation (fAPAR), Nitrogen (N), Gross Primary Productivity (GPP), Solar-induced chlorophyll fluorescence (SIF), soil moisture content (SMC) and Vcmax (maximum velocity of carboxylation)). 24
Figure 2.8. RTMs used to plant traits in terrestrial biomes. 25
Figure 2.9. Applications of RTM models on different vegetation types. 26
Figure 2.10. Variable retrieval methods of all evaluated studies involving the RTMs 27
Figure 3.1. Predicted against observed carotenoids using various algorithms. 53
Figure 3.2. Results of various algorithms in introducing the most important wavelengths: (a) RFR, (B) XGBoost, (C) PLSR, and (D) ANN. 54
Figure 3.3. RMSE of the three datasets (savanna, tropical forest, and combined) between measured and simulated spectra using PROSPECT-5. 55
Figure 3.4. Spectral matching between simulated PROSPECT-5 and measured spectral, (first column) and differences between simulated and measured spectra (second column).... 56
Figure 3.5. Relationships between observed and predicted carotenoids (μg/cm^2) for various spectral regions and bands selected from RFR in the Savanna landscape, derived from PROSPECT-5 inversion. 58
Figure 3.6. Relationships between observed and predicted carotenoids in the Dukuduku Tropical Forest using different spectral regions and bands selected from RFR. 60
Figure 3.7. Relationships between observed and predicted carotenoids (μg/cm^2) for various spectral regions and bands selected from RFR for both Savanna landscape and tropical forest dataset, derived from PROSPECT-5 inversion. 62

Figure 4.1. Location of burnt areas between 2010 and 2024 using false colour composite: A = 2021 (May), B = Never_exp, C = 2016 (February), D = 2015 (March), E = 2016 (April), F = 2010 (January). 75
Figure 4.2. Variation in (A) total phenols and (B) flavonoids in five areas that had fires and areas that have not experienced fires from 2010 to 2024. 76
Figure 4.3. Comparing the (A) average reflectance and (B) spectral coefficient of variation of five areas that experienced fire with areas that had not experienced fire during this period. 76
Figure 4.4. The relationships of LAI and phenolic compounds (A-B) and relationships of soil moisture and phenolic compounds (C-D). 78
Figure 4.5. Continuum-removed spectral features for the hotspot region: (A) total phenol in the range of 1650–1670 nm, and (B) flavonoid in the range of 400–430 nm. A red dotted line highlights the lowest absorption. 79
Figure 4.6. The correlation between total phenolics concentration and the spectral feature at 1660 nm (A) and between flavonoids concentration and the spectral feature at 415 nm (B). 80
Figure 4.7. Top four best-performing flavonoid spectral indices. 82
Figure 5.1. A depiction of the sampling protocol. 92
Figure 5.2. field spectral data spectral reflectance (n = 136). Due to high noise regions from 1824 to 1954 nm were excluded. 93
Figure 5.3. The canopy reflectance spectra obtained from AVIRIS-NG data. Specific spectral regions, 377 to 397 nm, 1364 to 1409 nm, and 1805 to 1970 nm, have been omitted due to high levels of noise. 94
Figure 5.4. Shows 100’000 simulated spectra from PROSAIL-4 (A and C) and (B) PROSAIL-D (B and D), for field spectral data (A and B) and AVIRIS-NG (C and D). 96
Figure 5.5. First-order Global Sensitivity Index (A) and Total Global Sensitivity Index (B) for PROSAIL-D input parameters. 99
Figure 5.6. An example of the differences between the measured and simulated spectra from PROSAIL-4 (A and C) and (B) PROSAIL-D (B and D), for field spectral data (A and B) and AVIRIS-NG (C and D). 100
Figure 5.7. RMSE of simulated and measured field spectral data data (A) and AVIRIS-NG (B) with the PROSAIL-4 and PROSAIL-D. 100
Figure 5.8. Illustrate the relationship between predicted and measured PAI (the model parameter LAI represents PAI). (A) PROSAIL-4 vs measured PAI for field spectral data plots, (B) PROSAIL-D vs measured PAI for field spectral data plots, (C) PROSAIL-4 vs measured PAI for AVIRIS image, and (D) PROSAIL-D vs measured PAI for AVIRIS image. 102
Figure 5.9. PAI maps derived from (A) PROSAIL-4 and (B) PROSAIL-D model. 103
Figure 6.1. Depiction of the sampling protocol 112
Figure 6.2. Example of AVIRIS-NG spectrum (a) Raw; (b) After excluding the noisy bands. 113

Figure 6.3. The simulated spectra using PROSAIL-LUT based on parameters ranges (A) and example of measured vs. PROSAIL-simulated spectra (B). 115
Figure 6.4. The relationships between field LAI and flavonoids. 116
Figure 6.5. Average spectral RMSE for the various flavonoid indices tested..... 118
Figure 6.6. Accuracy of top three best performing Flavonoids Indices from PROSAIL-D RTM: (A) $FI_{417,693}$, (B) $FI_{417,713}$, and (C) $FI_{412,713}$..... 119
Figure 6.7. Flavonoids maps derived from PROSAIL-D- $FI_{417,693}$ index showing distribution pattern in the GCFR. 120

Acronyms

AGB	Above-Ground Biomass
$AlCl_3$	Aluminum Chloride
ALIA	Average Leaf Inclination Angle
ANGERS	Airborne Near-Ground Experimental Remote Sensing System
Angle	Leaf Angle Distribution
ANN	Artificial Neural Networks
Anth	Anthocyanin
ASD	Analytical Spectral Device
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
BDPT	Bidirectional Path Tracing
BRF	Bidirectional Reflectance Factor
Cab	Chlorophyll
CAE	Catechin
Car	Carotenoid
CASI	Compact Airborne Spectrographic Imager
cbrown	Brown Pigments
CCC	Canopy Chlorophyll Content
Cm or LMA	Dry Matter Content
DART	Discrete Anisotropic Radiative Transfer
Deg	Degrees
DDblue	Double Difference Indices – Blue
DDnir	Double Difference Indices – Near Infrared
EnMAP	Environmental Mapping and Analysis Program
EWT, CWC, CAW, and Cw	Equivalent Water Thickness or Canopy/Leaf Water Content
fAPAR,	Fraction Of Absorbed Photosynthetically Active Radiation
fcover	Fraction Of Vegetation Cover
FI	Flavonoid Index
FLEX	Fluorescence Explorer
FLIGHT	Forest Light Interaction Model
fmc	Fuel Moisture Content
FPAR	Fraction Of Absorbed Photosynthetically Active Radiation
g	Gram
GCFR	Greater Cape Floristic Region
GEO	Group On Earth Observations
GIS	Geographic Information Systems
GPP	Gross Primary Productivity
GPR	Gaussian Process Regression
GPS	Global Positioning System
GPU	Graphics Processing Units
GSA	Global Sensitivity Analysis

GT	Gigatons
HyspIRI	Hyperspectral Infrared Imager
ISI	Institute Of Science Information
LAI	Leaf Area Index
LCC	Leaf Chlorophyll Content
LESS	Large-Scale Remote Sensing Data and Image Simulation Framework
Lidar	Light Detection and Ranging
LOPEX	Leaf Optical Properties Experiment
LUT	Look-Up Table
MC	Monte Carlo Raytracing
MeOH	Methanol
Mg	Microgram
Ml	Milliliter
mFI	Modified Flavonoid Index
MODIS	Moderate Resolution Imaging Spectroradiometer
N	Leaf Structure
N	Nitrogen
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
never_exp	Never Experience Fire
nm	Nanometre
NRMSE	Normalized Root Mean Square Error
ORNL DAAC	Oak Ridge National Laboratory Distributed Active Archive Centre
PAI	Plant Area Index
PCA	Principal Component Analysis
PCR	Principal Components Regression
PLSR	Partial Least Squares Regression
PRISMA	Precursore Iperspettrale Della Missione Applicativa
psi	Relative Azimuth Angle
PSM	Plant Secondary Metabolites
Psoil	Soil Coeffient
QGIS	Quantum Geographic Information System
R^2	Coefficient Of Determination
RFR	Random Forest Regression
RMSE	Root Mean Square Error
RR	Ridge Regression
RRMSE	Normalized Root Mean Square Error
RTM	Radiative Transfer Models
SAIL	Scattering By Arbitrarily Inclined Leaves
SANParks	South African National Parks
SAVI	Soil-Adjusted Vegetation Index
SCI	Science Citation Index
SCOPE	Soil Canopy Observation, Photochemistry and Energy Fluxes
sif	Solar-Induced Chlorophyll Fluorescence
Skyl	Diffuse/Direct Light

SLA	Specific Leaf Area
SMC	Soil Moisture Content
SMLR	Stepwise Multiple Linear Regression
SNAP	Sentinel Application Platform
SPAD	Soil Plant Analysis Development
SR	Simple Ratio
SSCI	Social Sciences Citation Index
SVM	Support Vector Machines
SWIR	Short-Wave Infrared
tts	Solar Zenith Angle
UAV	Unmanned Aerial Vehicles
USA	United States of America
UV	Ultraviolet
Vcmax	Maximum Velocity of Carboxylation
VI	Vegetation Indices
W_d	Dry Weight
W_f	Fresh Weight
WOS	Web Of Science
WPS	Weighted Photon Spread
µL	Microlitre
1D	One-Dimensional
3D	Three-Dimensional

Chapter One:
General Introduction

1.1. Introduction

Biomes and plant biodiversity are critical components of the global environment (Schimel et al., 2015; Odebiri et al., 2023). Typically, they comprise large naturally occurring ecological zones with diverse species that provide vital ecosystem goods and services such as global food security, herbal and pharmaceutical medicine, human and environmental health, sustainable development, and climate change mitigation (Schimel et al., 2015; Mngadi et al., 2022). Globally, biomes store more than 2000 gigatons (Gt) of carbon, making them a major driver of the global carbon cycle (Holmberg and Hellsten, 2016; Masson-Delmotte et al., 2021). However, anthropogenic activities such as changes in land use, nutrient enrichment, and habitat fragmentation have triggered the sixth major extinction catastrophe in the history of life, posing a massive existential threat to global plant biodiversity and associated benefits (Chapin III et al., 2000).

Due to the risks associated with high extinction rates, it is essential to carry out proper monitoring, modelling, and inventories in a timely and continuous manner. This ensures the provision of rapid and reliable information critical for management and policy decisions aimed at preventing further losses and restoring ecosystems (Silva et al., 2019). Additionally, the United Nations Decade of Ecosystem Restoration, launched in 2021, emphasized the importance of collective efforts by nations, governments, non-governmental organizations, and scientists to revive degraded ecosystems (Nelson et al., 2024). A central requirement for preventing losses and restoring plant biodiversity is the quantification and modelling of its functional diversity, measured using plant attributes (Ali et al., 2016; Zhang et al., 2012).

As a fundamental component of functional traits, secondary metabolites significantly influence ecosystem dynamics, mediate species interactions, and bolster ecological resilience (Tilman, 2001). Plant secondary metabolites are biochemical constituents produced by numerous plant species to adapt to various biotic (e.g., insects, bacteria, viruses) and abiotic (e.g., extreme temperatures, drought, flooding, light intensity, salinity, heavy metals) stressors (Chomel et al., 2016). Unlike animals, which can often evade threats, plants are permanently rooted in their environment and must continuously adapt to a wide spectrum of stresses, ranging from microscopic fungi and bacteria to larger herbivores (Dixon, 2001). These herbivores and

pathogens seek to exploit plants' nutrients and water within leaves, stems, fruits, and seeds. In response, plants have evolved an array of internal (biochemical) and external (biophysical) defence mechanisms that render them less palatable or even toxic, thereby reducing their susceptibility to predation and infection (Hiruma, 2019).

In addition to their importance in functional diversity, plants have also been shown to possess a wide range of medicinal properties that are of significant interest to human health, including potential anti-diabetic effects (Zhou et al., 2022; 2023). Hence, many plants rich in secondary metabolites are often adopted as drugs, medicines, and seasonings, because they are antimicrobial or insecticidal (Şenkal, 2020). Therefore, retrieving plant secondary metabolites can describe the capacity of a species to develop, reproduce, or survive in this harsh environment (Wu et al., 2020). Consequently, information on the distribution of secondary metabolites can enhance our knowledge and ability to study community substructures and ecological functioning for proper management and policy.

Determining the secondary metabolites in plant tissues can be a challenge. Conventional chemical analysis techniques for quantifying secondary metabolites, including high-performance liquid chromatography and ultraviolet spectrophotometry (Dong et al., 2005; Solovchenko et al., 2001), involve destructive sampling and laboratory analysis, are labour-intensive, expensive, and time-consuming. This makes them impractical for assessing plant phenolics over large spatial extents.

Fortunately, advances in remote sensing technology offer a more efficient and cost-effective alternative for assessing secondary metabolites. Remote sensing enables the acquisition of spatially and temporally continuous data, facilitating the mapping of secondary metabolite distribution across large areas (Skidmore et al., 2010; Xie et al., 2024; 2025). Additionally, secondary metabolites exhibit strong absorption features across various spectral wavelengths within the 400 to 2500 nm range, enabling their detection through non-destructive remote sensing techniques. For example, total phenols absorb light predominantly between 1620 and 1690 nm (Xie et al., 2024), while flavonoids strongly absorb shortwave blue light in the 410–430 nm range (Gitelson et al., 2017). Similarly, carotenoids absorb light primarily within the visible spectrum (Asner, 1998). These distinct absorption characteristics highlight the potential of remote sensing to assess and monitor secondary metabolite levels across different plant species and ecosystems, providing valuable insights into plant health, stress responses, and adaptive strategies.

Various methods have been employed to retrieve secondary metabolites from remote sensing data. These methods can be broadly classified into two categories: statistical approaches and physical approaches. In the statistical approach, either parametric or non-parametric regression is used to establish relationships between spectral data and specific variables (Verrelst et al., 2015). Vegetation indices are commonly utilized to enhance the correlation between a particular variable and spectral observations. However, indices such as the Normalized Difference Vegetation Index (NDVI) tend to saturate at high biomass densities (Mutanga and Skidmore, 2004). To strengthen the empirical connection between the variable of interest and spectral data, whether through derivatives, vegetation indices, or texture analysis, various regression models have been developed and are extensively documented in the literature (Dorigo et al., 2007).

According to Ali et al. (2021), the most widely used models include Stepwise Multiple Linear Regression (SMLR), Partial Least Squares Regression (PLSR), Principal Components Regression (PCR), Ridge Regression (RR), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Gaussian Process Regression (GPR), Bayesian Networks, and Random Forest Regression (RFR). These models are popular due to their simplicity and accessibility. However, they typically require in situ datasets for training and validation, which presents a significant limitation. Additionally, these methods are often site-specific, lacking generalization capabilities, and are not easily adaptable to different locations with varying vegetation types or sensor characteristics (Dorigo et al., 2007; Verrelst et al., 2015; 2019). Moreover, they tend to be computationally intensive and often perform poorly when applied to complex spectral datasets derived from hyperspectral or multispectral sensors (Vuolo et al., 2013).

In contrast, physically based approaches, such as RTMs, are considered more robust as they explain the cause-effect relationship (Verrelst et al., 2019). These models describe how sunlight interacts with vegetation by explaining the absorption, scattering, and emission of light using fundamental physical principles (Jacquemoud et al., 2009). According to Ali et al. (2021), RTMs simulate the reflectance of a landscape as seen by a sensor, considering various factors such as leaf properties, canopy structure (e.g., stand density, crown architecture, heterogeneity), illumination geometry, sensor configuration, and background conditions.

Over the years, numerous RTMs have been developed, each tailored to different levels of complexity and specific applications within remote sensing (Verrelst et al., 2015). These range

from simple one-dimensional (1D) models to more complex three-dimensional (3D) Monte Carlo ray tracing models, (see Widlowski et al., 2015, for a comprehensive comparison).

The PROSPECT model family is among the most widely utilized leaf-level RTM models (Sibiya et al., 2025a). Over time, this model has undergone successive updates and refinements aimed at improving its accuracy and applicability in simulating leaf optical properties (Lu et al., 2021). PROSPECT-4 represents an early iteration characterized by the inclusion of four key parameters: leaf structure (N), chlorophyll content, water content, and dry matter content (Feret et al., 2008). These parameters allow for effective simulation of typical photosynthetic vegetation under relatively stable and unstressed conditions.

Building upon PROSPECT-4, the subsequent introduction of PROSPECT-5 by Feret et al. (2008) expanded the model's parameter set to encompass additional pigment-related features, including brown pigments and carotenoids. This enhancement improved the model's ability to simulate more nuanced variations in leaf reflectance associated with different pigmentation levels, which are critical for understanding plant health and stress responses (Feret et al., 2008; Lu et al., 2021). Nonetheless, PROSPECT-4 and PROSPECT-5 are primarily optimized for healthy, photosynthetically active vegetation and may exhibit limitations when applied to environments characterized by complex secondary metabolites, senescence, or stress-induced physiological changes (Darvishzadeh et al., 2008).

To address these limitations, Feret et al. (2017) proposed PROSPECT-D, an advanced version that incorporated anthocyanins, a class of flavonoids associated with stress responses such as drought, aging, and nutrient deficiency, into the existing parameter framework of PROSPECT-5. The inclusion of anthocyanins enhanced the model's capacity to simulate vegetation under a broader range of environmental and physiological conditions, including stressful environments, different stages of vegetation development (juvenile and senescent phases), and environmental stressors that induce pigment accumulation (Lu et al., 2021; Feret et al., 2017).

Given that each iteration of the PROSPECT model involves distinct parameters and calibration datasets, their comparative performance must be evaluated across diverse biomes and environmental conditions. Such testing is crucial for determining the robustness, flexibility, and accuracy of these models in various ecological contexts (Jacquemoud et al., 2009). Overall, the progressive developments from PROSPECT-4 to PROSPECT-D reflect ongoing efforts to refine leaf optical property simulations, enabling more precise remote sensing applications and

advancing our understanding of plant physiology under varying environmental stresses (Feret et al., 2008; 2017).

At the canopy level, a widely adopted approach involves the integration of leaf optical models with canopy RTM to improve the simulation and retrieval of vegetation properties. One prominent example of this integrated modelling framework is PROSAIL, which combines PROSPECT (Jacquemoud and Baret, 1990) with SAIL (Scattering by Arbitrarily Inclined Leaves), a canopy RTM (Verhoef, 1985). PROSAIL leverages the detailed leaf optical properties provided by PROSPECT and extends them to the canopy level through SAIL, enabling the simulation of reflectance for entire plant stands.

PROSAIL has been extensively employed in numerous studies to estimate vital biophysical and biochemical parameters of vegetation, such as chlorophyll concentration, LAI, water content, and dry matter (Singh et al., 2025; Farmonov et al., 2025; Zhao et al., 2023; Darvishzadeh et al., 2008; 2019; Lunagaria and Patel, 2019). The theoretical review of PROSAIL, including its capability to retrieve plant attributes, has been elaborated by researchers like Jacquemoud et al. (2009) and Berger et al. (2018), who provided comprehensive frameworks for its application. Despite its widespread use, most research utilizing PROSAIL has focused on vegetation types such as grasslands, agricultural fields, and forests, primarily in environments where secondary metabolites are present in relatively low concentrations.

A notable gap in the current literature is the limited application of PROSAIL in environments rich in secondary metabolites such as flavonoids, carotenoids, and anthocyanins that play significant roles in plant stress responses, pigmentation, and ecological interactions (Sibiya et al., 2025b). These metabolites can substantially influence leaf and canopy reflectance, yet they are often overlooked in remote sensing models designed for more conventional vegetation types. Additionally, most studies prioritize the retrieval of basic plant attributes like LAI, chlorophyll, water, and dry matter content, with limited focus on secondary metabolites, which are crucial for understanding physiological status and environmental stress.

The present study aimed to bridge this gap by developing methodologies to retrieve plant secondary metabolites across various biomes using remotely sensed data integrated with advanced RTMs. By focusing on secondary metabolites such as flavonoids and carotenoids, this research seeks to enhance our capacity to monitor plant health, stress responses, and

adaptive strategies in diverse ecological contexts, thereby advancing the application of remote sensing in plant physiology and environmental monitoring.

1.2. Aims and objectives

The main aim of this research was to retrieve plant secondary metabolites across biomes, using remotely sensed data and Radiative Transfer Models. The objectives of the study were;

1. To review the utility of RTM on remotely sensed data in retrieving biophysical and biochemical properties of terrestrial biomes.
2. To determine the value of different wavelength ranges of non-imaging hyperspectral reflectance to estimate carotenoid content using the PROSPECT-5 model.
3. To investigate if leaf spectroscopy reveals post-fire phenolic chrono sequence in the Fynbos Biome.
4. To investigate if the inclusion of plant secondary metabolites in PROSAIL RTM improves remote sensing of Plant Area Index (PAI) in a phenolic-rich landscape.
5. To estimate flavonoids using PROSAIL-D inversion on imaging spectroscopy data.

1.3. Research questions

1. What is the effectiveness of radiative transfer models (RTMs) in concert with remotely sensed data to accurately retrieve biophysical and biochemical properties across various terrestrial biomes?
2. Can we effectively estimate carotenoids from field spectral data using PROSPECT-5 model and various spectral regions from the visible region?
3. Can leaf spectroscopy detect and characterize the post-fire phenolic chronological sequence within the Fynbos biome?
4. Does incorporating plant secondary metabolites into the PROSAIL model enhance the remote sensing of PAI in landscapes rich in phenolic compounds?
5. How accurately can flavonoid concentrations be estimated through PROSAIL-D inversion applied to the AVIRIS image?

1.4. Study site description

Kruger National Park (Savanna biome) is situated within the savanna biome (-24.88139° S, 31.398543° E) (Figure 1.1). It is the largest biome in Southern Africa, covering approximately 46% of the region's landscape. It is characterized by a herbaceous layer of various grass species and complemented by open tree canopies. Typically, the savanna undergoes a prolonged dry winter and a wet summer, with annual precipitation ranging from 235 to 1000 mm. The region of interest is primarily situated in the southern part of the Kruger National Park and encompasses two study regions; Sabi Sands Wildtuin and Bushbuckridge Municipal District.

The majority of Bushbuckridge is composed of granite soil types, while the Sabi Sands region still has gabbro intrusions ((see Cho et al. (2012) and Naidoo et al. (2012) spatial distribution of the landscape)).

Dukuduku indigenous coastal forest (Tropical Forest), is situated between Mtubatuba and St Lucia in the northern part of KwaZulu-Natal, South Africa (-28.41667° S, 32.28333 ° E) (Figure 1.1). The area covered approximately 6,000 hectares in the early 1950s, however, it dwindled due to significant deforestation. By 2011, its size had diminished to a mere 3,200 hectares, primarily due to conversion to farmlands and establishment of settlements (Ndlovu et al. 2011; Qabaqaba et al. 2023). The region receives consistent yearly rainfall, with an average annual precipitation of around 1,200 mm. The area was selected for investigation because it represents the largest remaining section of an indigenous forest ((see Cho et al. (2015) for the distribution of the study site)).

The Greater Cape Floristic Region (GCFR), situated within the Western Cape province of South Africa, specifically in the Overstrand Local Municipality, is an area recognized as one of the world's biodiversity hotspots owing to its high endemism and ecological diversity. The GCFR is known for its remarkable biodiversity, hosting nearly 10,000 plant species, with approximately 69% being endemic and found only in the specific region (Cardoso et al., 2025; Musungu et al., 2025; City of Cape Town Biodiversity Report, 2018). The research sites included Sandberg Nature Reserve (-34.577953° S, 19.693025° E), Smith Farm (-34.584551° S, 19.668460° E), and SANParks (South African National Parks) (-34.642366° S, 19.592418° E) (Figure 1.1). The region experiences a Mediterranean climate characterized by wet winters and dry summers. Approximately 80% of the annual precipitation occurs between April and September, with the dry season typically spanning November to March (Ndebele et al., 2022). The dominant vegetation across these locations consist primarily of *Proteaceae*, *Ericoid shrubs*, *Berzelia lanuginose*, *Erica species, Elegia mucronate, Platycaulos compressu,* and *Restionaceae*.

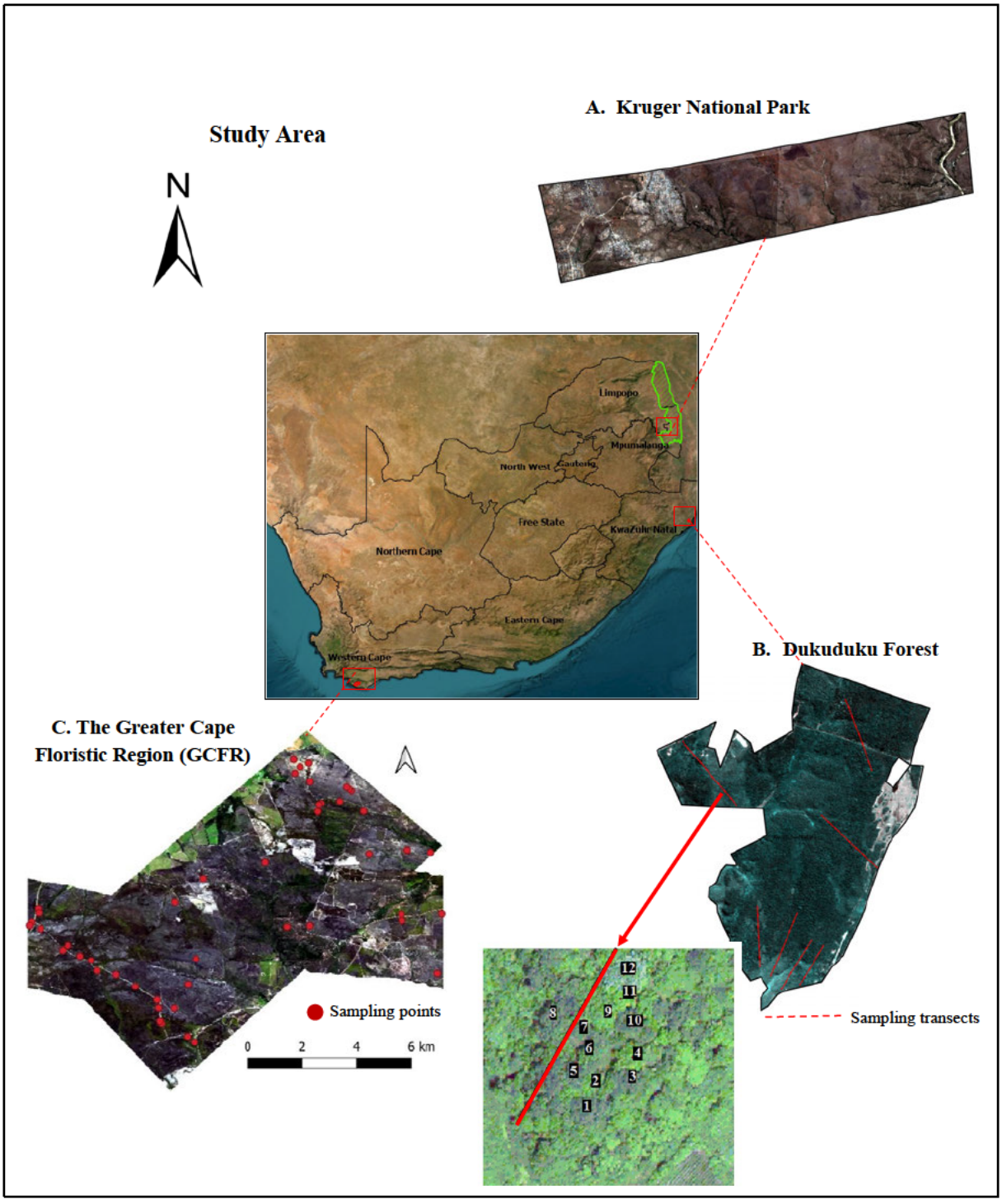


Figure 1.1. Map of South Africa showing the relative locations of the three study sites (Kruger, Dukuduku, and Greater Cape Floristic Region (GCFR)).

1.5. Thesis structure

The thesis is structured into seven chapters, with chapters one and seven as the introductory and synthesis sections, respectively. Chapters two to six comprise individual research articles; two of these have already been published in peer-reviewed journals, while the remaining three are currently under review. Each chapter is formatted as a standalone, peer-reviewed manuscript, beginning with an introduction and concluding with a summary that links to the next chapter. This format inherently results in some overlaps and repetition of theoretical concepts across chapters. Consequently, each chapter should be regarded as a self-contained unit of scholarly work, however, this modular approach should not detract from the overall coherence and integrative narrative of the thesis.

Chapter two: This chapter provides an overview of the use of RTMs to retrieve biophysical and biochemical properties in terrestrial biomes. The chapter presents trends, challenges and prospects of RTMs in quantifying biophysical and biochemical properties. Its further highlights research gaps and need to establish cost-effective methods for RTMs.

Chapter three: This chapter investigates the application of the PROSPECT-5 model to estimate carotenoid content by analysing different wavelengths within the field spectral data dataset. The study evaluates the spectral performance across three distinct datasets: savanna, tropical forest, and a combined dataset incorporating both biomes. The results indicate that the green spectral region consistently demonstrated the highest model performance across all datasets. In addition, bands selected through a Random Forest model, specifically the green, yellow, and red-edge regions, yielded the highest accuracy in the savanna dataset, with slightly lower performances observed in the combined and tropical forest datasets. Furthermore, the analysis revealed that the visible region (400–700 nm) exhibited particularly strong performance in estimating carotenoid content within tropical forests, followed by the combined datasets and savanna environments. These findings underscore the importance of targeted wavelength selection for radiative transfer model inversion, emphasizing that limiting the analysis to specific spectral regions rather than utilizing the entire spectrum can enhance estimation accuracy and efficiency.

Chapter four: Building upon chapter two, this chapter focuses on the assessment of phenolic compounds at the leaf level within a biome exceptionally rich in secondary metabolites, specifically, the Fynbos biome of the GCFR. Using field spectral data spectroscopy, we estimated the total phenols and flavonoids in relation to fire history over 14 years (2010-2024).

The results demonstrate that both fire disturbance and phenolic compounds are critical for the survival and resilience of the Fynbos biome. Areas that experienced recent fires in 2015, 2016, and 2021 exhibited elevated levels of phenolic compounds, whereas regions which last experienced fire in 2010 or areas that did not experience fire from 2010 to 2024 showed significantly lower concentrations. Additionally, employing the continuum removal technique, we identified that the strongest spectral signals for total phenols occurred at approximately 1660 nm, while flavonoids exhibited prominent signals at around 415 nm. Building upon these spectral insights, we further targeted specific wavelengths to develop indices for flavonoid estimation. Notably, the indices FI415,710 and FI420,710 demonstrated the highest correlations with flavonoid content, emphasizing the importance of subtle spectral features around 415 and 420 nm for accurate flavonoid retrieval. These findings underscore the potential of targeted spectral indices in identifying phenolic compounds associated with fire history and stress adaptation in phenolic-rich ecosystems.

Chapter five: Building upon the above chapters that focus on the leaf level, this chapter extends the investigation to the canopy scale, examining whether incorporating secondary metabolites into the PROSAIL model enhances the estimation of PAI in phenolic-rich landscapes. One of the earlier versions of PROSAIL include PROSAIL-4, which has been widely used and produced desired output. However, in the environment rich in secondary metabolites the former might not invert well due to lack of inclusion of secondary metabolites such as carotenoid and anthocyanins. To address the shortcomings, PROSAIL-D was introduced; which includes carotenoids and anthocyanins. Although secondary metabolites often have smaller absorption strengths than dominant constituents such as water and chlorophyll, the idea that adding them can improve the retrieval of structural parameters, such as plant area index (PAI), is therefore a hypothesis that requires testing. Thus, we compared the performance of PROSAIL-4, which does not account for secondary metabolites, with PROSAIL-D, a modified version that includes these compounds, particularly carotenoids and anthocyanins. Our results demonstrate that PROSAIL-D, which incorporates secondary metabolites, successfully retrieves PAI in phenolic-rich environments of the GCFR. This highlights the significant role these secondary metabolites play in influencing canopy reflectance and underscores the importance of including carotenoids and anthocyanins, key secondary compounds in RTM like PROSAIL-D, to improve biophysical parameter estimation in complex, metabolite-rich landscapes.

Chapter six: Building upon the previous chapter, this chapter focuses on the application of the PROSAIL-D model, proven effective in phenolic-rich environments, to estimate flavonoid content using inversion of AVIRIS hyperspectral data. Recognizing that chlorophyll can influence the spectral sensitivity to flavonoids, like the approach in Chapter Three, we further targeted specific wavelengths to develop spectral indices optimized for flavonoid estimation. The results reveal that the spectral index $FI_{417,693}$ achieved the highest accuracy in predicting flavonoid content, followed closely by $FI_{417,713}$. This underscores the critical importance of narrow spectral bands for reliable flavonoid retrieval, highlighting the potential of targeted spectral indices in enhancing remote sensing estimates of secondary metabolites within complex vegetation canopies.

Chapter Two:

The utility of Radiative Transfer Models (RTM) on remotely sensed data in retrieving biophysical and biochemical properties of terrestrial biomes: A systematic review

This chapter is based on;

Sibiya, B.S., Odindi, J., Mutanga, O., Cho, M.A. and Masemola, C., 2025. The utility of radiative transfer models (RTM) on remotely sensed data in retrieving biophysical and biochemical properties of terrestrial biomes: A systematic review. *Advances in Space Research*, 75 (10), 7424-7444.

Abstract:

Over the past few decades, there has been significant recognition of the value of RTMs for ecological remote sensing applications. This has led to various studies aimed at utilizing RTM techniques to quantify and map a range of biophysical and biochemical properties at different scales. Most literature reviews have predominantly focused on 1D models, such as PROSAIL, overlooking the more robust 3D models. This paper provides a detailed systematic review on the progress, gaps, and opportunities associated with both 1D and 3D RTM models in the context of remote sensing of terrestrial biomes. The review reveals a skewed distribution of research efforts between the Global North and South, with a significant concentration of studies conducted in the United States, China, and Germany, and fewer investigations in Africa. Furthermore, most studies have primarily utilized MODIS and Landsat sensors, focusing on plant attributes such as LAI and chlorophyll content, with studies predominantly conducted in grassland and forest landscapes. Overall, the findings indicate that PROSPECT and PROSAIL have been the most popular models over the past two decades. In the realm of 3D models, the Discrete Anisotropic Radiative Transfer (DART) and Forest Light Interaction Model (FLIGHT) models have been the most popular. These models have been primarily utilized through the look-up table (LUT) method, followed by the hybrid approach combining machine learning and RTMs. Understanding both 1D and 3D models offers an opportunity to assess the current state of research and identify future opportunities in the application of radiative transfer modelling for ecological remote sensing. By addressing the existing gaps and leveraging advancements in modelling techniques, researchers can enhance the accuracy and applicability of remote sensing on various ecosystems.

Keywords: Physical models, Radiative Transfer Model (RTM), Remote Sensing, biophysical and biochemical variables; Terrestrial biomes.

2.1. Introduction

The retrieval of biophysical and biochemical attributes is essential for effectively assessing and monitoring ecological applications. These attributes can be used as indicators to monitor changes in vegetation cover, such as a decline in the canopy, hence serving as an early warning sign of a deteriorating ecosystem (Cho et al., 2014a). Therefore, due to their significance for ecosystem function and relevance as indicators of ecosystem health, it is necessary to constantly map and monitor plant attributes in terrestrial biomes (Skidmore et al., 2021). The estimation of biophysical and biochemical attributes using remotely sensed data is achieved through three primary approaches: empirical, physical, and hybrid methods (Verrelst et al., 2015).

Most of the existing literature proposes empirical methods to find the relationship between plant attributes and remotely sensed data (Berger et al., 2018; Spafford et al., 2021). This is attributed to the empirical approach's inherent simplicity, which allows for straightforward implementation and often leads to satisfactory results (Verrelst et al., 2019). Whereas they produce desirable accuracy (e.g., RFR using spectral bands to predict LAI), empirical models require calibration for each specific dataset, as it is influenced by the characteristics of a respective database. For instance, the characteristics of the given database significantly influence the outcomes, which may lead to weaker results, particularly in large areas or with complex species components. This constraint might affect the reliability and adaptability of the method in different settings (Li et al., 2018; He et al., 2023; Darvishzadeh et al., 2008).

In contrast, physical-based models, such as RTMs, rely on fundamental physical principles to establish cause-and-effect relationships (Jacquemoud et al., 2009; Verrelst et al., 2019). These models enable the scientific community to comprehend how plants interact with light, including processes like absorption, scattering, and the emission of light (i.e., sun-induced chlorophyll fluorescence) (Verrelst et al., 2019). Over the last three decades, various RTMs models with varying degrees of complexity ranging from simple turbid media model (1D) to complex Monte Carlo ray tracing model (3D) (refer to the RAMI exercise for comparative analysis (Widlowski et al., 2015)) have been developed.

Briefly, 1D models treat plant canopies as uniform layers in which leaves are randomly placed small objects that can absorb and scatter light. These models assume that the canopy spreads out evenly in one direction and infinitely horizontal with limited height (Dorigo et al., 2007). They focus on the total leaf area, making them useful for studying light distribution in thick

vegetation like agriculture and grasslands, where individual leaves are much smaller than the plants' height. They are often called "*economically invertible*" because they are computationally cheap, simpler, and need fewer parameters, allowing for quick inversion (Verrelst et al., 2019). A well-known example is the PROSAIL model, which combines the leaf model PROPECT (Jacquemoud and Baret, 1990) with the canopy model SAIL (Verhoef, 1984). However, 1D models face challenges when applied to heterogeneous canopies, especially in fine-scale mapping, where the reflectance of each pixel is heavily influenced by its surroundings, such as crown shadows (Qi et al., 2023).

Conversely, three-dimensional (3D) models consider factors such as shadowing and clumping, as well as both vertical and horizontal variations. This complexity renders them particularly well-suited for representing forests or sparse canopies. Verrelst et al. (2019) describe these models as "*non-economically invertible*" due to their intricate nature and substantial computational resource requirements. They often necessitate numerous input variables and advanced mathematical modelling techniques. While such models can generate detailed representations, they incur a considerable computational cost. Examples of the most widely used 3D models include DART (Gastellu-Etchegorry et al., 1996) and FLIGHT (North, 1996).

The third category is the hybrid approach, which trains empirical regressing algorithms (e.g., Machine learning) on a database of spectral simulations generated by RTMs (Verrelst et al., 2015; Doringo et al., 2007; Tsele and Ramoelo, 2024). One example of hybrid models is those that combine RTM (e.g., PROSAIL, DART, and INFORM), with machine learning algorithms like ANNs, GPR, and Random RFR (Berger et al., 2018; He et al., 2023; Ali et al., 2021). The key advantage of the hybrid approach is its ability to leverage the simplicity and flexibility of the empirical approach while maintaining the physical foundations of radiative transfer (Verrelst et al., 2019). This leads to more accurate and robust estimates of vegetation properties. The growing use of hybrid models and RTMs has resulted in an expanding body of literature on these topics.

Over the past decades, numerous publications have reviewed RTMs and their applications in remote sensing. These reviews provide valuable syntheses of the existing knowledge on RTMs, with emphasis on state-of-the-art algorithmic approaches (Liang, 2005; Liang, 2007; Dorigo et al., 2007; Verrelst et al., 2015; Verrelst et al., 2019). Additionally, they have focused on the applications of the PROSAIL model within the agricultural sector (Ishaq et al., 2023). Notably, Jacquemoud et al. (2009) conducted the first systematic review on PROSAIL model between

1992 and 2007, evaluating the use of RTMs for estimating vegetation biophysical and biochemical properties, with a focus on the development and validation of RTMs. Building on Jacquemoud et al. (2009), Berger et al. (2018) focused on the retrieval of vegetation parameters using the PROSAIL model in conjunction with hyperspectral remote sensing data.

Previous reviews have mainly concentrated on the use of 1D models, such as PROSAIL, and largely overlooked the potential of 3D models. This study aims to address this gap by examining both 1D and 3D models across different types of terrestrial ecosystems. The importance of this investigation is further highlighted by the growing availability of high-temporal-resolution data from missions like Copernicus Sentinel-2 and Sentinel-3. Additionally, new as well as the planned hyperspectral missions, including EnMAP, PRISMA, FLEX, and HyspIRI, will provide more detailed environmental data (Verrelst et al., 2016). Moreover, advancements in airborne platforms, particularly unmanned aerial vehicles (UAVs), have expanded the range of remotely sensed data available for analysing vegetation (Kükenbrink et al., 2021; Yin et al., 2022). This wide variety of data necessitate the development of improved processing techniques that are accurate, reliable, and efficient. Given these advancements, it is crucial to re-evaluate RTMs to ensure they can effectively handle the growing body of new generation remotely sensed data.

Hence, this study provides a systematic review of the utility of RTMs in the analysis of remotely sensed data for the retrieval of biophysical and biochemical properties of terrestrial biomes. The objectives of this review are to; (i) Evaluate the current state-of-the-art applications of RTMs, (ii) Identify their strengths and limitations, and (iii) Explore opportunities for further advancements in this field. The insights gained from this review will be beneficial to researchers, remote sensing practitioners, and stakeholders involved in the development and implementation of operational satellite data processing workflows.

2.2. Methodology

2.2.1. Phase 1: Literature search

The first step in conducting a systematic review is identifying key terms to create a search string. For our study, we used the Artmo Toolbox (artmotoolbox.com) due to its popularity and includes a variety of RTMs, ranging from 1D to 3D models. We also incorporated terms from influential reviews by Verrelst et al. (2015; 2019), Berger et al. (2018), Jacquemoud et al. (2009), and Widlowski et al. (2015). These keywords were then used to formulate the PRISMA

(Table 2.1) (Page et al., 2021). It is worth mentioning that the search strategy was designed to find applied studies in specific biomes, rather than purely theoretical model development papers. Before accessing literature from various databases, we conducted iterative testing of the identified key terms and their variations. This process was repeated until we observed no significant changes in the search results, ensuring consistency and minimizing potential biases that might affect the findings of our study.

The literature review involved sourcing articles from Scopus, Web of Science (WOS), and Google Scholar, covering the period from 2000 to 2023. Initially, a total of 2,579 articles were identified: 1,405 from Web of Science, 907 from Scopus, and 267 from Google Scholar. These articles were then exported to EndNote for screening. To be eligible for inclusion, articles had to meet the following criteria:

i. Focus on RTMs applicable to terrestrial biomes.
ii. Examine various RTMs used to retrieve biophysical or biochemical properties.
iii. Employ Geographic Information Systems (GIS) or remote sensing methodologies.
iv. Published in peer-reviewed journals.
v. Article had to be written in English.

During the initial screening, 2,579 articles were evaluated, and 950 identified as duplicates and subsequently removed. This left 1,629 articles, which were assessed based on their titles and abstracts to determine their relevance to remote sensing-based RTMs.
During this assessment, we excluded ineligible articles, including 74 that were not written in English. From the remaining 964 articles, 41 were excluded because they were not available in full text, leaving 923 articles for further eligibility assessment. Out of these (923), in total, 754 articles were excluded from consideration due to their failure to meet the established selection criteria. Ultimately, 169 articles that met our selection criteria were used for data extraction as illustrated in Figure 2.1.

Table 2.1: Key searched terms.

Forest AND/OR Radiative Transfer Model	Tundra AND/OR Radiative Transfer Model	Boreal Forest AND/OR Radiative Transfer Model	Grassland AND/OR Radiative Transfer Model	Savanna AND/OR Radiative Transfer Model
Forest AND/OR PROSPECT	Tundra AND/OR PROSPECT	Boreal Forest AND/OR PROSPECT	Grassland AND/OR PROSPECT	Savanna AND/OR PROSPECT
LIBERTY AND/OR Forest	LIBERTY AND/OR Tundra	LIBERTY AND/OR Boreal Forest	LIBERTY AND/OR Grassland	LIBERTY AND/OR Savanna
Fluspect AND/OR Forest	Fluspect AND/OR Tundra	Fluspect AND/OR Boreal Forest	Fluspect AND/OR Grassland	Fluspect AND/OR Savanna
SAIL AND/OR Forest	SAIL AND/OR Tundra	SAIL AND/OR Boreal Forest	SAIL AND/OR Grassland	SAIL AND/OR Savanna
FLIGHT AND/OR Forest	FLIGHT AND/OR Tundra	FLIGHT AND/OR Boreal Forest	FLIGHT AND/OR Grassland	FLIGHT AND/OR Savanna
INFORM AND/OR Forest	INFORM AND/OR Tundra	INFORM AND/OR Boreal Forest	INFORM AND/OR Grassland	INFORM AND/OR Savanna
SCOPE AND/OR Forest	SCOPE AND/OR Tundra	SCOPE AND/OR Boreal Forest	SCOPE AND/OR Grassland	SCOPE AND/OR Savanna
PROSAIL AND/OR Forest	PROSAIL AND/OR Tundra	PROSAIL AND/OR Boreal Forest	PROSAIL AND/OR Grassland	PROSAIL AND/OR Savanna
PARAS AND/OR Forest	PARAS AND/OR Tundra	PARAS AND/OR Boreal Forest	PARAS AND/OR Grassland	PARAS AND/OR Savanna
GeoSAIL AND/OR Forest	GeoSAIL AND/OR Tundra	GeoSAIL AND/OR Boreal Forest	GeoSAIL AND/OR Grassland	GeoSAIL AND/OR Savanna
FLiES AND/OR Forest	FLiES AND/OR Tundra	FLiES AND/OR Boreal Forest	FLiES AND/OR Grassland	FLiES AND/OR Savanna
RAPID AND/OR Forest	RAPID AND/OR Tundra	RAPID AND/OR Boreal Forest	RAPID AND/OR Grassland	RAPID AND/OR Savanna
DART AND/OR Forest	DART AND/OR Tundra	DART AND/OR Boreal Forest	DART AND/OR Grassland	DART AND/OR Savanna

2.2.2. Phase 2: Data extraction

The eligible data (n = 169) were extracted from EndNote and transferred to an Excel spreadsheet. This spreadsheet included essential information such as the author, title, year of publication, journal, keywords, and abstract. Additionally, it contained supplementary details like the study site (country), sensors used, landscape (vegetation type), the RTM employed, and plant attributes. The extracted data were then converted into numerical variables to facilitate further analysis.

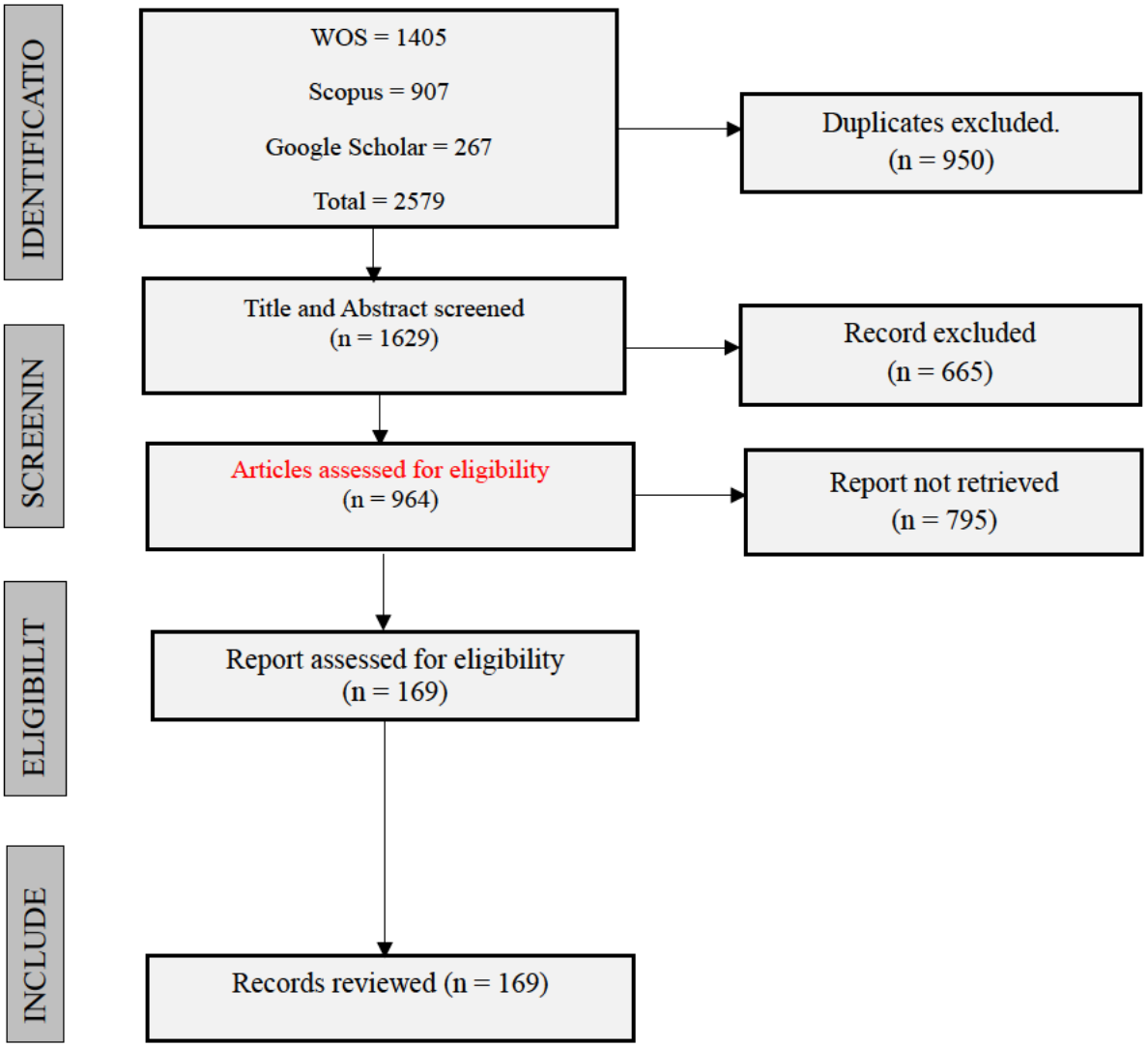


Figure 2.1. Studies selected for review.

2.2.3. Phase 3: Analysis

The selected articles were subjected to quantitative and qualitative analyses. To assess the progress of remote sensing technology related to RTMs, basic statistical frequencies and trend analyses were performed using Excel. A bibliometric analysis was conducted to examine topical concepts and trends in characterizing RTMs with remotely sensed data. This analysis aimed to evaluate literature characteristics and identify connections among key terms within the research field (Sibanda and Bacela, 2024).

Trends were identified using VOSviewer that generates a network of key terms visualized as linked clusters (refer to Van Eck and Waltman (2010) for VOSviewer functionality). The VOSviewer map shown in Figure 2.2 was created using the full counting method based on the co-occurrence of terms in titles, abstracts, and keywords. All tables, graphs, and illustrations were generated following the above-mentioned procedure.

2.3. Results

2.3.1. Analysis of keywords frequency, growth, and co-occurrence

In analysing the key concepts taken from the titles, abstracts and keywords of the retrieved studies, the network map was produced using Vosviewer software with four clusters (Figure 2.2). The green cluster produced terms such as “Radiative Transfer Model,” “PROSPECT,” “SAIL,” “PROSAIL,” “LUT,” “DART,” “canopy level,” “chlorophyll”. The cluster indicates the utility of RTM to understand how light interacts with plant canopies and interpret vegetation reflectance in terms of biophysical and biochemical properties (e.g., chlorophyll). The cluster shows that over the past decades, PROSPECT and SAIL model, respectively, known as PROSAIL, are the most popular models (Jacquemoud et al., 2009). Furthermore, the cluster shows that among the 3D RTM model, the DART model is the most utilized and the models have been tested using mostly LUT approach (Gastellu-Etchegorry, 2008).

The second cluster (blue) had “Grassland” “ndvi,” “fmc,” “performance,” and “MODIS”. This cluster indicates that most studies were conducted in grasslands to retrieve different plant traits (e.g., fuel moisture content). Also, the “NDVI” is one of the widely used vegetation indices since the beginning of remote sensing. In addition, inclusion of terms such as “MODIS” showcase its wide usage for studying terrestrial biomes globally (Timothy et al., 2016; Mngadi et al., 2022; Thenkabail et al., 2002).

The third cluster (yellow) had key concepts like “LAI”, “leaf chlorophyll content (LCC)”, “canopy chlorophyll content (CCC)”, “rmse”, “sentinel”, and “landsat”. This cluster shows that LAI, LCC, and CCC were the most retrieved plant traits (Berger et al., 2018; Ishaq et al., 2023). The incorporation of words such as Landsat and Sentinel images shows the link between freely available remotely sensed data in retrieving plant attributes such as LAI and chlorophyll content. Additionally, the cluster included words like "root mean square error (rmse)" that pertain to the accuracy evaluation of model performance in retrieving plant attributes.

Lastly, the red cluster includes “hyperspectral”, “lidar”, “gpp”, “sif”, “canopy structure”, “plant function trait”, “scope”, showing that the hyperspectral data was the most used in retrieving plant functional trait (i.e., “gpp”, “sif”) using RTMs (e.g., scope).

In summary, the reported clusters highlighted key RTMs, plant attributes, vegetation type, and sensors, as well as hybrid approaches used for inversion.

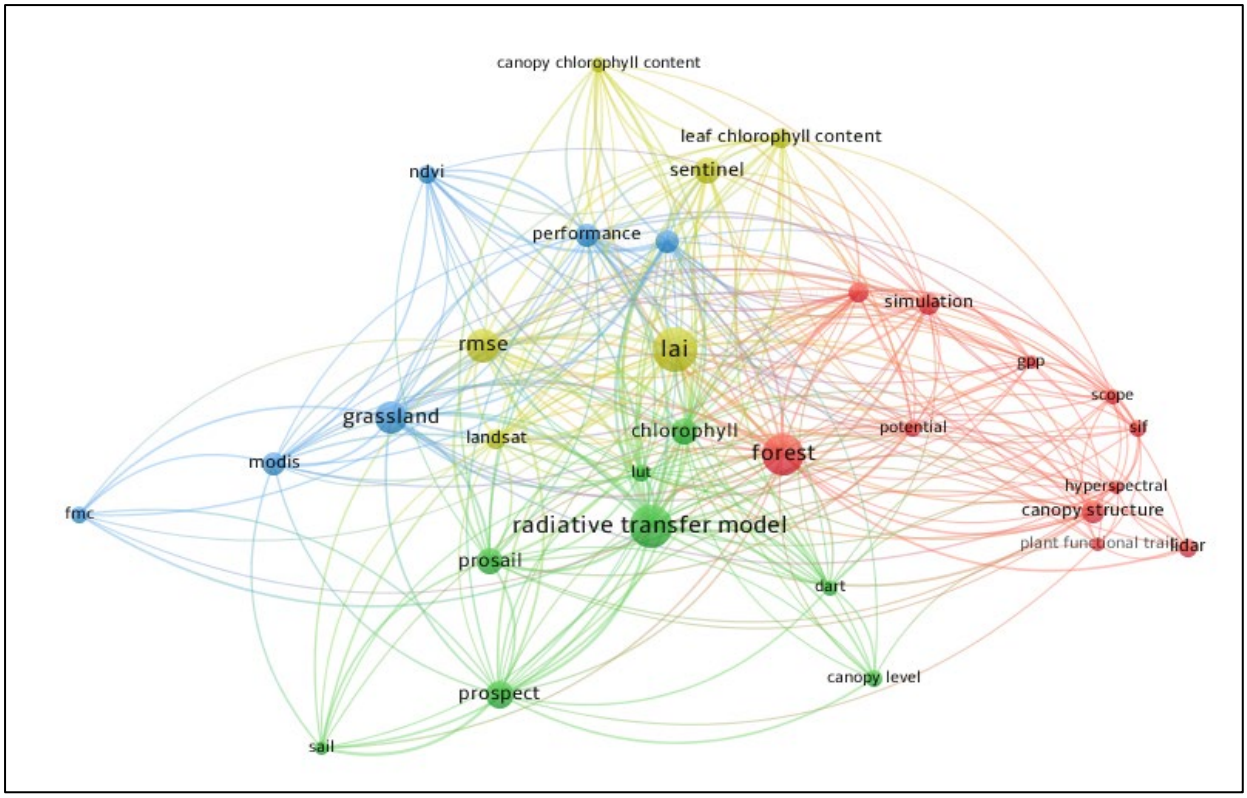


Figure 2.2. Topical concepts of RTM and biophysical and biochemical property retrieval from terrestrial biomes derived using data from titles, abstracts, and keywords.

2.3.2. Global trend and distribution of RTM publications

There has been a steady increase in the number of articles published from 2000 to 2023 that focus on retrieving the plant traits of terrestrial biomes (Figure 2.3). From 2000 to 2014, the number of publications using RTM models ranged from 1 to 8 per year. A noticeable increase is seen in 2019 with 14 publications, which is also a peak.

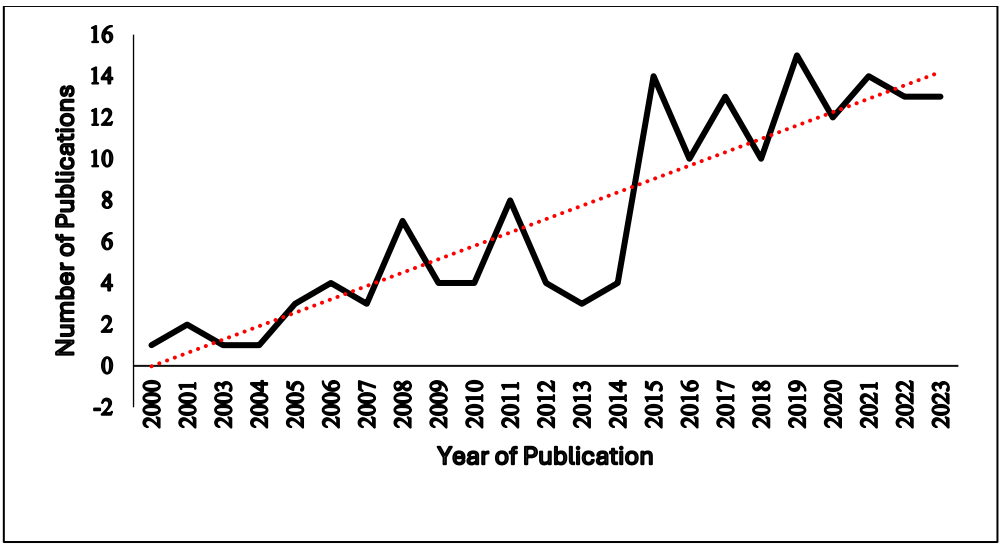


Figure 2.3. Frequency of studies from year 2000 to 2023.

Most of the published articles were published in the northern hemisphere, i.e., United States of America (USA) (27), China (22), and Germany (18). In contrast, Africa recorded the lowest number of articles, with only six (Figure 2.4).

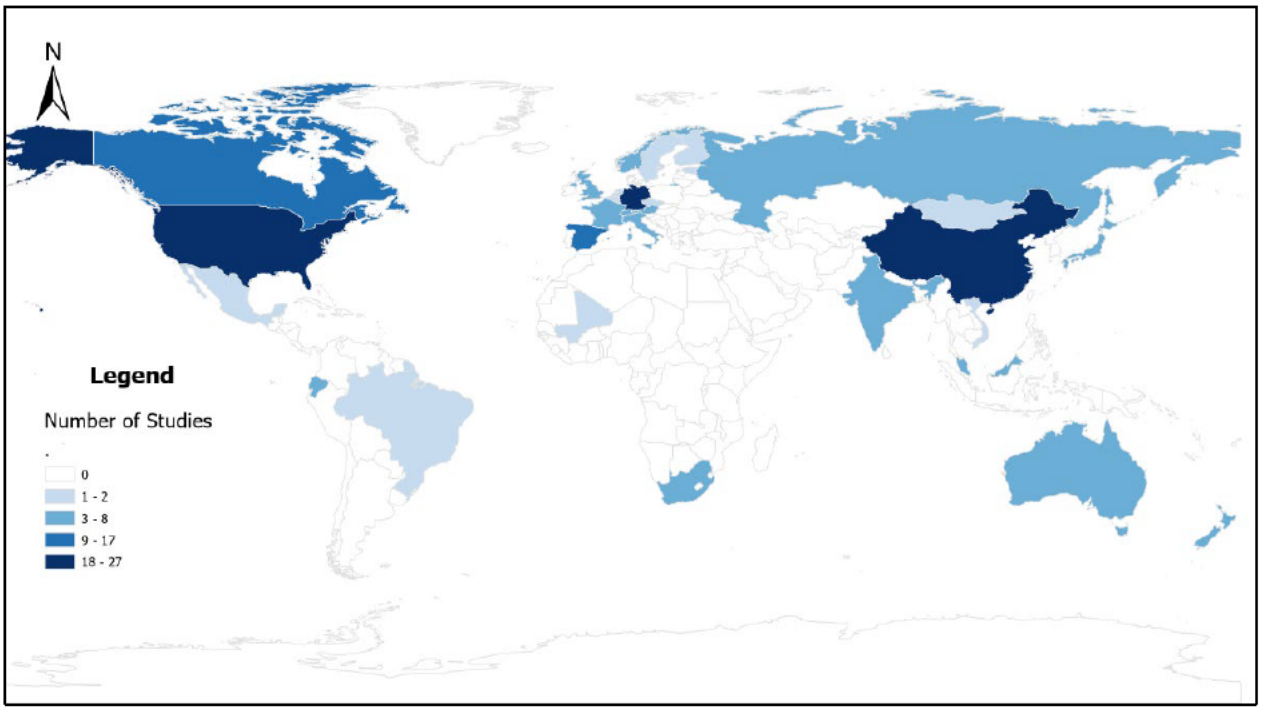


Figure 2.4. Global distribution of the selected published articles on the use of RTMs to retrieve plant attributes in terrestrial biomes.

2.3.3. Sensor datasets used

A noticeable increase was observed in the use of remote sensors of RTMs to retrieve plant traits in terrestrial biomes. The literature reviewed identified 196 sensor types (Figure 2.5). As indicated in Figure 2 (blue cluster), the most widely used sensors were MODIS (19%) and Landsat (13%) (yellow cluster). Furthermore, several studies (11%) have employed portable hyperspectral devices (e.g., field spectral data) for *in-situ* collection of spectra to retrieve terrestrial biomes biophysical and biochemical characteristics. Also, there is a growing interest in the use of recently launched sensors such Environmental Mapping and Analysis Program (EnMAP) (Guanter et al., 2015) and PRecursore IperSpettrale della Missione Applicativa (PRISMA) (Labate et al., 2009), as well as the utilization of UAVs for vegetation analysis and monitoring. Lastly, there is a noticeable rise in the use of Sentinel-2 data (yellow cluster) in the use of Sentinel-2 data for retrieving plant attributes using RTMs in terrestrial biomes.

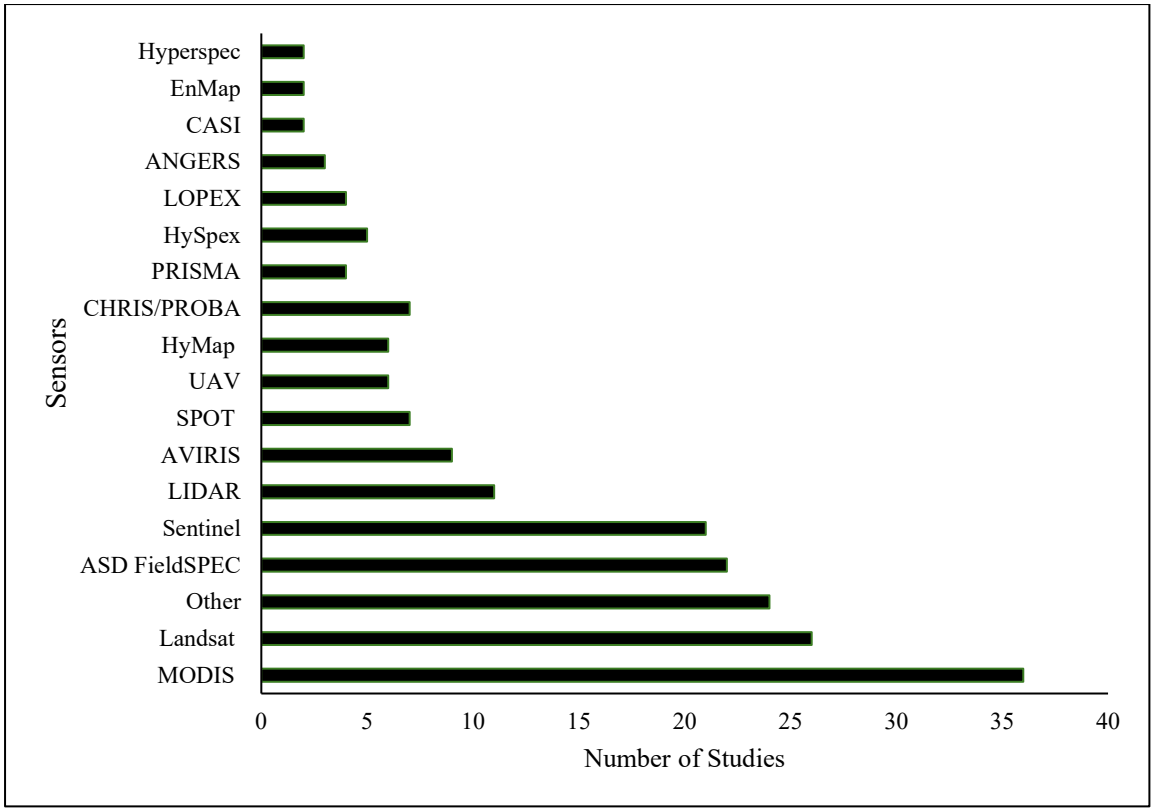


Figure 2.5. The frequency of use for sensor type (sensors used only once are classified as other).

2.3.4. Vegetation types and biophysical and biochemical variables

Figure 2.6 shows the terrestrial biomes examined. The Chaparral biome, Mangrove (grouped into Forest), and shrublands were not included in the initial search, but were picked from the retrieved literature. The forest biome from the red cluster (e.g., tropical, deciduous, and temperate forest) characterized by heterogeneous canopy architecture had the highest percentage of 44%. This was closely followed by the grassland biome (blue cluster) characterized by homogeneous canopy architecture with 37%. The two biomes accounted for 81% of all the landscapes examined in the study. Generally, literature analysis showed that RTM was successfully used for various vegetation types from homogenous to heterogeneous canopy architecture.

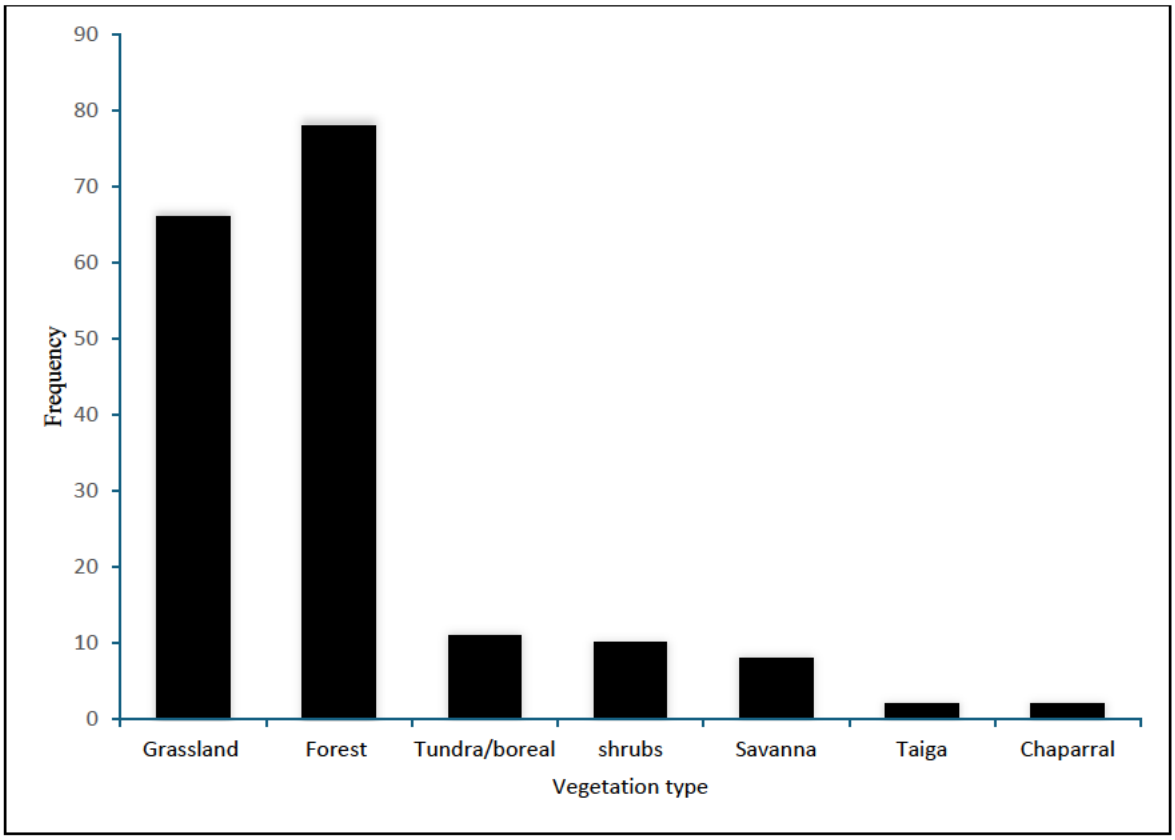


Figure 2.6. Vegetation types analyzed by the evaluated studies. The category 'natural' includes forest, grassland, Savanna, Tundra/boreal, shrubland, Taiga, and Chaparral biomes.

As indicated by the yellow cluster (Figure 2.2), the LAI was the most retrieved attribute and closely followed by chlorophyll content (LCC and CCC) (Figure 2.7). Lastly, the FPAR, fAPAR, SMC, and Vcmax, respectively, were estimated in only a few studies.

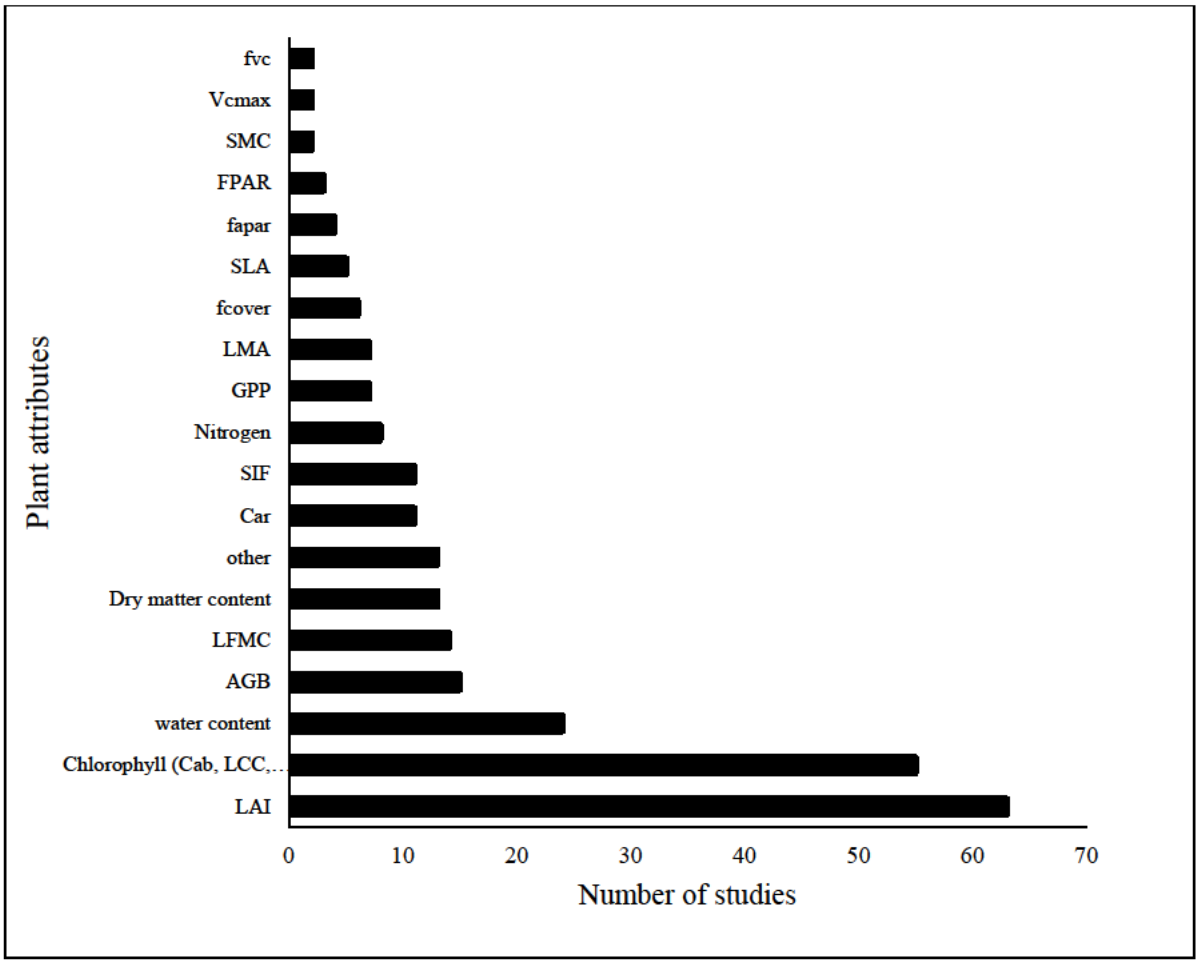


Figure 2.7. Biophysical and biochemical variables predicted using the RTMs ((leaf and canopy chlorophyll content (Cab, LCC and CCC), carotenoid content (Car), dry matter content (Cm or LMA), Equivalent water thickness or canopy/leaf water content (EWT, CWC, CAW, and Cw), specific leaf area (SLA) and fuel moisture content (FMC), above ground biomass (AGB), fraction of vegetation cover (fcover), fraction of absorbed photosynthetically active radiation (fAPAR), Nitrogen (N), Gross Primary Productivity (GPP), Solar-induced chlorophyll fluorescence (SIF), soil moisture content (SMC) and Vcmax (maximum velocity of carboxylation)).

2.3.5. Applications of RTM

The application of RTM in terrestrial biomes using remote sensing applications is commonly dependent on the landscape type, ((i.e., homogenous (1D) or heterogeneous (3D)). Results showed the popularity of 1D models for retrieving plant traits compared to the 3D models (Figure 2.8). As shown in Figure 2.2 (green cluster), the PROSAIL (46%) and PROSPECT (14%) were the most popular in retrieving terrestrial biomes' plant traits. Furthermore, among the several 3D models, the DART (11%) was the most popular model. Meanwhile, a noticeable change was identified in the number of studies that utilized less intensive 3D RTMs such as INFORM (6%). Such models show a comparable performance to complex models and are thus considered physically sound when retrieving plant terrestrial biomes traits.

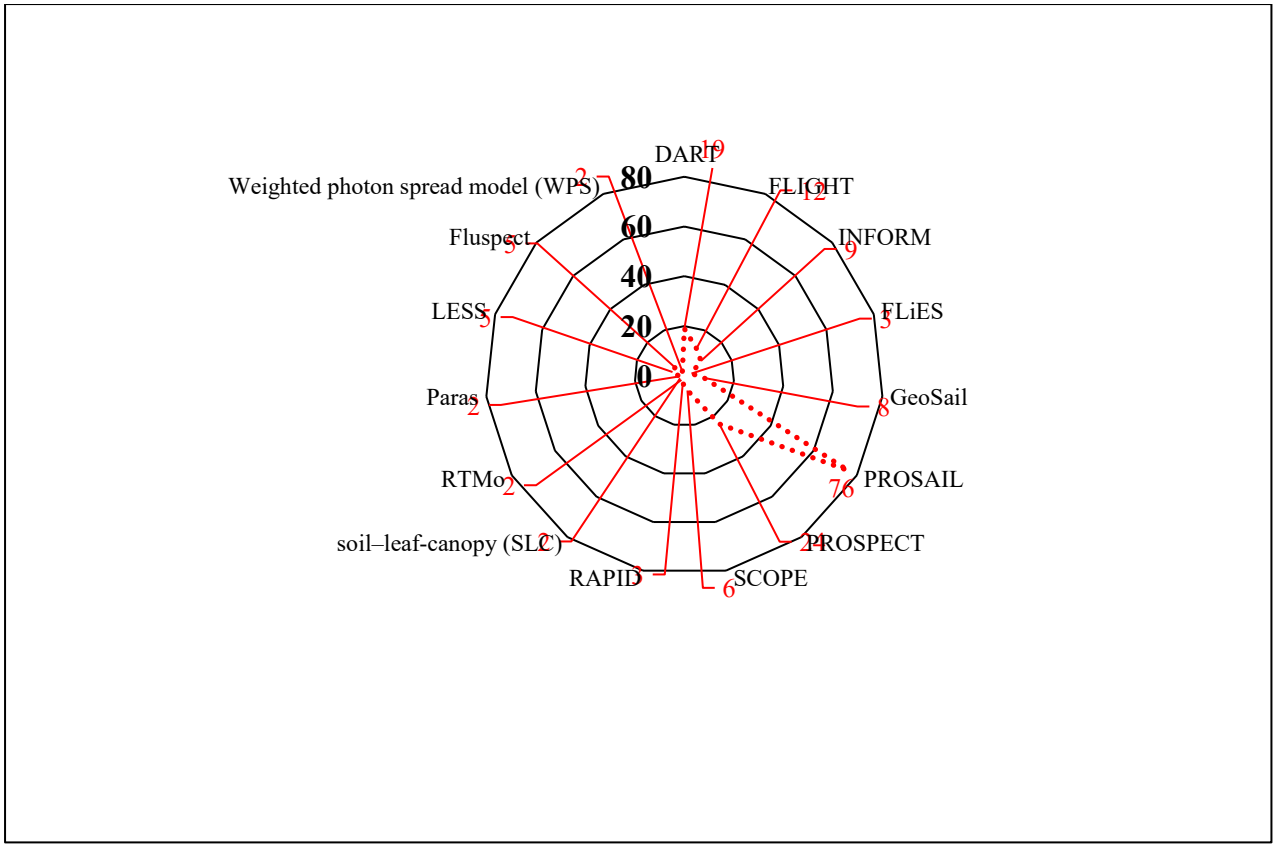


Figure 2.8. RTMs used to plant traits in terrestrial biomes.

Figure 2.9 depicts that the PROSAIL model was predominantly utilized in grassland vegetation, with a significant application also observed in forest landscapes. Conversely, the PROSPECT model, known for its proficiency in vegetation studies, was predominantly employed in forest biomes. Additionally, the DART, FLIGHT, and INFORM model, respectively, showed a focus on forest landscapes in their applications. These findings indicate that different models have been strategically chosen based on their strengths and capabilities for specific vegetation types and landscapes, highlighting the importance of considering appropriate model selection in diverse ecosystems.

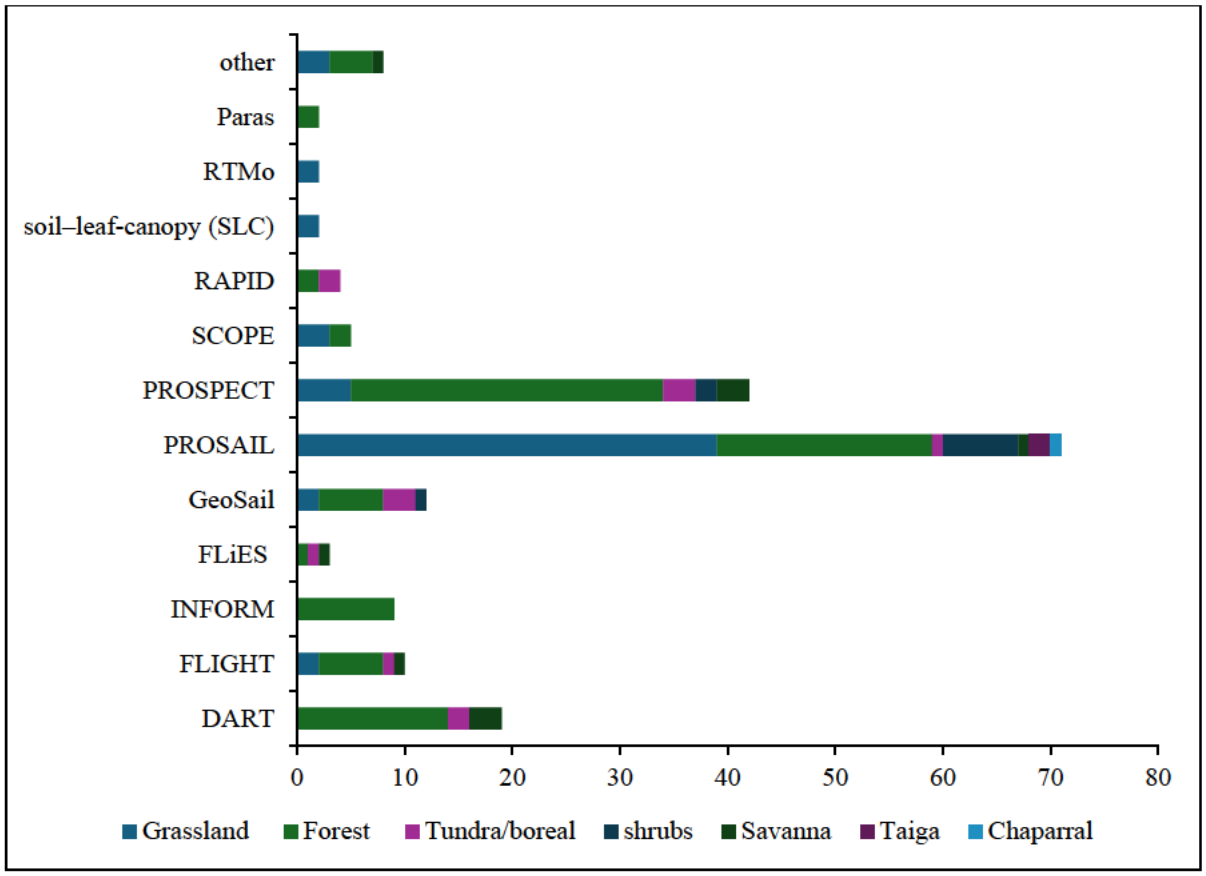


Figure 2.9. Applications of RTM models on different vegetation types.

2.3.6. The hybrid approaches

The following techniques were used to perform the inversion of RTM for retrieving plant traits variables: (i) parametric: applies vegetation indices and relate them to the variable of interest, (ii) LUTs, and (iii) Hybrid approach: combining statistical approach and physical models. Figure 2.10 shows the use of the techniques. Overall, LUTs had the highest frequency, followed by machine learning algorithms and Parametric.

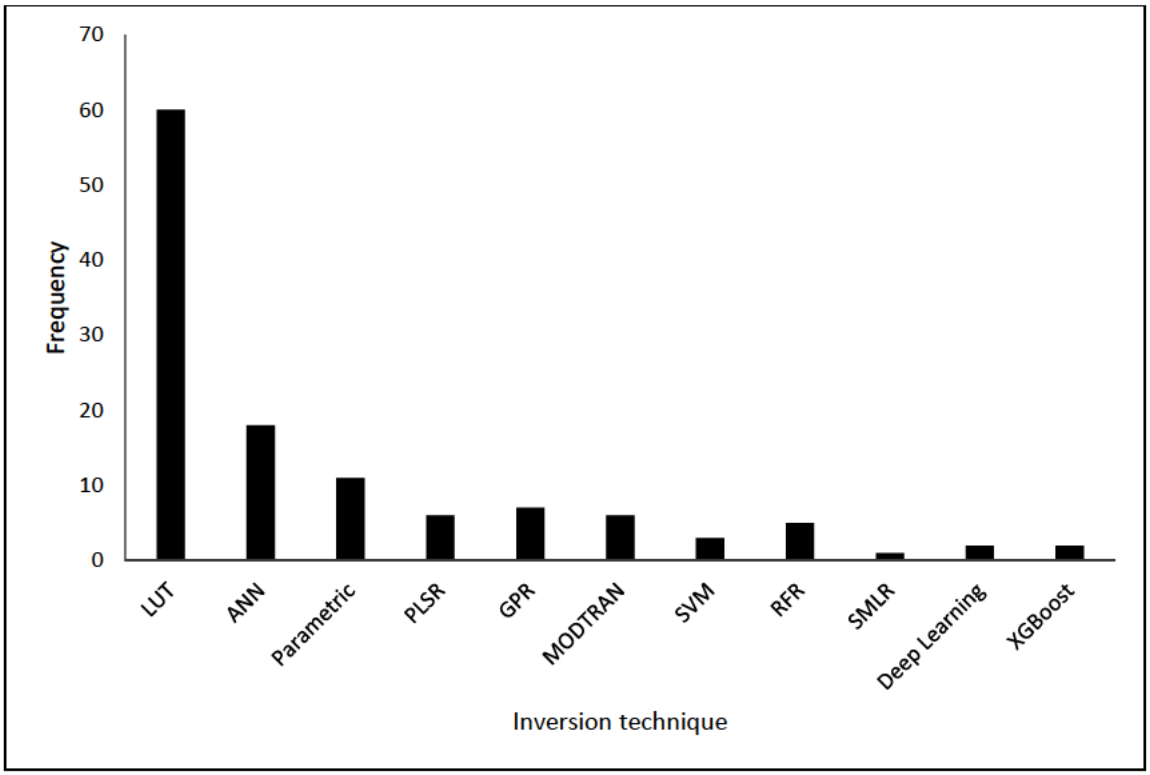


Figure 2.10. Variable retrieval methods of all evaluated studies involving the RTMs.

2.4. Discussion

2.4.1. Global distribution publication trends

The results presented in Figure 2.4 indicate that a significant proportion (16%) of the published articles in the field of remote sensing originate from the USA, for example, Adeline et al. (2016); Banskota et al. (2015); Bayat et al. (2018); Bayat et al. (2020) and others. This prevalence can be attributed to the country's substantial advancements in remote sensing technology. For instance, the USA is responsible for the ownership and operation of approximately 61% of the satellites currently in orbit (Zhao et al., 2022). These datasets are extensively utilized for monitoring terrestrial ecosystems.

This observation aligns with the findings of Zhuang et al. (2013), who conducted a comprehensive review of remote sensing literature utilizing data from the Science Citation Index (SCI) and Social Sciences Citation Index (SSCI) spanning from 1991 to 2010. Their study highlighted the USA's significant international collaborations in remote sensing research. Furthermore, Duan et al. (2020) evaluated the application of remote sensing in terrestrial Protected Areas and concluded that the USA had the highest number of publications from 1991 to 2018 within the Web of Science database. Overall, these findings underscore the extensive availability of remote sensing data in the USA, derived from various satellite missions such as

NASA's Landsat and MODIS, in addition to airborne and ground-based sensors. Remote sensing techniques and RTMs can effectively simulate and interpret these measurements, thereby enhancing researchers' understanding of the Earth's surface.

Figure 2.4 further illustrates that the United States is closely followed by China, which accounts for 13% of the published articles for example, He et al. (2017); He et al. (2019); Jiang et al. (2021); Jin et al. (2021); Luo et al. (2019) and others. This could be attributed to the fact that China has vast grassland ecosystems, with more than 85% of studies conducted in grassland in this study (Dai et al., 2016). China has also made significant advancements in ecological science and technology, including the development of its satellite systems like the Gaofen and Huanjing series (Yuan et al., 2022). China has been actively involved in international collaborations, such as the Group on Earth Observations (GEO), which promotes the use of Earth observation data for sustainable development (Wang, 2022). China has also designed and launched environmental and disaster monitoring satellites like the HJ-1, which provide data for ecological monitoring, including grassland ecosystems (Meng et al., 2017; Masenyama et al., 2022). These advancements collectively contribute to China's growing influence in the domain of remote sensing and ecological research.

In contrast, results show few publications in the developing regions such as Africa (Cho et al., 2014; Masemola et al., 2016; Pinheiro et al., 2006; Kennedy et al., 2020; Sibiya et al., 2025b; Tsele and Ramoelo, 2024). This can be attributed to financial constraints, lack of technical expertise, limited access to technology, and infrastructure. Also, limited awareness among policymakers, researchers, and other stakeholders on the potential benefits of using remote sensing for, among others, land-use planning, agriculture, forestry, natural resource management, and disaster response. Overall, a significant increase was observed in the studies that utilized RTM based on remote sensing of terrestrial biomes globally (Figure 2.4). Over the last two decades, there has been a significant development in remote sensing, providing ecologists and policy makers with multiplatform remotely sensed data to monitor terrestrial biomes.

2.4.2. Earth observation sensors

Figure 2.5 demonstrates that MODIS and Landsat sensors are the most frequently utilized for mapping and monitoring terrestrial biomes on a global scale (Zhang et al., 2023; Bayat et al., 2018; Bayat et al., 2020; Casas et al., 2014; Cheng et al., 2006; Cho et al., 2014; Dasgupta et al., 2009; Fang and Liang, 2005; He et al., 2017; He et al., 2019; Luo et al., 2019; Luo et al.,

2019). The prominence of these sensors can be attributed to their extended operational periods in orbit, making them the most readily available options for researchers. Additionally, the data generated by these sensors are freely accessible, which facilitates widespread adoption for the mapping and monitoring of biophysical and biochemical properties (Timothy et al., 2016; Mngadi et al., 2022). These findings align with the observations of Masenyama et al. (2022), who reported that MODIS and Landsat accounted for the highest number of publications in their review of remote sensing technologies used to quantify ecosystem services in grasslands.

Furthermore, the results of the present study indicate a significant number of investigations that have explored the utility of handheld hyperspectral devices (Ali et al., 2016; Arellano et al., 2016; Arellano et al., 2017a; Arellano et al., 2017b; Clevers and Kooistra, 2011; Darvishzadeh et al., 2008; Pacheco-Labrador et al., 2021; Vohland and Jarmer, 2008; Wang et al., 2015; Dehghan-Shoar et al., 2023a; Dehghan-Shoar et al., 2023b; Sibiya et al., 2024). These findings are consistent with those reported by Soubry et al. (2021), who noted that a substantial proportion of studies (16.5%) utilized *in situ* hyperspectral sensors. These sensors provide a comprehensive range of electromagnetic spectrum sections, which enhances the accuracy and effectiveness of estimating and mapping the biophysical and optical properties of vegetation (Cerasoli et al., 2018). Furthermore, *in situ* hyperspectral sensors are equipped with multiple electromagnetic spectrum bands that facilitate data acquisition near the canopy, thereby significantly minimizing potential errors associated with remote sensing measurements.

In addition to the specifications of the field spectral data field, Figure 2.5 indicates that the hyperspectral data utilized in these studies were predominantly acquired via airborne sensors, particularly the Compact Airborne Spectrographic Imager (CASI) (Melendo-Vega et al., 2018; Croft et al., 2013), HySpex (Wang et al., 2017; Wang et al., 2018; Ali et al., 2016), the Airborne Imaging Spectrometer HyMap (Atzberger et al., 2013; Atzberger et al., 2015; Darvishzadeh et al., 2010; Darvishzadeh et al., 2011; Dorigo et al., 2009), the Airborne Near-Ground Experimental Remote Sensing System (ANGERS)/Leaf Optical Properties Experiment (LOPEX) (Feret et al., 2008; Féret et al., 2011; Féret et al., 2019; He et al., 2023), the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (Adeline et al., 2016; Banskota et al., 2015; Casas et al., 2014; Cheng et al., 2006), and CHRIS/PROBA (Kennedy et al., 2020; Schlerf and Atzberger, 2012; Yanez-Rausell et al., 2015; Verrelst et al., 2012). Hyperspectral data are crucial for detailed environmental monitoring and land management, as they allow for precise identification of materials and conditions across diverse ecosystems. This information is essential for applications such as biodiversity assessment, land-use planning, and resource

management, ultimately contributing to sustainable environmental practices (Berger et al., 2018).

The recent launch of advanced sensors like EnMAP and PRISMA has significantly enhanced vegetation analysis and monitoring (Han et al., 2023; Schucknecht et al., 2023; Pascual-Venteo et al., 2024; Verrelst et al., 2021). EnMAP, featuring a hyperspectral sensor with 242 bands across the visible, near-infrared, and shortwave-infrared regions, enables detailed spectral analysis at a spatial resolution of 30 meters (Guanter et al., 2015). This is a notable improvement over sensors like AVIRIS, which provide fewer bands and lower resolution, making them less effective in detecting subtle spectral differences. PRISMA shares similar advantages with its wide spectral range, improving land cover and vegetation monitoring compared to sensors like HYMAP that have limited capabilities (Labate et al., 2009).

Both EnMAP and PRISMA incorporate advanced atmospheric correction methods within their processing chains, reducing the need for extensive post-processing compared to sensors like CHRIS/PROBA. They also promote modern data accessibility practices, offering easier public access and integration into various applications (Verrelst et al., 2019). Furthermore, the missions of EnMAP and PRISMA highlight interdisciplinary collaboration, facilitating partnerships among researchers, policymakers, and practitioners, which reflects a more comprehensive understanding of ecological complexities than found with earlier sensors.

The deployment of UAVs has significantly heightened global interest in vegetation analysis and monitoring, as illustrated in Figure 2.5. This trend is supported by a growing body of literature, including studies by Juszak et al. (2017), Kükenbrink et al. (2021), Singh et al. (2023), Hernández-Clemente et al. (2012), Yin et al. (2022), Zhu et al. (2023), Zhou et al. (2023). UAVs, commonly known as drones, are increasingly becoming popular for mapping and monitoring ecosystems.

These aerial platforms are equipped with advanced technologies, such as multispectral, hyperspectral imaging systems, and Light Detection and Ranging (LIDAR) systems. Such capabilities provide significant flexibility and cost-effectiveness compared to traditional spaceborne methods (Sibanda et al., 2021). Drones can operate at low altitudes, allowing for the acquisition of high-resolution data with exceptional speed and accuracy. This enables near-real-time information collection regarding land use, agricultural practices, and environmental hazards (Ndlovu et al., 2024; Xiang et al., 2019; Pillay et al., 2024). The incorporation of drone

technology into ecological research and management represents a substantial advancement, fostering more timely and informed decision-making processes.

However, despite the valuable data drones can provide, significant challenges remain, particularly concerning statutory regulations governing operations. Many countries impose stringent restrictions on where and how UAVs can be operated. For instance, in some regions of the Global South, regulations dictate the take-off mass, maximum flight altitude, and designated operational areas (Lally et al., 2019). In South Africa, for instance, the Civil Aviation Authority stipulates that remotely piloted or toy aircraft should not be operated within 50 meters of any person or group (Sibanda et al., 2021).

To fully leverage the potential of these sensors, reliable and efficient data retrieval processing techniques are essential. Techniques such as RTMs will be crucial in ensuring that the data collected is accurate, robust, and processed swiftly, enabling effective analysis and application in environmental monitoring.

2.4.3. Biophysical and biochemical properties

Figure 2.7 illustrates that the LAI is the plant attribute used most, for example Qin et al. (2024); Binh et al. (2022); Zhao et al. (2023); Darvishzadeh et al. (2011); Darvishzadeh et al. (2008) and others. This finding is consistent with many studies in the literature. For instance, Berger et al. (2018) noted that many studies utilized LAI when evaluating the capabilities of the PROSAIL Model to determine biophysical and biochemical variables from various land cover types. In another study, Ishaq et al. (2023), found that LAI was the most utilized plant attribute when conducting a systematic review of RTMs for Crop Yield Prediction and Crop Traits Retrieval.

This can be attributed to easier retrievability than other attributes and its influence over the spectrum variations dominating some spectral regions (Verrelst et al., 2015). Notably, LAI serves as a critical determinant of changes across the spectra, suggesting that it is the primary driver of variations in specific spectral regions (Verrelst et al., 2016). Moreover, a direct correlation has been established between LAI and flavonoid content. For example, Zhou et al. (2023) employed the SAIL model to simulate LAI to estimate flavonoid concentrations in Ginkgo plantations, using UAV-borne LiDAR and hyperspectral data. In addition, LAI plays a vital role in the detection of plant stress and other Eco physiological processes. Consequently, the quantification of LAI is essential for long-term global monitoring of climate impacts on

vegetated ecosystems, as well as for identifying dynamic changes in productivity (Zheng and Moskal, 2009).

Even with many studies estimating LAI, issues have arisen when using the PROSAIL model. For example, the reflectance values can become saturated, especially at higher LAI values, due to the complex structure of the specific vegetation analysed (Berger et al., 2018:13). On the other hand, during the earliest and latest growth stages, when LAI is low, the background reflectance dominates the signal, strongly influencing the LAI retrieval (Verger et al., 2011). A solution proposed by Danner et al. (2017), was to use two different background types in the inversion - normal soil reflectance for the vegetative stage, and a spectrum of non-photosynthetic vegetation for the maturing stages of wheat. Also, incorporating PROSAIL-M proposed by Proctor et al. (2017) for senescence vegetation, can help overcome the issue of background dominance that causes problems for LAI retrieval at the earliest and latest growth stages.

LAI is closely followed by chlorophyll content (Cab, LCC, and CCC), as demonstrated in Figure 2.7 (Darvishzadeh et al., 2019; Verrelst et al., 2010; Adeline et al., 2016; Ali et al., 2021; Ali et al., 2020; Arellano et al., 2016; Berni et al., 2011; Clevers and Kooistra, 2011; Croft et al., 2013; Darvishzadeh et al., 2008). This is attributed to the easy retrievability of chlorophyll content data, particularly through non-destructive techniques such as SPAD chlorophyll meters (Raymond et al., 2014), which has contributed to its widespread use in remote sensing studies (Zhang et al., 2022). Chlorophyll content is the second most frequently estimated variable in remote sensing applications, owing to its close relationship with foliar chemistry parameters, notably nitrogen (Lehnert et al., 2014).

This strong correlation between chlorophyll and nitrogen content enables the use of chlorophyll as an indicator of plant health and early stress detection (Berger et al., 2018). The ability to monitor chlorophyll content non-destructively and its sensitivity to changes in plant physiology make it a valuable parameter for remote sensing-based vegetation monitoring and assessment. It is important to note that the estimation of the CCC produced lower accuracy than those measured at the leaf scale (Ali et al., 2021; Darvishzadeh et al., 2008; Darvishzadeh et al., 2010). This is due to other structural factors like LAI and Average Leaf Inclination Angle (ALIA), which have a strong influence on the canopy-level spectral signal. These structural variables can interfere with the chlorophyll signal at the canopy scale (Berger et al., 2018).

2.4.4. Application of Radiative Transfer Model (RTM)

2.4.4.1. One-dimensional (1D) models

This review noted that the one-dimensional or turbid media models, with PROSAIL and PROSPECT RTMs, are the most widely adopted (Figure 2.8). These findings are similar to Jacquemoud et al. (2009) who conducted a systematic review of PROSPECT and SAIL (known as PROSAIL combined) applications using data from Web of Science and Institute of Science Information (ISI) between 1992 and 2002. In the study, Jacquemoud et al. (2009) found that among all the models published, PROSAIL was the most popular. The model's popularity could be attributed to their inherent simplicity that require few input parameters, hence quick inversion (Verrelst et al. 2019).

The PROSAIL model was predominantly used in grassland vegetation (55%) as shown in Figure 2.9 (Atzberger et al., 2013; Atzberger et al., 2015; Cho et al., 2014; Clevers and Kooistra, 2011; Darvishzadeh et al., 2008; Darvishzadeh et al., 2010; Qin et al., 2024; Zhang et al., 2023; Zhu et al., 2023; Dehghan-Shoar et al., 2023; Tsele and Ramoelo, 2024). This is due to grasslands being characterized by relatively homogeneous and open canopies, with a lower density of vegetation. The PROSAIL model is known for its ability to accurately simulate the radiative transfer in sparse canopies, making it an appropriate choice for grassland vegetation where light penetration and interaction with the underlying soil surface are significant factors. In contrast, PROSAIL model has also been employed in forested areas (28%) (Figure 2.9). The PROSPECT model has been combined with different versions of the SAIL model (i.e., GeoSail, 2M-SAIL, 4SAIL, or 4SAIL2) e.g. (Singh et al., 2023; Zhang et al., 2005; Shi et al., 2018; Yebra et al., 2018a; Yebra et al., 2018b; Zhang et al., 2006; Koetz et al., 2007; Luo et al., 2019a; Luo et al., 2019b; Huemmrich, 2001), such as to account for canopy heterogeneity (Jacquemoud et at., 2009). Hence, allowing for its application in various vegetation types, including forests. Despite its original formulation for homogenous like environment, the PROSAIL model effectively captures radiative transfer processes in forest canopies, encompassing light absorption, scattering, and transmission.

On the other hand, the PROSPECT model has been widely used in publications. Its versatility allows for adoption in various landscapes (e.g., grassland, forest, savanna, shrub) (He et al., 2024; Cho et al., 2014; Arellano et al., 2016; Arellano et al., 2017; Ali et al., 2016; Féret et al., 2011; Féret et al., 2019; Proctor et al., 2017; Sibiya et al., 2024). This wide applicability contributes to its widespread usage within the scientific community. One well-known version is PROSPECT-5 (Feret et al., 2008), which is commonly employed. However, PROSPECT-5

is limited in accurately simulating spectra of leaves containing decay pigments due to the absence of specific absorption coefficients. To overcome this limitation, a new parameter, anthocyanins, which belongs to the flavonoid family, resulted in the development of the PROSPECT-D model (Feret et al., 2017) and recent version PROSPECT-PRO (Feret et al., 2021). Another modification to the PROSPECT model called PROSPECT-5M was carried out by Proctor et al. (2017), enabling the model to simulate the senescent and decomposition stages more accurately. These advancements and modifications over time, and continued evolvement have contributed to the wide usage and development of the RTMs.

2.4.4.2. Three-dimensional (3D) models

The DART and FLIGHT models are the most used tools in remote sensing, standing out among other 3D models (Figure 2.8). DART is particularly notable for its wide-ranging capabilities and accuracy in simulating how light interacts with different landscapes. It has been validated through numerous studies, demonstrating its reliability in remote sensing research. Key studies include those by Adeline et al. (2016); Banskota et al. (2015); Gascon et al. (2004); Juszak et al. (2017); Kükenbrink et al. (2021); Liu et al. (2019); Lukeš et al. (2020); Malenovský et al. (2008); Miraglio et al. (2019); Pinheiro et al. (2006); Regaieg et al. (2021); Yanez-Rausell et al. (2015). The basic version of DART, called DART-FT, uses a discrete ordinates method and has been thoroughly tested across various landscapes, making it a popular choice among researchers (Malenovský et al., 2008; Gascon et al., 2004).

However, its high computational demands make it challenging to scale up (Regaieg et al., 2021). To address these issues, the new DART-Lux version incorporates advanced algorithms like Bidirectional Path Tracing (BDPT) and Monte Carlo integration. These advancements enhance computational efficiency and decrease memory usage. Recent validation studies, such as those by Regaieg (2023), show that DART-Lux performs similarly to DART-FT while offering significant improvements. This positions DART-Lux as a promising tool for simulating complex landscapes in remote sensing.

In contrast, the FLIGHT model ranks second among 3D models after DART. It excels in simulating diverse environments, from grasslands to intricate forest structures, thanks to its foundation in the Monte Carlo ray tracing method, leading to several publications, for example, Berni et al. (2011); Bye et al. (2017); Hernández-Clemente et al. (2017); Leonenko et al. (2013); Melendo-Vega et al. (2018); Montesano et al. (2015); North et al. (2010); Verrelst et al. (2012). This versatility allows FLIGHT to accurately predict vegetation canopy reflectance.

Its efficient geometric representation of tree crowns ensures that simulations are both realistic and resource-efficient. The model's participation in the RAMI exercise further confirms its robustness in practical applications (Darvishzadeh et al., 2008).

However, most 3D models, such as FLIGHT, often face challenges when simulating multilayered ecosystems (Melendo-Vega et al., 2018). These ecosystems involve complex interactions between soil, understory vegetation, and tree canopies, which can complicate the accuracy of simulations. To overcome these limitations, several studies have adopted coupled modelling strategies. For instance, Melendo-Vega et al. (2018) integrated the 1D PROSAIL model with FLIGHT to enhance the simulation of spectral properties in multilayered tree-grass ecosystems. Similarly, Hornero et al. (2021) developed FluorFLIGHT by combining FLUSPECT and FLIGHT to assess the contribution of understory sun-induced chlorophyll fluorescence. Hernández-Clemente et al. (2017) also employed FluorFLIGHT to address issues related to aggregated pixels in heterogeneous forest canopies, highlighting the significant impact of background and shadow effects on model performance. The FluorFLIGHT model demonstrated substantial improvement in accuracy when accounting for aggregation effects, with the coefficient of determination (R^2) increasing from 0.42 (without consideration) to 0.70 (with consideration), indicating a more precise representation of the ecosystem dynamics.

These Monte Carlo Raytracing (MC) based models are crucial for accurately simulating complex multilayered ecosystems by effectively modelling interactions between vegetation layers and enhancing the representation of spectral properties and ecosystem dynamics. However, traditional MC models require significant processing time, limiting their practicality. To enhance simulation speed for Bidirectional Reflectance Factor (BRF) distributions, a variance reduction technique called "photon spread" has been introduced. Zhao et al. (2015) developed a new MC model named Weighted Photon Spread (WPS), which combines the concepts of weight reduction and photon spread.

This model has been shown to accurately reproduce the angular distribution of canopy BRFs, as confirmed by comparisons with field measurements and the RAMI Online Model Checker. Zhao et al. (2016) further advanced this approach by integrating fluorescence modelling with the Weighted Photon Spread method, resulting in the FluorWPS model. FluorWPS effectively captures both spectral and angular distributions of Sun-Induced Fluorescence (SIF), achieving a R^2 greater than 0.92 and a root mean square error (RMSE) below 0.066 $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$.

This demonstrates its reliability and potential for accurate modelling in remote sensing applications.

While the 3D models are promising in simulating heterogenous and complex canopy, using 3D models usually have difficulties, including the construction of realistic 3D vegetation scenes and the computational requirements to efficiently generate training datasets. Therefore, the emergence of LIDAR technology has attracted significant attention and demand for its ability to capture highly precise 3D structural information (Coops et al., 2021). This has resulted in many studies utilizing LIDAR for structural information, which is crucial information for the parameterization for 3D models (Zhao et al., 2024; Berni et al., 2011; Zhou et al., 2023; Bye et al., 2017; Du et al., 2021; Jin et al., 2021; Koetz et al., 2007; Kükenbrink et al., 2021; Montesano et al., 2015; North et al., 2010; Shi et al., 2018; He et al., 2024; Zhao et al., 2024). However, with the increase in the use of 3D models, there is a corresponding rise in computational complexity, which necessitates the need to enhance the 3D RTMs for greater efficiency. Several approaches have been proposed to address this complexity, such as Graphics Processing Units (GPUs) and semi-empirical methods.

The GPU plays a crucial role in enhancing the efficiency of 3D radiative transfer models (Bailey et al., 2014). By leveraging the parallel processing capabilities of GPUs, researchers can significantly accelerate the computation of photon transport in complex media, enabling the handling of larger datasets and more intricate geometries (Owens et al., 2007). For instance, Bailey et al. (2014), successfully developed a scalable and efficient model for radiative transfer in complex canopies, specifically by utilizing optimized GPU ray tracing techniques.

On the other hand, the semi-empirical method is an innovative approach that blends modern 3D radiative transfer simulations with traditional analytical equations (Jiang et al., 2020). This method has been utilized in various studies to enhance the accuracy and efficiency of remote sensing. For example, Qi et al. (2023) introduced a new method called Semi-LESS for 3D modeling of light interactions with plant canopies. This method allows for quick and accurate calculations of how much light is reflected by plants, considering different leaf properties and soil types across various wavelengths. According to Zhao et al. (2024), one of the key advantages of Semi-LESS is its low demand for computer power. It only requires one simulation for each specific plant setup, making it very efficient.

Compared to the base LESS, Semi-LESS is 320 times faster for these calculations. This significant improvement indicates that Semi-LESS has great potential for applications in

remote sensing, which involves studying the Earth's surface through satellite or aerial imagery. Similarly, Zhao et al. (2024) also used Semi-LESS method, which combines a semi-empirical approach with a 3D RTM and a 1D residual network. This method effectively retrieves LCC from images taken by UAVs and LiDAR data.

To effectively utilize these RTMs, inversion methods become essential. Inversion can be performed using techniques such as LUT inversion or hybrid approaches which combine machine learning and RTM models.

2.4.5. Hybrid inversion

2.4.5.1. Look-up table (LUT)

Over the past decade, the LUT method has gained significant traction in remote sensing research of RTM (Adeline et al., 2016; Ali et al., 2016; Atzberger et al., 2015; Banskota et al., 2015; Bye et al., 2017; Darvishzadeh et al., 2008, 2010, 2019; Dorigo et al., 2009; He et al., 2019). This popularity can be attributed to several key advantages, including faster computations compared to iterative optimization techniques, the ability to avoid local minima, a common challenge in optimization, and the incorporation of various constraints and a priori information, which enhances the accuracy and reliability of estimates (Berger et al., 2018). The foundational study by Weiss et al. (2000) introduced the LUT approach for inverting the PROSAIL model, analyzing the necessary LUT size and the optimal number of best-fit solutions for reliable estimates of biophysical variables such as LAI and chlorophyll. This pioneering work sparked a series of follow-up studies that further refined LUT-based inversion methodologies (Darvishzadeh et al., 2008; Adeline et al., 2016; Ali et al., 2021; Féret et al., 2011, 2019; Kennedy et al., 2020). These studies revealed that multiple combinations of biophysical variables can yield similar reflectance signatures, highlighting the need for a priori information to improve estimation accuracy. However, the LUT technique can be computationally intensive, particularly when handling large datasets (Kimes et al., 1998). As the availability of updated hyperspectral data (e.g., EnMAP, PRISMA) continues to rise, researchers may soon encounter the limitations of LUT-based methods.

2.4.5.2. Machine learning-based

Machine learning-based retrieval algorithms that use various RTMs in a hybrid approach are likely to become more popular than traditional look-up table methods in the coming years. This trend is supported by studies showing their high accuracy and transferability for example, Ali et al. (2021); Ali et al. (2020); Ali et al. (2016); Döpper et al. (2022); Fang and Liang. (2005);

Féret et al. (2019); Jin et al. (2021); Masemola et al. (2016) and others. Figure 10 indicates that ANNs are the most used method after LUT. This popularity can be traced back to the mid-1990s, when ANNs first emerged as a solution for handling large datasets (Verrelst et al., 2015; Dorigo et al., 2007). Over time, other machine learning algorithms, such as GPR, partial least squares regression (PLSR), support vector machines (SVM), and RF, along with recent deep learning techniques, have also gained traction. These methods combine the flexibility of statistical approaches with the physical principles of radiative transfer.

However, a challenge with hybrid approaches is that RTM data may contain outliers, which can negatively impact model performance (Verrelst et al., 2016). Tsele et al. (2024) noted that these outliers could fall outside the typical reflectance range. To address this issue, active learning has been proposed. This approach aims to: (i) eliminate non-diverse and potential outliers from the RTM-simulated reflectance samples, and (ii) optimize the training dataset to include only the most informative samples necessary for improving retrieval accuracy (Pasolli et al., 2012).

Active learning has been effectively applied in agricultural settings (Verrelst et al., 2016; 2020; Berger et al., 2021; and others), however, studies in natural ecosystems are limited. For example, Binh et al. (2022) used an active learning-based PROSAIL model to retrieve LAI from Sentinel-2 data, and Tsele et al. (2024) applied active learning to retrieve LAI and LCC in grassland ecosystems. This highlights a need for more research on active learning applications in natural ecosystems.

2.4.5.3. Vegetation Indices

Vegetation indices (VIs) are commonly used to estimate the biophysical and biochemical properties of plants using various radiative transfer models (RTMs). However, their popularity has decreased over time, as shown in Figure 2.10. Many studies still favor VIs because they are simple and quick, yielding meaningful results (Broge and Leblanc, 2001; Biriukova et al., 2020; Casas et al., 2014; Clevers & Kooistra, 2011; Adeline et al., 2016; Dusseux et al., 2011; Miraglio et al., 2019; Shi et al., 2018; Verrelst et al., 2010; Hernández-Clemente et al., 2012; Zhu et al., 2023; Zarco-Tejada et al., 2004; Lehnert et al., 2014; Cheng et al., 2006). Despite their advantages, there is a noticeable trend away from parametric methods like VIs. This shift may be attributed to their long history of use and known limitations, such as a lack of transferability (Baret and Buis, 2008; Berger et al., 2018). Moreover, VIs do not make full use of the spectral resolution available, especially in hyperspectral datasets (Atzberger et al., 2015).

2.4.6. Findings, challenges, and prospects

The findings of this review suggest some future research directions and knowledge gaps that researchers should consider. Addressing these will help improve the scope and accuracy of using RTM to retrieve biophysical and biochemical properties in terrestrial biomes:

- There is evidence of uneven distribution of research conducted between the Global North and South. Only a few studies utilized RTMs to retrieve terrestrial biomes plant traits in developing regions such as Africa.
- In recent years, significant efforts have focused on providing freely available data for monitoring various ecosystems. The launch of new sensors, such as EnMAP and PRISMA, has created an unprecedented data stream for land monitoring, accessible to a diverse user community. Additionally, UAVs equipped with multispectral, hyperspectral, and LiDAR technology have become increasingly popular in this field. However, the effective use of these advanced platforms often requires specialized expertise and substantial computational resources. Unfortunately, such resources may not be widely available in many countries, especially in developing regions like Africa.
- Most studies on plant attributes have primarily focused on metrics such as LAI and chlorophyll levels, while very few have examined secondary metabolites like flavonoids. A significant reason for this oversight is that no RTM currently incorporates flavonoids as an input parameter, making it challenging to simulate their effects. However, recent work by Zhou et al. (2023) has demonstrated a direct correlation between leaf area index and flavonoids using the SAIL model. This finding opens the door for further research to explore models like PROSAIL in relation to flavonoids. Such exploration could enhance our understanding of how chlorophyll affects flavonoid production, providing valuable insights for plant ecology and remote sensing applications (Zhou et al., 2023).
- Significant advancements have been made in RTMs and sensors for quantifying SIF. The development of FLUSPECT, which directly incorporates fluorescence, allows for the simulation of SIF in both 1D and 3D scenarios. The 1D models, which assume vegetation consists of homogeneous layers, include combinations like FLUSPECT with SAIL, resulting in FluorSAIL (Zarco-Tejada, 2005). Other notable 1D models include FLSAIL (Rosema et al., 1991) and SCOPE (Van der Tol et al., 2009), as well as mSCOPE (Yang et al., 2017), which considers heterogeneity. In contrast, 3D SIF models provide a more realistic representation by accounting for the heterogeneity of

vegetation, addressing the limitations of 1D models. Some key 3D models include FliES-SIF (Sakai et al., 2020), FluorFLIGHT (Hernandez-Clemente et al., 2017; Hornero et al., 2021), FluorWPS (Zhao et al., 2016), DART (Malenovsky et al., 2021), FluCVRT (Kallel, 2020), and DART-lux (Regaieg et al., 2023), all of which simulate canopy types for SIF assessment. Moreover, forthcoming sensors, such as the Fluorescence Explorer (FLEX), offer improved spatial resolution (300 m) and spectral resolution (0.1 nm), presenting new opportunities to assess the dynamics of actual photosynthesis through SIF measurements. These advancements pave the way for better understanding and monitoring of plant health and photosynthetic processes.

- The emergence of DART-lux has significantly enhanced the ability to simulate large landscapes in remote sensing. In contrast to DART-FX, which is computationally demanding and poses challenges for scaling up, DART-lux offers a more efficient approach. This advancement allows researchers to analyze extensive areas more effectively, facilitating comprehensive studies of vegetation dynamics and improving the overall understanding of ecosystem processes (Regaieg et al., 2021).
- Using 3D models in remote sensing presents various challenges, particularly in constructing realistic 3D vegetation scenes and meeting the computational demands for efficiently generating training datasets. LiDAR technology has emerged as a valuable alternative for obtaining detailed 3D information about vegetation structures. To address computational requirements, model acceleration is crucial. This can be achieved using more powerful hardware, such as GPUs, which enhance processing capabilities (Bailey, 2014). Additionally, empirically based acceleration methods have been developed to improve efficiency in model performance (Qi et al., 2023; Zhao et al., 2024). These approaches help streamline the simulation process, making it more feasible to utilize advanced 3D models in large-scale studies.
- Equally important, recent advancements in less intensive 3D models have shown comparable performance in retrieving various plant traits. Models like INFORM and PARAS appear particularly suitable for studies requiring efficient computational resources while maintaining accuracy (Yanez-Rausell et al., 2015; Binh et al., 2022; Ali et al., 2021; Ali et al., 2020; Ali et al., 2020; Ali et al., 2016; Darvishzadeh et al., 2019; Schlerf and Atzberger., 2012). Given their effectiveness, these less intensive models should be prioritized for exploration in future research, as they can facilitate broader

access to 3D remote sensing applications without the high demands associated with more complex models.

- The use of hybrid inversion methods, which combine machine learning techniques with RTMs, is expected to become more popular than traditional LUT methods in the coming years. However, these hybrid approaches may encounter challenges related to outliers, which can adversely affect model performance. To mitigate this issue, active learning strategies have been proposed as a solution (Binh et al., 2022; Tsele et al., 2024). Active learning can enhance the robustness of hybrid models by selectively querying the most informative data points, thereby improving accuracy and reducing the impact of outliers on the overall results.
- Finally, there are multiple free open-source software platforms available that facilitate the execution of various RTMs. One prominent example is the ARTMO toolbox (artmotoolbox.com), which operates within the MATLAB software environment, consisting of various models ranging from 1D to 3D. Also, Sentinel Application Platform (SNAP) toolbox, which is a freely accessible software package featuring plugins and algorithms for running the PROSAIL model (Ali et al., 2020). Additionally, QGIS, a free and open-source GIS software, enables users to run the PROSPECT, PROSAIL, and INFORM models. The EnMap-box plugin in QGIS offers a user-friendly interface. Moreover, several built-in packages in the Python programming language, such as prosail, pyPro4SAIL, and pyprosail, are available for users to leverage for RTM applications. Further resources include DART (https://dart.omp.eu/), Eradiate (https://www.eradiate.eu/), and the LESS model (https://lessrt.org/), all of which provide valuable tools for researchers working with radiative transfer modeling. These platforms significantly enhance accessibility and facilitate a wide range of applications in remote sensing and vegetation analysis.

2.5. Conclusion

This study sought to systematically review RTM's effectiveness in retrieving biophysical and biochemical properties from terrestrial biomes. The results of this study indicate that among all the models published, PROSPECT and PROSAIL have been the most widely used over the last two decades. This is attributed to their exceptional qualities, such as robustness, simplicity, and reliability. Additionally, the model (both PROSPECT and PROSAIL) can be downloaded from various portals and is freely available online. Six biomes were identified in the reviewed studies, with forest and grassland having the highest frequencies, respectively. The most

retrieved variables include LAI and chlorophyll (Cab, LCC, CCC), mostly retrieved using the LUT method established and tested using various RTM inversions. LUT is closely followed by ANNs and other machine learning regression algorithms, which shows an increasing trend attributed to availability, effectiveness, and speed.

It is recommended that with the newly developed spaceborne imaging spectroscopy missions such as, EnMAP, CHRIS/PROBA, and PRISMA, studies should aim to adopt fast, generic, transferable, and hybrid inversion based on the RTM model and MLRAs that use approaches such as Gaussian processes regression (Verrelst et al. 2015, Berger et al. 2018). Furthermore, new satellite sensors with enhanced time, spatial, and wavelength resolution require suitable methods for faster processing that provide users with up-to-date maps of various plant traits. In this regard, it takes retrieval approaches that are both independent from in-situ data and computationally quick to analyze enormous volume of the most diverse data. Based on our findings, the RTM models combined with non-parametric methods appear to be suited to monitor and map terrestrial biomes.

2.6. Chapter summary

This chapter provides an overview of how different RTM models are used in terrestrial environments around the world to estimate biophysical and biochemical properties of terrestrial biomes. It highlights the shift from simpler 1D models to more complex 3D models that can better simulate how light interacts with plant canopies. The research shows that most studies using 1D models rely on PROSAIL and PROSPECT, while 3D models often utilize DART and FLIGHT. Additionally, there is a growing use of less complex 3D models such as INFORM, PARAS, and LESS. The most commonly retrieved plant properties across landscapes include LAI and chlorophyll content, especially in grasslands and forests, frequently estimated using methods like LUT and ANN. The chapter also summarizes the main challenges and future opportunities in using RTMs for remote sensing. One notable gap is that very few studies have used RTMs to estimate secondary metabolites and used RTMs in phenolic-rich landscape. To address this, the next chapter focuses on testing different wavelengths at the leaf level to estimate secondary metabolites like carotenoids using the PROSPECT model. This approach will help determine if focusing on a limited number of wavelengths can improve model performance.

Chapter Three:
Determining the value of different wavelength ranges of non-imaging hyperspectral reflectance to estimate carotenoid content using the PROSPECT-5 model

This chapter is based on:

Sibiya, B., Cho, M., Mutanga, O., Odindi, J., Masemola, C., and Bonnet, W., 2025. Determining the value of different wavelength ranges of non-imaging hyperspectral reflectance to estimate carotenoid content using the PROSPECT-5 model. *International Journal of Remote Sensing*, *46*(4), pp.1696-1719.

Abstract:

Carotenoids are important plant attributes offering valuable insights into the physiological condition of vegetation and serve as essential indicators for early identification of plant stress. Generally, carotenoids can be extracted using radiative transfer model (RTM) remote sensing techniques like PROSPECT that utilize the entire spectral domain (400 – 2500 nm) to retrieve carotenoid information. However, such inversions suffer from ill-posed nature due to model uncertainties. Literature suggests that selecting appropriate bands improves the RTM inversion. Hence, this study proposed a wavelength selection approach using various regions in the visible portion of the electromagnetic spectrum and bands selected by the random forest algorithm to estimate carotenoids using the PROSPECT-5 model. This study utilized three distinct datasets - savanna, tropical forest, and a combination of two. The green spectral region demonstrated the strongest performance in the tropical forest dataset (R^2=0.90, RMSE=0.71) than the savanna (R^2=0.70, RMSE =1.19) and combined (R^2=0.72, RMSE=1.11) datasets, respectively. The bands (green, yellow, and red-edge) selected by the random forest model produced the highest accuracy in the savanna dataset (R^2=0.84, RMSE=0.99), followed by combined (R^2=0.80, RMSE=1.20) and tropical forest (R^2=0.78, RMSE=1.33), respectively. Lastly, the visible region, demonstrated strong performance in the tropical forest (R^2=0.84, RMSE=0.85), followed by combined datasets (R^2=0.72, RMSE=1.15) and savanna (R^2=0.68, RMSE=1.27), respectively. The findings suggest that carotenoid retrieval should be limited to the visible portions of the spectrum as it exhibited strong performance in estimating carotenoid content across the savanna, tropical forest, and combined datasets.

Keywords: Field spectral data, carotenoids, PROSPECT-5, and Radiative transfer Model (RTM)

3.1. Introduction

Savannas and indigenous forests are distinct ecosystems with unique vegetation, biodiversity, and ecological functions. The savanna, which consists of herbaceous layer of grass species and open layer of trees, is the largest biome in Southern Africa, covering approximately 46% of the landscape. In South Africa, it covers one-third (399,600 km^2) of the county's landmass (Minasny et al., 2017). On the other hand, indigenous forests, accounting 0.56% of the country's landscape, are characterized by contiguous tree crowns within small patches, typically less than 1 km^2 (Lawes et al., 2004; Cho et al., 2015). These ecosystems are crucial for carbon sequestration and climate change mitigation (Odebiri et al., 2023). They also support biodiversity, offer habitats for wildlife, and provide a range of socio-economic benefits and services, including, food, wood, and medicine (Madonsela et al., 2017). However, these ecosystems face significant threats. For instance, over the last decades, approximately 20% of South Africa's biomes have undergone conversion to pasture, farmland, and urban developments (Odebiri, 2022). Consequently, intensive management initiatives are necessary to preserve the savanna and the remaining indigenous forest patches to prevent further loss.

To prevent losses and restore plant biodiversity, quantifying and modelling its functional diversity, measured using plant attributes, is paramount (Zhang et al., 2012). Carotenoids are essential plant attributes valuable for ecosystem functioning and plant physiology status, which can provide valuable information to effectively assess how anthropogenic and natural factors such as climate change affect plant adaptation and functioning (Féret et al., 2017; Zhang et al., 2021). Generally, carotenoids are regarded as reliable indicators of vegetation's physiological status (Huang et al., 2018). According to Zhou et al. (2017), a decrease in chlorophyll at various stages of a plant's growth cycle indicates that it is experiencing environmental stress. On the other hand, alterations in carotenoid concentrations indicate vegetation's physiological conditions. Literature also reveals that carotenoid content changes when plants are exposed to intense sunlight, high temperatures, low nitrogen, and onset of leaf senescence (Huang et al., 2018). Therefore, quantifying carotenoid content is essential for early detection of vegetation stress and detecting the senescence cycle, as well as elucidating photoprotection and light adaptation mechanisms.

In the past decades, remote sensing, with the benefits of being non-destructive, has been one of the most cost-effective approaches for retrieving plant attributes such as carotenoids (Zhang et al., 2021; Sibiya et al., 2021; Mutanga and Skidmore, 2004). The approach has also shown

great potential to quickly estimate carotenoids over large spatial extents (Sonobe and Hirono, 2023; Asner et al., 2002).

Previous studies have demonstrated progress in estimating carotenoid content using remotely sensed data, both at the leaf and canopy levels (Huang et al., 2018). However, much of this research has primarily focused on the development of spectral indices for such estimation (Chappelle et al., 1992; Datt, 1998; Blackburn, 1998, Gitelson et al., 2002, 2006; Hernández-Clemente et al., 2012). Spectral indices or empirical models generally require calibration for each specific dataset, as it is influenced by the characteristics of the given database. This constraint affects the reliability and adaptability of the method in different settings (Zhou et al., 2017).

On the other hand, physical models such as the PROSPECT (Feret et al., 2008) are rooted in understanding physical processes governing the interaction between photons and plant biochemical characteristics, which make them suitable for retrieving plant attributes such as carotenoids, as they are more robust and transferable than empirical models (Cho et al., 2014a; Wan et al., 2022). The PROSPECT-5 radiative transfer model, developed by Feret et al. in 2008, is one of the most used version within the PROSPECT family of models. This model incorporates two types of pigments related to photosynthesis in vegetation: chlorophylls and carotenoids. The leaf samples used to calibrate the PROSPECT-5 model are primarily green in color. As a result, this model is effective at simulating spectra and estimating the biochemical and biophysical properties of photosynthetically active leaves (Lu et al., 2021; Proctor et al., 2017). Given that the primary aim of this study was to estimate carotenoids, the PROSPECT-5 model provides a robust framework as it can accurately simulate the reflectance characteristics of vegetation by incorporating the optical properties of these key pigments (Lu et al., 2021). Specifically, the PROSPECT-5 model provides a valuable tool for estimating carotenoid content from leaf-level spectral data, yielding insights into the health and functioning of the studied vegetation, crucial for a wide range of remote sensing and ecological applications.

Generally, the PROSPECT inversion is employed to simulate wavelength across the entire optical domain (400 – 2500 nm), which is frequently unavailable and demands high computational processing resources (He et al., 2023). However, considering the entire spectral range, including less informative wavelengths, the inversion process may introduce unnecessary noise or inaccuracies. To address these issues and enhance the accuracy of

retrieving leaf biochemical traits, a carefully selected subset of wavelengths or bands most sensitive to the desired properties for the inversion process is necessary (Song et al., 2011, 2021). According to the literature, selecting suitable wavelength improves the inversion and reduces biased variable estimation caused by measurement and model uncertainty (Yang et al., 2021; Sun et al., 2019). For instance, literature reveals that the unique absorption properties of carotenoids in the visible portion of the spectrum allow for quantifying carotenoids content using remote sensing approaches. Hence, this study sought to investigate band selection method in the visible region of the spectra using PROSPECT-5 model to estimate carotenoids in the savanna and tropical forest using ANN inversion.

The adoption of an ANN as a hybrid approach in the PROSPECT-5 model offers several benefits compared to other machine learning algorithms. Firstly, ANNs are recognized for their superiority in capturing complex non-linear relationships within the data, a crucial advantage when working with the intricate radiative transfer processes modelled by PROSPECT-5 (Verrelst et al., 2015, 2019). Unlike more rigid parametric models, the flexible data-driven nature of ANNs allows them to adapt and learn the underlying patterns in the spectral data, leading to more accurate estimations of leaf biochemical constituents like carotenoids.

In addition to the wavelength selection method in the visible region of the electromagnetic spectrum, the present study investigated the performance of various machine learning algorithms for feature selection. RFR was chosen for its ability to handle nonlinear relationships and provide insights into feature importance (Mngadi et al., 2021). XGBoost was selected due to its superior predictive performance and capacity to identify the most informative wavelengths (Sauer et al., 2021). PLSR was employed due to its effectiveness in dealing with high-dimensional data and identifying latent variables (Yang et al., 2023). ANN was utilized for its powerful nonlinear modelling capabilities and potential to uncover complex spectral patterns (Odebiri et al., 2022). The wavelengths identified through these feature selection techniques were subsequently incorporated into the PROSPECT-5 model to enhance the model's performance and accuracy.

Hence, this study's objectives were to i) perform feature selection using different algorithms such random forest, XGBoost, PLSR, and ANN, ii) conduct the PROSPECT-5 inversion by employing the bands selected from the algorithm that exhibited the best performance, and iii) utilize different regions within the visible spectrum (blue, green, yellow, red, and red-edge)

during the PROSPECT-5 inversion process and evaluate the model's performance using three datasets: savanna, tropical forest, and a combined dataset.

3.2. Methods

3.2.1. Field survey

Similar sampling and methodological approach were employed at both the Dukuduku Tropical Forest and Savanna study sites. For the Dukuduku Tropical Forest, seven transect lines were randomly established using a pan-sharpened WorldView-2 satellite image. Each transect spanned a minimum distance of 1 km, and the tree crowns visible on the enhanced image were assigned numerical identifiers along the transects. The coordinates of these numbered tree crowns were then extracted and uploaded into a handheld Garmin Vista Cx etrex GPS device. In the Savanna landscape (Kruger National Park), the coordinates of the tree crowns were extracted from the satellite imagery using a network of roads and paths, focusing on a range of 0 to 200 m away from the road or path. These coordinates were also downloaded into the same handheld GPS device. During the survey, the pre-selected tree canopies were located and identified using a GPS device. Some tree species were found to be dominant in each landscape, as reported in previous studies (Cho et al., 2015; Cho et al., 2012; Naidoo et al., 2012).

Field data was collected in May 2010 and August 2011 for the Savanna landscape and Dukuduku Tropical Forest, respectively, to record tree species and canopy gap. Leaf samples were collected using the detached branch technique, targeting the predominant tree species in the canopy layer (upper tree layer). In addition to the canopy tree species, the data collection recorded different types of canopy gaps such as grass and shrub covers, bare soil, and recently burned patches (i.e. burned scars) along each survey transect. The tree species that were sampled are listed in Table 3.1.

Table 3.1: Leaf datasets.

Ecosystem type	Species name	No. of points Spectral sample per species	Average carotenoids content and standard deviation
savanna trees	Leaves of different ages for 8 species (*Combretum hereroense*, *Combretum molle*, *Combretum collinum*, *Euclea natalensis*, *Terminalia sericea*, *Sclerocarya birrea*, *Pterocarpus rotundifolius* and *Lannea discolour*)	78	Greater Kruger National Park area, Mpumalanga Province, South Africa.
tropical trees	Species of broadleaved trees, coastal forest (*Albizia adianthifolia, Strchynos spp, Acacia, Ziziphus mucronate, Trichilia dregeana, Strychnos spp, Lannea spp, Ficus spp, Celtis Africana, Balanites maughamii*, and *Bridelia micrantha)*	82	St Lucia, South Africa

3.2.2. Spectral measurement and wet chemical analysis

In total, 82 and 78 leaf samples were collected from the Dukuduku Tropical Forest and Savanna landscape, respectively. These samples were then subjected to spectral measurement and wet chemical analysis within 24 hours. The spectral reflectance and carotenoids contents of leaves were measured using ASD FieldSpec3(R) spectrometer (Analytical Spectral Devices, Boulder, CO, USA) equipped with a leaf contact probe (ASD Analytical Spectral Devices. 2016). A spectralon panel was used as a reference standard for the spectral measurements. The spectrometer provides a spectral range of 350 to 2500 nm with a 1 nm resolution, covering the visible, near infrared, and shortwave infrared regions of the electromagnetic spectrum. For each leaf sample, two reflectance measurements were taken and averaged to account for spatial heterogeneity within the leaf structure (Main et al., 2011). The contact probe had a 25 mm diameter and a 10 mm instantaneous field of view, with its own halogen lamp light source used for the measurements. It is important to note that the full spectral range was utilized in two ways in this study. Firstly, carotenoids exhibit specific absorption properties in the visible portion of the spectrum. Therefore, this study focused on analyzing the visible region of the spectrum (400-745 nm). Secondly, various machine learning algorithms, including RFR, XGBoost, PLSR, and ANN, were used to perform feature selection across the entire spectral range (i.e., 2100 bands). The wavelengths identified as most important by the best-performing algorithm were then further incorporated into the PROSPECT-5 RTM inversion (Feret et al., 2008).

After the spectral measurements, the leaf samples were used for wet chemical analysis to determine carotenoid concentration. Here, a leaf borer with a diameter of 18 mm was used to

collect a sample from the same area of the leaf that had been measured. The wet-lab extraction technique of Lichtenthaler and Wellburn (1983) was used to determine carotenoid concentration. Briefly, fresh weight of each leaf sample was recorded, pigments extracted in 100% acetone, and the extracts centrifuged in a microcentrifuge to pellet cell debris. According to Cho et al. (2014), its absorbance was measured with a UV–VIS spectrophotometer (WPA) at 661.2 nm, 644.8 nm and 470 nm. In these expressions A_λ denotes the absorbance at wavelength λ nm. Chlorophyll a (Ca) and chlorophyll b (Cb) were calculated from the red-region absorbances using the following equations:

$$C_a = 11.24A_{661.2} - 2.04A_{644.8} \quad (1)$$

$$C_b = 20.13A_{644.8} - 4.19A_{661.2} \quad (2)$$

Total carotenoids (Cx+c) were then estimated from the blue region absorbance with correction for the chlorophyll contributions:

$$C_{x+c} = 1000A_{470} - 2.27C_a - 81.4C_b / 227 \quad (3)$$

3.2.3. Statistical Analysis

The field spectral data generates high-dimensional data due to its ability to capture many spectral bands (i.e., 2100 bands). Due to the high correlation between independent and dependent variables, this poses a challenge for many regression analyses, such as multi-collinearity. Dealing with such high-dimensional hyperspectral data requires a robust algorithm to handle its unique characteristics. Thus, RFR, XGBoost, PLSR, and ANN were tested to estimate carotenoid content.

3.2.3.1. Random Forest Regression (RFR)

In this study, regression analysis was conducted using the random forest algorithm to identify informative wavelength for estimating carotenoids. RFR is a popular machine-learning technique for feature selection and prediction. In the context of carotenoid analysis, RFR –can help identify the specific wavelengths most informative for predicting carotenoid concentrations. RFR works by constructing an ensemble of decision trees, where each tree is trained on a random subset of the available wavelengths (Ließ et al., 2016; Mngadi et al., 2021). During the training process, the algorithm assesses the importance of each wavelength by measuring how much it contributes to the model's overall accuracy or predictive power. The importance of each wavelength is typically measured using metrics such as Gini importance or Mean Decrease Impurity (Breiman, 2001).

3.2.3.2. XGBoost

XGBoost regression is an advanced gradient-boosting algorithm widely used for regression tasks. XGBoost leverages the collective strength of numerous individual weak decision trees to generate a robust and accurate predictive model (Chen and Guestrin, 2016). XGBoost regression is important for feature selection because it can automatically rank the importance of features based on their contribution to the overall predictive performance. This algorithm operates through an iterative process of constructing decision trees, with each subsequent tree aimed at rectifying the errors made by the previous ones (Hengl et al.,2017). During this iterative process, XGBoost assigns weights to features based on their ability to minimize the objective function, which measures the model's error. By examining the accumulated weights of features across multiple trees, XGBoost identifies the most important features that impact prediction accuracy. Consequently, it enables efficient feature selection and facilitates more accurate regression modelling.

3.2.3.3. Partial Least Squares Regression (PLSR)

PLSR is a well-established technique that combines the dependent and independent variables to select the optimal variables (Song et al., 2011). In this case, PLS makes it possible to find spectral bands closely correlated to carotenoids. PLS regression is a superior method because the analysis is not done band-by-band, but a continuous spectrum is treated as a single measurement. Also, the major advantage of PLS compared to Principal Component Analysis (PCA) is that the dependent variable is considered when performing dimensionality reduction and eliminating overlapping wavelengths (Lehmann et al. 2015). Both independent variables (X) and dependent variables (Y) are reduced to principal components. Once the PLS has been performed, several components are formed, each explaining a portion of the variance.

3.2.3.4. Artificial Neural Network (ANN)

ANNs are computational models that draw inspiration from the structure and functioning of the human brain. These networks are composed of interconnected nodes, often called "neurons," organized into layers. Each neuron carries out weighted computations and transfers the results to the subsequent layer. ANNs are important for feature selection because they can automatically learn and extract relevant patterns and relationships from complex data, enabling the identification of the most informative features (Xu et al., 2020). By leveraging their ability to capture non-linear relationships and adapt to different data types, ANNs facilitate efficient and effective feature selection, improving model performance, interpretability, and computational efficiency.

The best-performing algorithm wavelengths were further incorporated in the PROSPECT-5 inversion. The model's accuracy was assessed by comparing predicted values to the measured in the test set. Validation involved calculating the root mean squared error (RMSE) and coefficient of determination (R-squared). The entire dataset (n = 160) was divided, allocating 70% as the training set and 30% as the testing dataset.

3.2.4. Simulated datasets using the PROSPECT-5 model

Utilizing the PROSPECT-5 model, 100,000 randomly simulated spectra were generated using six input variables: leaf structure parameter (N), LChl, LCar, brown pigments (Cbrown), equivalent water thickness, and Leaf mass per area. All input parameters for the PROSPECT-5 model are detailed in Table 3.2. N was fixed at 1.2 based on preliminary inversions and sensitivity tests showing negligible changes in retrieved pigments for N ≈ 1.0–1.5 under our noise conditions. Cab (chlorophyll a+b) was constrained to 20–80 µg cm^{-2} to represent peak summer chlorophyll for sun-exposed leaves at our savanna and tropical-forest sites. Car (carotenoids) was constrained to 2–15 µg cm^{-2} based on field pigment assays. Cbrown was set to 0 because only visually green, non-senescent leaves were used. Cm (dry matter content) was constrained to 0.001–0.016 g cm^{-2}, consistent with measured LMA/dry-mass values and published ranges for the two biomes. Lastly, pigments absorb light between 400 and 700 nm in the visible spectrum, water has a strong absorption in the 1000–2500 nm near-infrared, and dry matter is present throughout the entire wave range.

Therefore, the water content value was set constant as this study sought to measure carotenoid content in the visible range, where water content did not influence the outcome. Hence, a synthetic databased was constructed using the 100,000 spectra and matching parameter value sets. For the experiment, six optical domains that include, visible (400- 700 nm), blue (450-510 nm), green (510- 580 nm), yellow (585- 625 nm), red (630- 690 nm), red-edge (705- 745 nm), and bands were selected using RFR (Figure 3.2a). Table 3.2 displays the input parameter values employed in the PROSPECT-5 simulations (Feret et al., 2008). Therefore, in this study, we used ANN inversion methods on the selected wavelengths to determine a robust, accurate, and dependable approach for the inversion of PROSPECT-5 from field spectral data. The ANN with three layers (input, hidden, and output layer) was utilized in Jupyter Notebook using the sklearn package in the Python 2.7 version.

The precision of the retrievals was demonstrated through the comparison of the *in-situ* data and the prediction using coefficient of determination (R^2) (equation 4) and root mean square error (RMSE) (equation 5).

$$R^2 = 1 - \frac{\sum(Y_{measured} - Y_{predicted})^2}{\sum(Y_{measured} - Y_{measured})^2} \quad (4)$$

$$RMSE = \frac{\sqrt{\sum_{i-1}^{N}(Y_{measured} - Y_{predicted})^2}}{N} \quad (5)$$

Table 3.2: Input parameters of PROSPECT.

			Range	
Parameter	**Symbol**	**Unit**	**Min**	**Max**
Leaf structure parameter	(N)	NA	1.2	1.2
Leaf chlorophyll content	LChl	µg/cm2	20	80
Leaf carotenoid content	LCar	µg/cm2	2	15
Equivalent water thickness	EWT	g/cm2	0.021	-
Leaf mass per area	LMA	g/cm2	0.001	0.016
Brown	Cbrown	NA	0	0

3.3. Results

3.3.1. Algorithms prediction performance

Among the four tested models, the RFR was the most robust in estimating carotenoids, indicating an R^2 of 0.82 and RMSE of 1.436. This was followed by XGBoost (R^2 = 0.80 and RMSE = 1.501) and ANN (R^2 = 0.80 and RMSE = 1.641). Lastly, PLSR was the least robust, with an R^2 of 0.75 and RMSE of 1.897. Figure 3.1 illustrates the correlation between the observed and the estimated carotenoids for the constructed models.

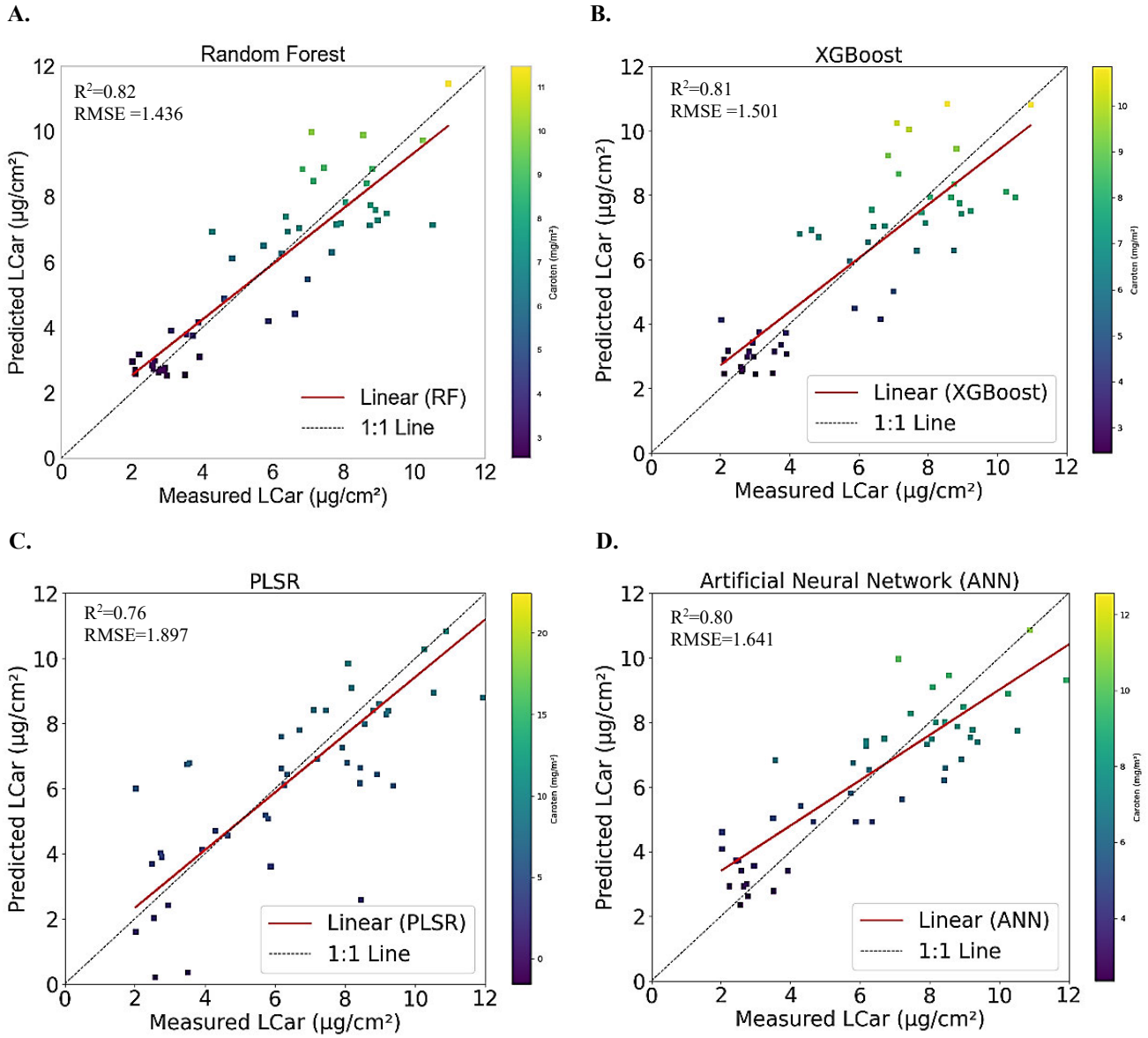


Figure 3.1. Predicted against observed carotenoids using various algorithms.

3.3.2. Variable importance selection using the RFR

For demonstration purposes, Figure 3.2 illustrates the most influential wavelength by various algorithms. However, since the RFR was the most robust, we focused on the important wavelengths of the model to estimate carotenoids (Figure 3.2a). The most influential wavelength ranges include the yellow (581, 587, 589, 595, 599, 600, 605, 606, 609, 610, 617), green (557, 558, 570), and red-edge (703, 708) regions of the spectrum. The selected bands by random forest were further tested in the PROSPECT-5 model to constrain the inversion.

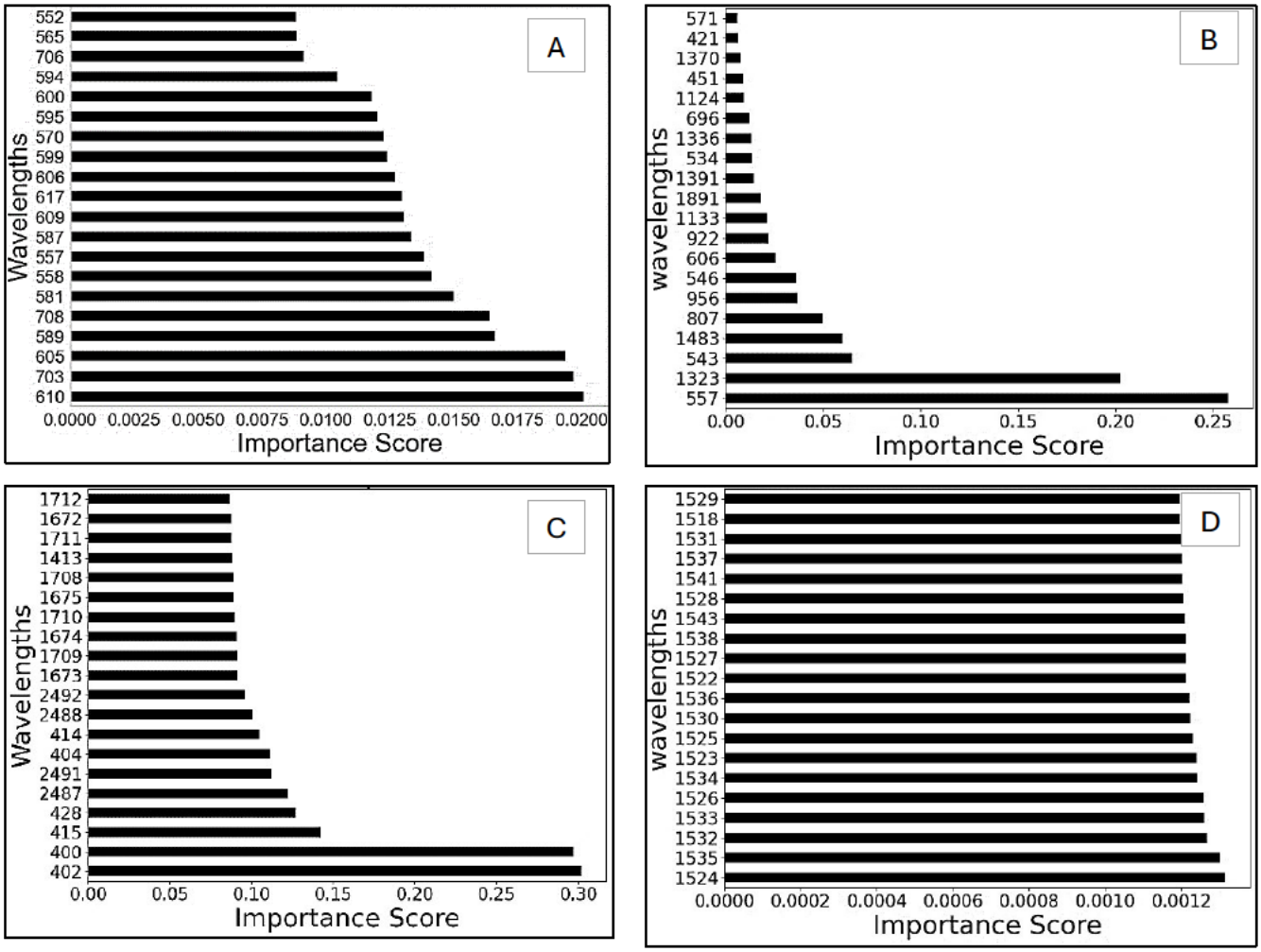


Figure 3.2. Results of various algorithms in introducing the most important wavelengths: (a) RFR, (B) XGBoost, (C) PLSR, and (D) ANN.

3.3.3. Results of PROSPECT-5 model for Savanna, tropical forest, and Combined datasets

Figure 3.3 illustrates the RMSE between measured and simulated reflectance for the savanna, tropical forest, and combined datasets, respectively. The results indicate that the tropical forest dataset produced the lowest RMSE among the three, with an average RMSE of approximately 1.55 and a maximum value below 2.58. This was followed by the combined dataset, which had an average RMSE of 1.86 and a maximum value of 2.65. The savanna dataset exhibited the highest RMSE, with an average of 1.98 and a maximum of 2.85.

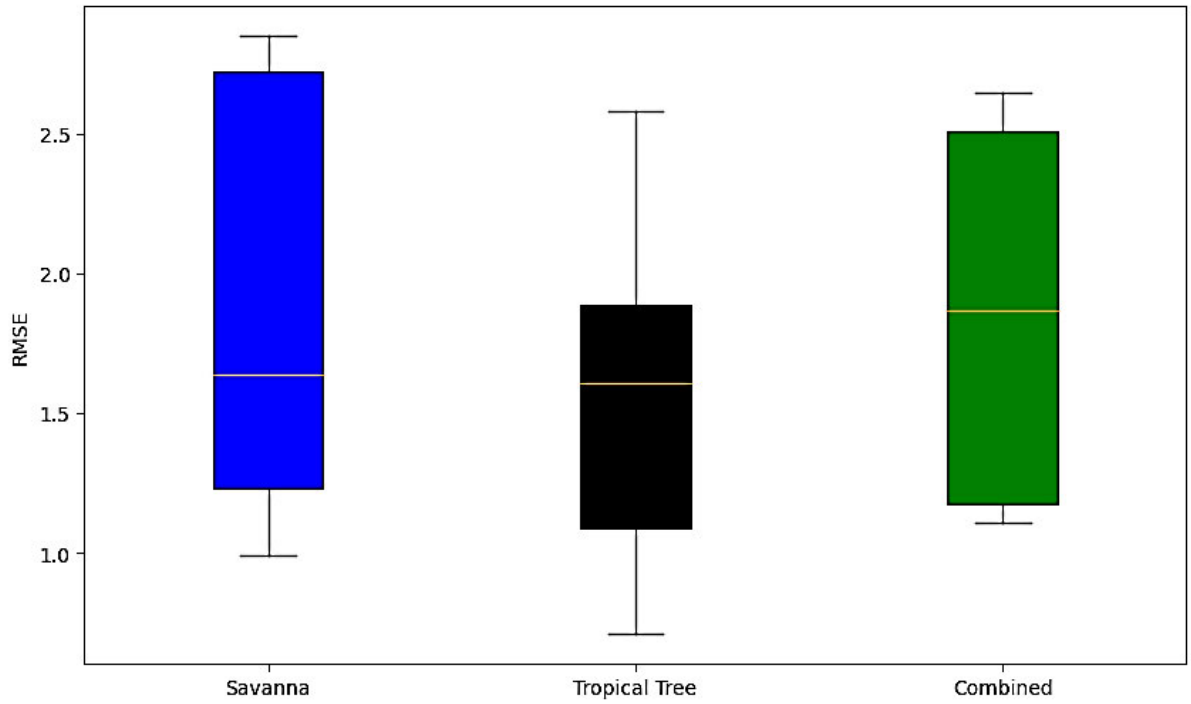


Figure 3.3. RMSE of the three datasets (savanna, tropical forest, and combined) between measured and simulated spectra using PROSPECT-5.

To further investigate the factors influencing the performance of the different models, the spectral regions with the lowest RMSE among the three datasets were identified and the simulated and measured spectra compared (Figure 3.4). The findings reveal that the green region had the lowest RMSE of 0.71 across all models (Figure 3.4C1-2), followed by the visible region with an RMSE of 0.85 (Figure 3.4c). The bands selected by the RFR had an RMSE of 0.99 (Figure 3.4g), respectively. Among the selected bands with the lowest RMSE, the majority were from the tropical forest dataset, including the visible, green, yellow, red, and red-edge regions. The blue region and the bands selected by the RFR were derived from the savanna dataset. The analysis of individual performance among the three datasets, which includes Savanna, tropical tree and combined, respectively, is shown in the next section.

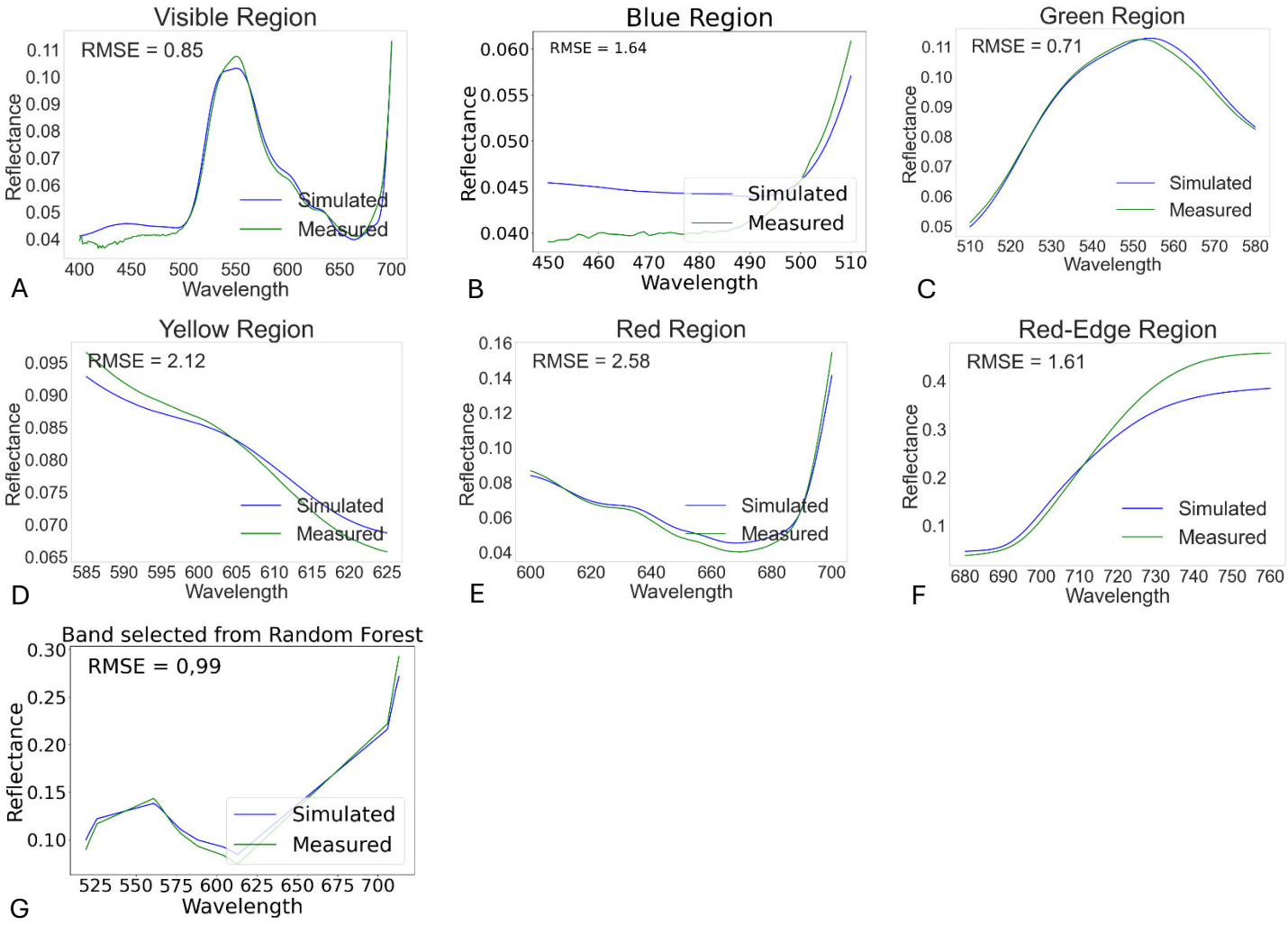
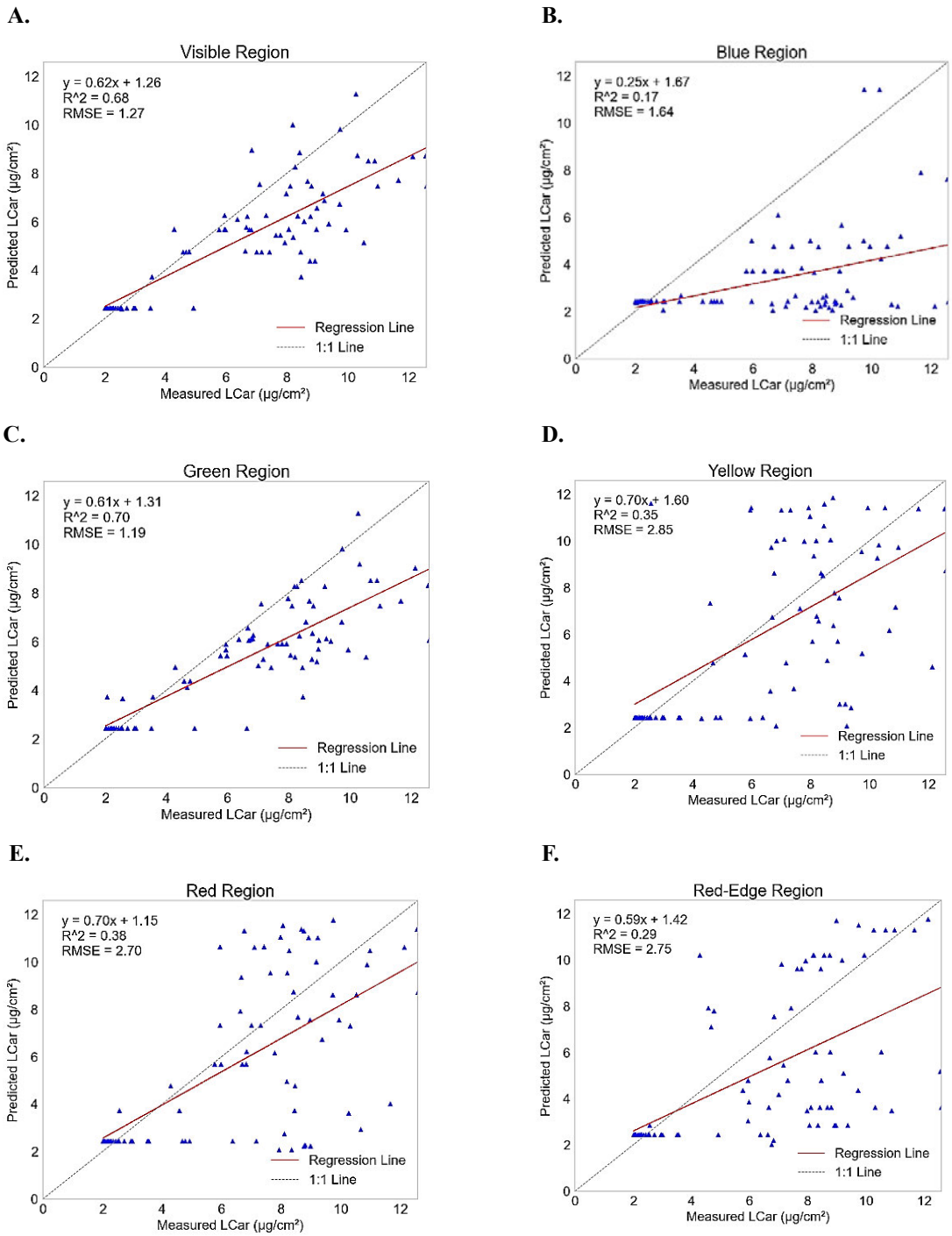


Figure 3.4. Spectral matching between simulated PROSPECT-5 and measured spectral, (first column) and differences between simulated and measured spectra (second column).

3.3.3.1. Savanna

The analysis of the savanna landscape using various spectral regions (visible, blue, green, yellow, red, and red-edge), and the bands selected from RFR showed varying performance levels. The key findings are as follows: the bands selected by the RFR demonstrated the highest overall performance, with an R^2 of 0.84 and a RMSE of 0.99 (Figure 3.5g). This approach outperformed the other spectral regions, suggesting that the band selected from the RFR is the most effective in estimating carotenoids. This was followed by the green region which also exhibited strong performance, with an overall accuracy of $R^2 = 0.70$ and an RMSE of 1.19 (Figure 3.5c). The visible region produced a respectable overall accuracy of $R^2 = 0.68$ and an RMSE of 1.27 (Figure 3.5a).

On the contrary, the blue region showed the lowest performance among all the models, with an overall accuracy of $R^2 = 0.17$ and an RMSE of 1.64 (Figure 3.5b), closely followed by the red region with an overall accuracy of $R^2 = 0.38$ (Figure 3.5e). These results highlight the importance of carefully selecting the appropriate spectral regions for analysing savanna landscapes, as different regions can perform differently. The band selected from the random forest model demonstrated the highest overall accuracy and the lowest RMSE, making it a potentially valuable approach to estimating carotenoids.



G.

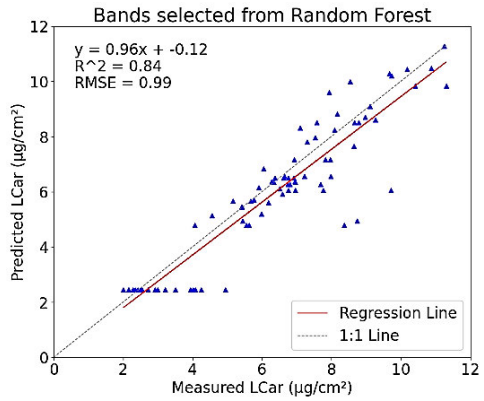


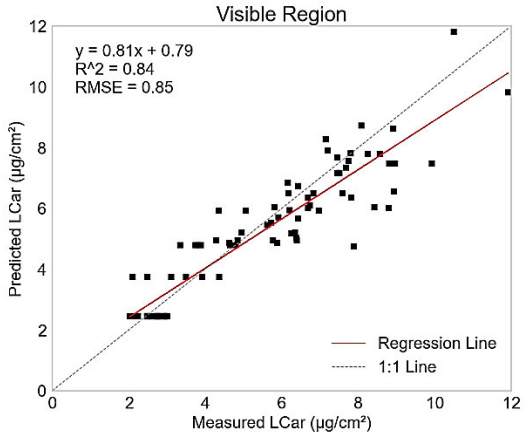
Figure 3.5. Relationships between observed and predicted carotenoids (µg/cm^2) for various spectral regions and bands selected from RFR in the Savanna landscape, derived from PROSPECT-5 inversion.

3.3.3.2. Tropical Forest

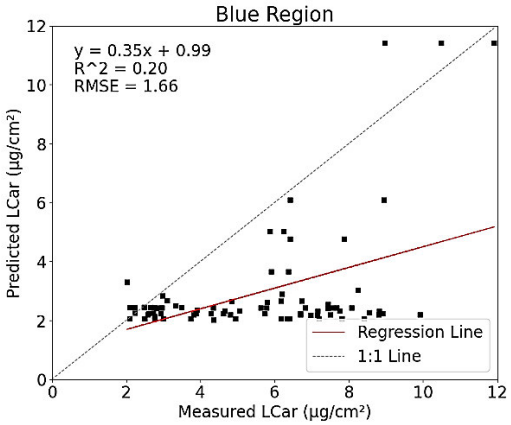
The analysis of the Dukuduku indigenous tropical forest using various spectral regions and bands selected from the RFR shows that the green region had the highest overall performance, with an R^2 of 0.90 and an RMSE of 0.71 (figure 3.6c). The visible region closely followed the green region, with an overall accuracy of $R^2 = 0.84$ and an RMSE of 0.85 (Figure 3.6a). The band selected from RFR had an overall accuracy of $R^2 = 0.78$ and an RMSE of 1.33, which places it as the third-best performing approach in this context (Figure 3.6g).

On the other hand, the blue region showed the lowest performance, with an overall accuracy of $R^2 = 0.20$ and an RMSE of 1.66 (Figure 3.6b) and was closely followed by the red region with an overall accuracy of $R^2 = 0.30$ and an RMSE of 2.58 (Figure 3.6e). These results highlight the importance of carefully selecting spectral regions for analysing tropical forest landscapes, as different regions can perform differently. The green region demonstrated the highest overall accuracy and the lowest RMSE, making it a valuable approach for this type of analysis.

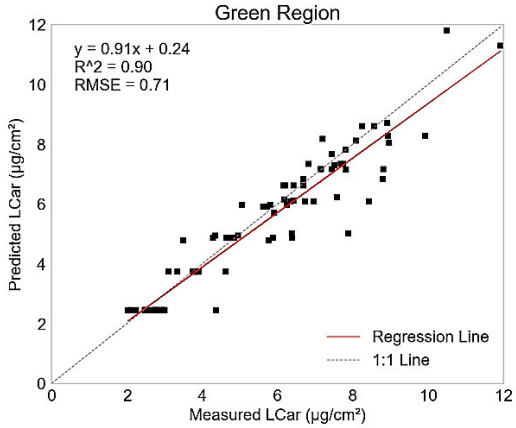
A.


Visible Region
y = 0.81x + 0.79
R^2 = 0.84
RMSE = 0.85
Regression Line
1:1 Line
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)

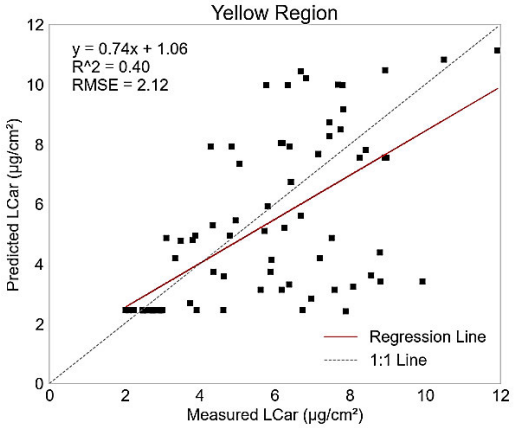
B.


Blue Region
y = 0.35x + 0.99
R^2 = 0.20
RMSE = 1.66
Regression Line
1:1 Line
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)

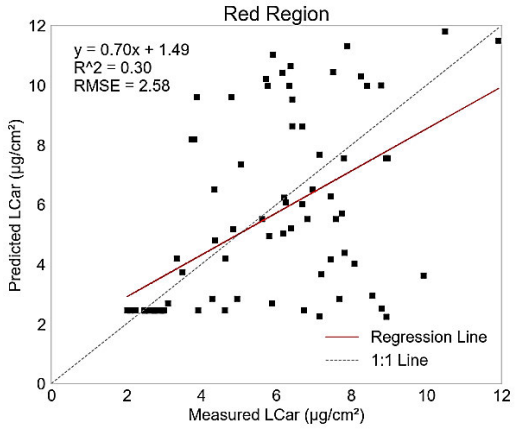
C.


Green Region
y = 0.91x + 0.24
R^2 = 0.90
RMSE = 0.71
Regression Line
1:1 Line
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)

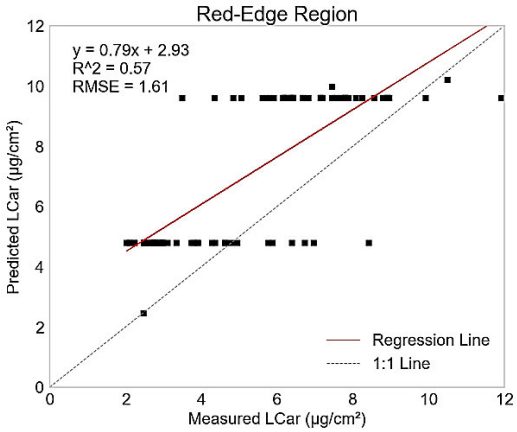
D.


Yellow Region
y = 0.74x + 1.06
R^2 = 0.40
RMSE = 2.12
Regression Line
1:1 Line
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)

E.


Red Region
y = 0.70x + 1.49
R^2 = 0.30
RMSE = 2.58
Regression Line
1:1 Line
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)

F.


Red-Edge Region
y = 0.79x + 2.93
R^2 = 0.57
RMSE = 1.61
Regression Line
1:1 Line
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)

G.

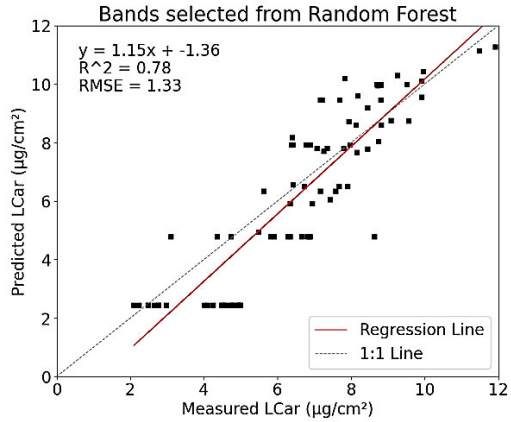


Figure 3.6. Relationships between observed and predicted carotenoids in the Dukuduku Tropical Forest using different spectral regions and bands selected from RFR.

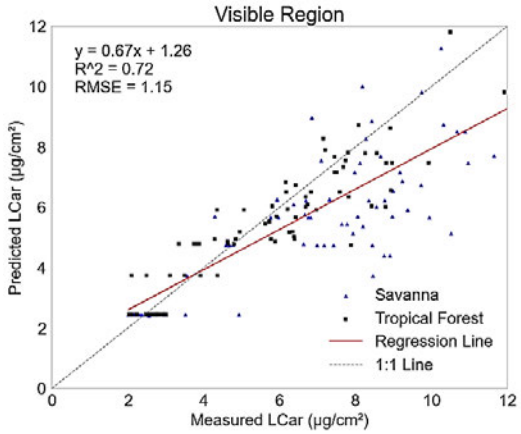
3.3.3.3. Combined

The study further evaluated the performance of both a savanna landscape and the Dukuduku tropical forest. By combining the results from these two distinct ecosystems, the analysis provides a comprehensive assessment of the efficacy of the different spectral domains. The band selected from the RFR demonstrated the highest overall performance, with an R^2 of 0.80 and an RMSE of 1.20 (Figure 3.7g). The green region also exhibited strong performance, with an overall accuracy of R^2 = 0.75 and an RMSE of 1.11 (Figure 3.7c). The visible region produced a respectable overall accuracy of R^2 = 0.72 and an RMSE of 1.15 (Figure 3.7a).

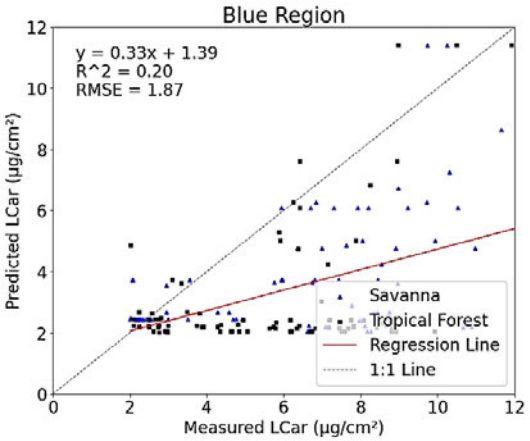
Lastly, the blue region showed the lowest performance, with an overall accuracy of R^2 = 0.20 and an RMSE of 1.87 (Figure 3.7b) and was closely followed by the red-edge region which had an overall accuracy of R^2 = 0.33 and an RMSE of 2.51 (Figure 3.7f). These results highlight the importance of carefully selecting the appropriate spectral regions for analysing both savanna landscapes and tropical forests, as different regions can perform differently. The band selected from RFR demonstrated the highest overall accuracy and the lowest RMSE, making it a potentially valuable approach for this type of analysis.

Overall, these findings highlight the importance of carefully evaluating the performance of different spectral regions and models when estimating carotenoid content in both savanna and tropical forest environments. The band selected from RFR, the green and visible region of spectra emerged as the most promising approaches for these applications.

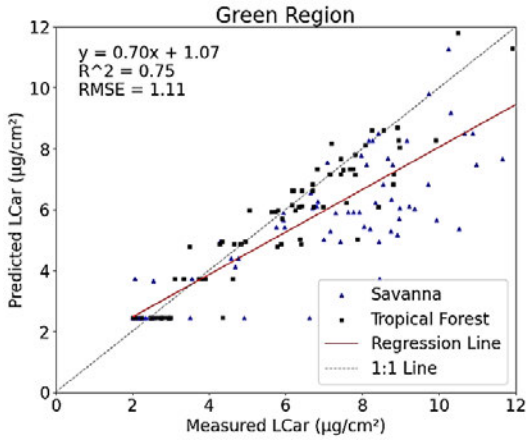
A.


Visible Region
y = 0.67x + 1.26
R^2 = 0.72
RMSE = 1.15
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)
Savanna
Tropical Forest
Regression Line
1:1 Line

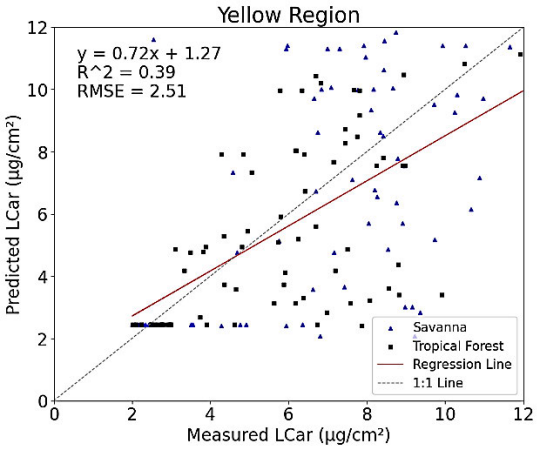
B.


Blue Region
y = 0.33x + 1.39
R^2 = 0.20
RMSE = 1.87
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)
Savanna
Tropical Forest
Regression Line
1:1 Line

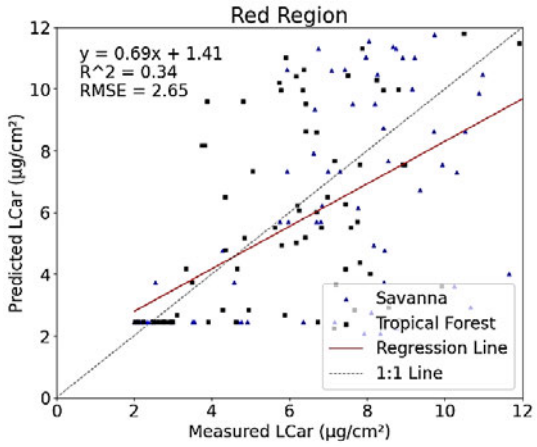
C.


Green Region
y = 0.70x + 1.07
R^2 = 0.75
RMSE = 1.11
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)
Savanna
Tropical Forest
Regression Line
1:1 Line

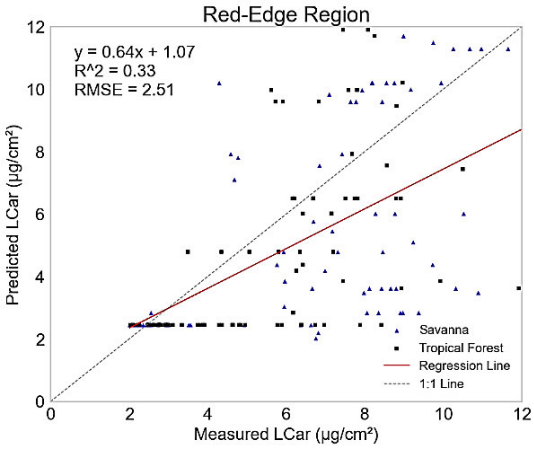
D.


Yellow Region
y = 0.72x + 1.27
R^2 = 0.39
RMSE = 2.51
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)
Savanna
Tropical Forest
Regression Line
1:1 Line

E.


Red Region
y = 0.69x + 1.41
R^2 = 0.34
RMSE = 2.65
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)
Savanna
Tropical Forest
Regression Line
1:1 Line

F.


Red-Edge Region
y = 0.64x + 1.07
R^2 = 0.33
RMSE = 2.51
Predicted LCar (µg/cm²)
Measured LCar (µg/cm²)
Savanna
Tropical Forest
Regression Line
1:1 Line

G.

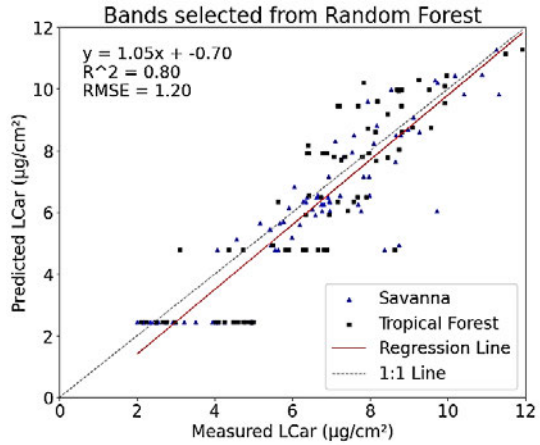


Figure 3.7. Relationships between observed and predicted carotenoids (µg/cm^2) for various spectral regions and bands selected from RFR for both Savanna landscape and tropical forest dataset, derived from PROSPECT-5 inversion.

3.4. Discussion

Traditionally, the PROSPECT model uses the entire domain reflectance (400 - 2500 nm) to estimate leaf contents. However, using waveband characteristics can significantly reduce the need for complete domain reflectance and improve the inversion performance. In this study, we proposed a wavelength selection approach utilizing PROSPECT-5-ANN models to estimate carotenoids.

The analysis of three distinct datasets, encompassing savanna, tropical forest, and a combination of the two (both savanna and tropical forest) revealed that the green region demonstrated the strongest overall performance. This is evidenced by an R^2 of 0.90 and an RMSE of 0.71, which were observed in the tropical forest dataset. This high correlation between the green reflectance and carotenoid content is consistent with findings from numerous studies that have highlighted the green region as a sensitive indicator of carotenoid dynamics across a range of plant functional types and biomes (Gamon et al., 1997; Gitelson et al., 2002; Sims and Gamon, 2002; Stylinski et al., 2002). Studies by Gitelson et al. (2002), Blackburn (2007), and Gamon et al. (1997) for instance demonstrated similar strong correlation between carotenoid levels and reflectance in the green spectral region. These and other works have helped establish the green region as a sensitive indicator of carotenoid status in plants, with applications in remote sensing, plant stress monitoring, and other vegetation applications.

The strong correlation between carotenoid levels and green reflectance observed in the tropical forest dataset can be attributed to the relatively stable leaf pigment composition and optical properties in the evergreen ecosystem. Hence, the tropical forest showed the lowest RMSE

among all three tested data to estimate carotenoids (Figure 3.3). Furthermore, the high and consistent chlorophyll content in the Dukuduku forest may reduce the interference of other pigments (e.g., anthocyanins) on the green reflectance signal, allowing the green region to directly capture the variations in carotenoid content. In the savanna ecosystem, the green region also exhibited the second-best performance, with an R^2 of 0.70 and an RMSE of 1.19. A similar trend was observed in the combined dataset, where the green region again demonstrated the second-best performance, with an R^2 of 0.72 and an RMSE of 1.11. These findings further corroborate the importance of the green region as a sensitive indicator of carotenoid dynamics across different biomes.

The bands selected by the RFR produced the second-highest accuracy, and this was observed in the savanna ecosystem, with an R^2 of 0.84 and RMSE of 0.99. For the combined dataset, the band selected by the RFR produced the highest accuracy, while for the Dukuduku tropical forest, it produced the third-highest accuracy. The high performance of the bands chosen by the RFR may be attributed to the ability of the RFR to gain a holistic understanding of the interconnections between vegetation attributes throughout the entire spectrum, allowing for the identification of wavelengths that contain more conclusive information for estimating carotenoid content (Ghosh and Behera, 2018; Breiman, 2001). The wavelengths selected by the RFR included the green, yellow, and red-edge regions of the visible spectrum, which aligns with this study's findings, suggesting the visible region's importance for carotenoid estimation.

The visible region produced the second-highest accuracy for the Dukuduku tropical forest dataset (R^2= 0.84 and RMSE=0.85) and the combined dataset, (R^2= 0.72 and RMSE=1.15) and the third-highest performance for the savanna ecosystem (R^2= 0.68 and RMSE=1.27). This suggests that the visible region may be a suitable alternative for estimating carotenoids, particularly in situations where more advanced modelling techniques like RFR are not feasible or accessible. The visible region has been used extensively in literature for retrieving carotenoid content from remotely sensed data, as carotenoids influence the reflectance characteristics in this region of the spectrum (Blackburn, 2007; Gitelson and Merzlyak, 2004; Gamon et al., 1997; Sonobe and Wang, 2018). The visible region demonstrated strong performance in estimating carotenoid content across the Dukuduku tropical forest and combined (Dukuduku and savanna) datasets, achieving the second-highest accuracy. This suggests that the visible region effectively captures the distinct optical signatures created by varying carotenoid concentrations in the leaves, likely due to the high and consistent chlorophyll content in the evergreen Dukuduku ecosystem reducing interference from other pigments.

The visible region's sensitivity to carotenoid content remained robust even when integrating data from contrasting vegetation types, indicating its broad applicability across diverse biomes. In the savanna ecosystem, the visible region exhibited the third-highest performance. While not as strong as in the other datasets, the visible region's performance was still considerable, highlighting its potential for use in savanna environments. It is important to note that while the savanna dataset exhibited considerable accuracy in estimating carotenoid content, it was not as strong as the performance of the tropical forest dataset, and combined, respectively. The savanna dataset produced the highest RMSE among the three datasets (Figure 3.3). This can be attributed to the landscape's complex structure.

The lower performance in Savanna leaves is likely due to specific traits (e.g., higher wax content, different specific leaf, or sun-stress adaptations) that PROSPECT-5 fails to simulate perfectly (Madonsela, 2018). In contrast, the Dukuduku tropical forest is composed of a continuous, evergreen leaves, which likely provides a more consistent and predictable spectral response for the PROSPECT-5 model to accurately simulate carotenoid content.

The findings reveal a consistent pattern of poor performance by the blue and red spectral regions across the different datasets and models examined. For the combined dataset, the blue region exhibited the lowest overall accuracy ($R^2 = 0.20$) and the highest RMSE (1.87), closely followed by the red-edge region ($R^2 = 0.33$, RMSE = 2.51). A similar trend was observed in the savanna ecosystem, where the blue region again showed the lowest performance ($R^2 = 0.20$, RMSE = 1.66), followed by the red region ($R^2 = 0.30$, RMSE = 2.58). The poor performance of the blue and red spectral regions could be attributed to the vegetation characteristics of the studied ecosystems. In Dukuduku tropical forest, the evergreen nature of the vegetation likely results in high chlorophyll content, which would lead to strong absorption of blue and red wavelengths. Similarly, in the savanna ecosystem, if the data was collected during the summer season, the vegetation may have been at its peak greenness, with a high chlorophyll content. Under such conditions, the blue and red wavelengths would be heavily absorbed, limiting their ability to capture the nuances of carotenoid variations. In contrast, the visible region, which encompasses the green wavelengths, may have been better suited to detect the distinct optical signatures created by varying carotenoid concentrations, even in the presence of high chlorophyll levels. The influence of seasonal and ecosystem-specific characteristics on the performance of different spectral regions highlights the importance of understanding the underlying biophysical and biochemical processes governing vegetation reflectance, which can help optimize the selection of spectral bands and the development of more robust remote

sensing-based approaches for monitoring vegetation properties such as carotenoid content across diverse biomes.

Based on this study's findings, we recommend using machine learning algorithms such as RFR and visible sub-spectral domain to select characteristic wavelengths, instead of using the full spectral domain when performing the inversion for carotenoids. The PROSPECT model can accurately estimate carotenoid concentrations by focusing on the band selection method, which aligns with the carotenoid's absorption bands. The visible range is particularly relevant for carotenoids, as they exhibit strong absorption peaks within this region. Therefore, restricting the analysis to the bands selected by RFR or visible spectrum enhances the model's sensitivity to carotenoid content and improves the inversion results.

The PROSPECT-5 model has demonstrated promising results in simulating carotenoid content at the leaf level. However, to further investigate this approach, it is necessary to evaluate its performance at the canopy scale. At the canopy level, additional parameters need to be considered, such as leaf area, soil properties, leaf angle, and various angular effects (observer zenith, relative azimuth, solar zenith, and hotspot). To account for these canopy-level factors, the PROSPECT model can be combined with a canopy model, such as SAIL, to create the PROSAIL framework. This integrated approach can provide a more comprehensive understanding of the interactions between leaf-level properties and canopy-level factors, and their collective influence on the remote sensing signal.

Future studies should explore options of model inversion such as the mean of several solution (5, 10, 20, 50). Literature shows that the selection of single best fitting spectra may not be the optimal strategy of inverting the LUT as this is prone to ill-poseness. Also, future studies should test the proposed wavelength selection approach within the PROSAIL framework to evaluate its performance in estimating carotenoid content at the canopy scale, where the impacts of noise and other confounding factors can be more realistically simulated. Additionally, existing carotenoid-related indices and their first-order derivatives should be tested on the current dataset across different ecosystems. This cross-validation and comparative analysis can help identify the most robust and versatile methods for carotenoid estimation in diverse biomes.

3.5. Conclusion

In this study, the wavelength selection approach was applied to estimate carotenoids using the PROSPECT-5. The following conclusions can be drawn:

- Among the tested algorithms for feature selection, RFR surpassed all other models with an R^2 of 0.82 and RMSE of 14.36.
- The analysis of three distinct datasets, encompassing savanna, tropical forest, and a combination of the two (both savanna and tropical forest) revealed that the green region, the band selected by the RFR, and visible region outperformed all other models.

Overall, this study is the first to use the band selection method with the PROSPECT model to simulate carotenoids. The results are valuable in determining the methods or bands valuable in estimating carotenoids.

3.6. Chapter summary

This chapter examined how different wavelengths within the field spectral data can be used to estimate carotenoid content using the PROSPECT-5 model across three different datasets: savanna, tropical forest, and a combined dataset that includes both biomes. The results showed that the green spectral region (510-580 nm) provided the best performance across all three datasets. Wavelengths selected through a RFR (i.e., green, yellow, and red-edge bands) achieved the highest accuracy for the savanna, followed by the combined dataset, and then the tropical forest. The visible region (400–700 nm) also performed well, especially within tropical forests, combined, and savanna datasets, showing promising results as well. These findings highlight the importance of choosing specific spectral regions for RTM inversion. Focusing on targeted wavelengths, rather than using the entire spectrum, can improve both the accuracy and efficiency of carotenoid estimation. Building on this, the next chapter also at the leaf level restricted the wavelengths in the form of vegetation indices. This approach aimed to increase the sensitivity of flavonoid content, especially regarding the influence of chlorophyll between 410 and 430 nm, on landscapes rich in phenolic compounds, such as the Fynbos biome of the GCFR.

Chapter Four:

Tracking phenolic Changes in Fynbos Biome: A spectroscopic Approach to Post-Fire Recovery

This chapter is based on:

Sibiya, B., Cho, M., Mutanga, O., Odindi, J., Masemola, C., Van Staden, J. and Vambe, M., 2025. Leaf spectroscopy reveals post-fire phenolic chrono sequence in the Fynbos Biome. *International Journal of Remote Sensing*, Under Review, Manuscript ID: TRES-PAP-2025-1611 (252838703)

Abstract:

Fire and secondary metabolites are vital for the health and function of ecosystems. In the fynbos biome, decomposition rate is very low, that dead plant accumulates over time until fire consume all dead material, allowing for new vegetation to sprouts. However, new vegetation faces a major challenge due to poor soil conditions particularly limited nitrogen and phosphorus. Hence, plant survive these stressful conditions by producing secondary metabolites. Together, fire and secondary metabolites play a crucial role in the survival and regeneration of Fynbos biome. Previous studies often study these two elements as separate entities, therefore, there is still a big gap in understanding how post-fire affects the concentration secondary metabolites using remote sensing. In this research, using leaf spectroscopy, we tested this hypothesis in the fynbos biome; a fire-driven and secondary metabolites rich landscape. The results showed that both fire and secondary metabolites are important for ecosystem rejuvenation. Areas that experienced recent fires in 2015, 2016, and 2021 had high levels of phenolic compounds. In contrast, regions with limited fire (2010 and never experience fire ("never_exp")) had low concentrations of phenolics. Using the Continuum removal technique, we identified that the strongest spectral signals for total phenols occurred at 1660 nm, and 415 nm for flavonoid. Since chlorophyll also absorbs light around this region (410–430 nm), we tested and adjusted various spectral indices to improve the sensitivity for flavonoids. Notably, $FI_{415,710}$ and $FI_{420,710}$ showed the highest correlations with phenolic content, with an R^2 of 0.55 and an RRMSE of around 1.98-1.99%. Overall, this study provides an understanding of how fire and secondary metabolites work together to sustain ecosystem health, highlighting their combined role in ecosystem resilience and regeneration.

Keywords: Phenolic compounds, total phenol, flavonoid, leaf spectroscopy, fynbos biome.

4.1. Introduction

Globally, fire and plant secondary metabolites (PSMs) play a crucial role in maintaining ecosystem health and functioning. Fire is an essential ecological process that plays a huge role in numerous ecosystems such as fynbos, chaparral, sclerophyll communities, grassland, and savanna (Kraaij and van Wilgen, 2014; Keeley et al., 1981; Collins, 1990; Belsky, 1992; Morgan, 1999). It helps break down dead plants by removing old vegetation, which then allows new, nutrient-rich plants to grow (Smit et al., 2024). In environments where natural decomposition is slow, causing litter to accumulate over time, fire breaks down the build-up and creates favourable conditions for seed germination and nutrient cycling (Bengtsson et al., 2011). This process enhances forage quality for herbivores and is vital for the regeneration of ecological communities (Neary and Leonard, 2020).

Similarly, PSMs serve as defence compounds that help plants survive various biotic and abiotic stressors, including herbivory, pathogens, heavy metals, drought, UV radiation, and air pollution (Wand, 1995; van der Merwe and Marais, 1994; Porter, 1992; Kumar et al., 2020; Campbell et al., 2004; Nandave et al., 2009). Plants respond to stress by producing specific secondary metabolites, including alkaloids, phenolic compounds, and terpenes (Chomel et al., 2016). For example, research indicates that the production of phenolic compounds increases when plants face stress, which enhances their resilience (Chomel et al., 2016; Xie et al., 2024).

In the fynbos biome, both fire and PSMs are crucial for the survival and regeneration of plant communities. For instance, the slow decomposition rates characteristic of its nutrient-poor soil result in litter accumulation in the ecosystem for an extended period until it is consumed by periodic fires, which play a vital role in nutrient cycling and community structure (Stock and Allsopp, 1992; Mitchell et al., 1986). Fire not only remove old vegetation but also provides opportunities for seed germination, enabling new growth (Bond et al., 2005; Kraaij and van Wilgen, 2014; Smit et al., 2024).

However, this new growth faces significant challenges due to nutrient-poor, highly leached, and acidic sandy soils (Magadlela et al., 2015; Bengtsson et al., 2011). In the fynbos, plants cope with stress factors by producing secondary metabolites. The literature indicates that this ecosystem is particularly rich in phenolic compounds, which serve as chemical defence mechanisms (Wand, 1995; an der Merwe and Marais, 1994). Thus, while fire acts as a physical driver of ecosystem change by consuming old vegetation, PSMs provide a biochemical support system for newly emerging vegetation following fire events. Together, these processes

contribute to the overall adaptability and survival of plant communities in fire-prone ecosystems, highlighting the complex interplay between physical and chemical factors in ecological resilience. While a substantial body of research has explored the significance of fire and PSMs in the functionality of various ecosystems, these elements are often studied as separately entities (Skidmore et al., 2010; Zhou et al., 2022; Xie et al., 2024; 2025; Kraaij and van Wilgen, 2014; Kozlowski, 2012). Therefore, there is still a big gap in understanding how post-fire affects the levels and variety of secondary compounds in plant communities. Understanding PSMs post-fire is essential for grasping their role in ecosystem recovery, resilience, and overall health, as well as for informing conservation and management strategies in fire-prone environments. This was tested in the fynbos biome found in the GCFR, a fire-driven ecosystem and very rich in phenolic compounds, using remotely sensed data (Bengtsson et al., 2011; Kumar et al., 2020; Nandave et al., 2009).

Remote sensing is one of the most cost-effective and efficient methods for studying the post-fire phenolic compound sequence over time. Its advantages include being non-destructive, covering large areas, and offering high spectral and temporal resolution, allowing researchers to examine changes over different periods (Skidmore et al., 2010; Tshabalala et al., 2021; Zhou et al., 2023; De Klerk, 2008). This capability enables continuous monitoring and analysis of vegetation recovery and chemical responses over time, including revisiting previous conditions to better understand the dynamics of ecosystem resilience.

Furthermore, phenolic compounds possess distinct spectral features that can be detected by advanced sensors like the field spectral data (Zhou et al., 2022; Sibiya et al., 2024b). These sensors have highly detailed spatial and spectral resolution due to their narrow bandwidths, enabling them to capture precise spectral information from plant canopies (Darvishzadeh et al., 2008). The primary advantage of hyperspectral sensors is their ability to collect data that closely resembles laboratory measurements, making them highly effective for detailed analysis of post-fire phenolic chrono sequence (Paramanik et al., 2025). In addition, phenolic compounds absorb specific wavelengths within the 400 to 2500 nm spectral range, making it possible to non-destructively using remote sensing. For instance, total phenols absorb light predominantly between 1620 and 1690 nm (Xie et al., 2024), while flavonoids strongly absorb shortwave blue light in the 410–430 nm range (Gitelson et al., 2017).

However, in this region (410 – 430 nm) chlorophyll also absorbs this light spectrum (Gitelson et al., 2017). To better detect and quantify flavonoids, it is important to minimize the influence

of chlorophylls to enhance their sensitivity. Consequently, various studies have proposed the formulation of vegetation indices that directly relate to flavonoid absorption characteristics (Gitelson et al., 2017; Zhou et al., 2022, 2023). In this study, we tested and modified these spectral indices in the fynbos ecosystem to evaluate their effectiveness in this unique, phenolic fire-adapted ecosystem. This study aims to evaluate and modify these spectral indices within the context of the fynbos ecosystem, a unique phenolic fire-adapted environment. The main aim of this research is to utilize leaf spectroscopy to investigate the post-fire phenolic chrono sequence in the Fynbos Biome. The specific objectives of this study are as follows: (i) to assess the variability of phenolic compounds during post-fire recovery over the past 14 years (2010–2024) in the Fynbos biome; and (ii) to evaluate the efficacy of flavonoid indices to study post-fire recovery.

4.2. Materials and methods

4.2.1. Field sampling

The field data collection was conducted from March 3 to March 8, 2024. The experimental sampling design was based on areas that previously experienced fire events between 2010 and 2024. To locate burned sites, satellite imagery from Landsat-7 (2010 - 2015) and Sentinel-2 (2015 - 2024) was employed from Google Earth Engine (GEE). Fire scars were mapped based on visual interpretation of Landsat-7 and Sentinel-2 image false colour composite (Sato et al., 2016) (Figure 4.1). This approach was deemed sufficient, hence, burn indices were not employed (Filipponi, 2018). Therefore, a total of 190 samples were collected from areas that experienced fire, as well as from areas that did not experience fire during this period, designated as "never-exp." The collection dates and the corresponding number of points collected are as follows: 33 points in March 2015, 60 points in February 2016, 12 points in April 2016, 16 points in May 2021, and 37 points in January 2010. Additionally, 32 points were collected from the "never-exp" areas.

To ensure accurate sampling, a high resolution submeter (50 cm) accurate handheld Trimble Global Navigation Satellite System (GNSS) was employed to mark the sampling points, with each point situated at least 200 meters from the road to mitigate road edge effects (Delgado et al., 2007; Lottering and Mutanga, 2012). Upon arrival at each plot, two 10-meter tapes were laid perpendicular to one another to form a quadrant as shown in Figure 3. Within each querent, leaf Area Index (LAI) was measured using a LiCOR 2200C plant canopy analyser (LI-COR GmbH, Germany) and Soil moisture content was measured using HH2 moisture probe (Delta-

T soil moisture sensors, United states) recorded as a percent at each sample point. Additionally, leaf samples were randomly harvested from various canopy levels, top, middle, and lower branches, to capture intra-canopy variability from all four quadrants. In total 17922.2g of leaves were harvested from dominant species and were placed in zipped plastic bags and then preserved in a cooler box with ice. The leaf samples were then transported back to a lab for spectral measurements. The spectral measurements were taken on fresh leaves harvested. Thereafter, the sample leaves were dried in the oven at 80 ◦C for about 48 h until constant weight.

4.2.2. Leaf reflectance measurements

A FieldSpec 3(R) spectrometer was used to measure the leaves' spectra. This device can detect light with wavelengths between 350 and 2500 nm, with the ability to record measurements at every 1 nm interval (ASD Analytical Spectral Devices, 2016). This comprehensive spectral range allowed for the acquisition of detailed information about the leaf's optical properties. The leaf samples were stored in a dark room to ensure a controlled and consistent measurement environment. A halogen lamp was used as the light source, and a dark board used as the background to eliminate any extraneous light interference. This setup allowed the researchers to capture the true reflective properties of the leaves without external light contamination. Before analysing the leaf samples, the leaves were placed on the dark board, and a white reference panel used for calibration. Then, a special probe used to measure the leaf's reflectivity. Each leaf sample was measured five times to obtain a reliable dataset and then averaged (Cho et al., 2014). After measurement, brown labelled bags were used to place samples before being dried in an oven at 80 °C. This post-measurement treatment was necessary to preserve the samples for further analysis.

4.2.3. Biochemical analysis

The preserved samples were transferred to labelled 30 ml vials following the grinding process using a coffee grinder. Subsequently, 0.5 g of the leaf powder was added to a centrifuge tube, along with 5 mL of a solution consisting of 50% water and 50% methanol (MeOH) prepared in a reagent bottle. The samples were then sonicated in cold water for 20 min using an ultrasonic bath to extract compounds. Thereafter, the samples were centrifuged for another 20 min to separate the solid leaf powder from the liquid extract. The liquid extract was then carefully removed from the centrifuge tube and stored in a new labelled centrifuge tube. This extracted liquid was used for analysing total phenols and flavonoids as outlined by Vambe et al. (2024). The total phenols and flavonoids were assessed using the Folin-Ciocalteu and aluminum

chloride assays (Markar, 1999). Table 4.1 outlines the specific assay procedures followed for quantifying the total phenol and total flavonoid in the leaf extracts.

Table 4.1: Chemical Assay Procedures for Quantifying Total Phenol and Flavonoid.

Total phenol	**Flavonoids**
• In a test tube placed in a rack, mix 50 µL of the sample, 950 µL of H_2O, 500 µL of Folin-C, and 2.5 mL of sodium carbonate using pipette.	• In a test tube placed in a rack, mix the following using a pipette: 250 µL of the sample, 1 mL of H_2O, 75 µL of sodium nitrite ($NaNO_2$), 75 µL of Aluminum chloride ($AlCl_3$), 0.5 mL of sodium hydroxide (NaOH).
• Garlic acid (0.1mg/ml) was used as a standard.	• Catechin (0.1 mg/mL) was used as a standard
• Incubate the reaction mixtures in the dark for 30 minutes.	
• Absorbance was then read at 725 nm using a spectrophotometer.	• Absorbance was read at 725 nm using a spectrophotometer.
• The assay was done in triplicate and results expressed as mean ± standard error.	• The assay was done in triplicate and results expressed as mean ± standard error.

4.2.4. Spectral data processing

Prior to conducting spectral pre-processing on the leaf reflectance data, it was crucial to tackle the inadequate signal observed within certain spectral bands, specifically, the spectral range between 350 - 399 nm and 2200 – 2500 nm were removed. Following the removal of the low-quality spectral regions to enhance the smoothness, the Savitzky-Golay method with a second-degree polynomial using a 15 moving window was used, (Savitzky and Golay, 1964; Xie et al., 2024). Thereafter, we used the Continuum removal method to pre-process the spectra (Clark and Roush, 1984). The continuum was calculated using the ConvexHull approach, which involves fitting a convex polygon to the upper envelope of the smoothed spectrum. The application of Savitzky-Golay filtering and continuum removal techniques is the well-known pre-processing workflow to enhance the quality and interpretability of the spectral data by reducing noise, preserving spectral features, and isolating distinct absorption traits associated with the biochemical and structural attributes of the leaves (Mutanga et al., 2005; Kokaly and Clark, 1999). The application of these pre-processing steps is crucial for subsequent feature extraction, spectral modelling, and the development of robust predictive algorithms for various leaf and plant-related applications.

4.2.5. Calculation of vegetation indices

Flavonoids absorb shortwave blue light in the 410 to 430 nm range but do not absorb red or red-edge light (Gitelson et al., 2017; Agati et al., 2015). Conversely, chlorophyll absorbs blue, red, and red-edge light, with the red-edge spectrum shown to be particularly sensitive to chlorophyll content (Gitelson et al., 2017). To improve the detection of flavonoids and minimize the influence of other pigments, especially chlorophyll, Gitelson et al. (2017) and Zhou et al. (2022) developed flavonoids indices. In this study, we tested and modified these formulas from Zhou et al. (2022) to enhance their effectiveness. This was achieved by replacing 420 nm wavelength with 415 nm as shown in Table 4.2.

Table 4.2: Utilized Vegetation Indices.

Group	Index	Formulation	Reference
$FI_{\lambda1,\lambda2}$	$FI_{420,710}$	$R\frac{-1}{420}$ - $R\frac{-1}{710}$	Zhou et al. (2022) modified from Gitelson et al. (2017)
	$FI_{415,710}$	$R\frac{-1}{415}$ - $R\frac{-1}{710}$	Modified from Zhou et al. (2022)
$FI_{\lambda1,\lambda2,800}$	$FI_{420,710,800}$	$(R\frac{-1}{420}$ - $R\frac{-1}{710})$ x 800	Zhou et al. (2022) modified from Gitelson et al. (2017)
	$FI_{415,710,800}$	$(R\frac{-1}{415}$ - $R\frac{-1}{710})$ x 800	Modified from Zhou et al. (2022)
	FRI	$(R\frac{-1}{410}$ - $R\frac{-1}{460})$ x 800	Merzlyak et al. (2005)
$mFI_{\lambda1,\lambda2}$	$mFI_{420,710}$	$(R_{420} - R_{445})^{-1} - (R_{710} - R_{445})^{-1}$	Zhou et al. (2022) modified from Gitelson et al. (2017)
	$mFI_{415,710}$	$(R_{415} - R_{445})^{-1} - (R_{710} - R_{445})^{-1}$	Modified from Zhou et al. (2022)
	$mFI_{420,690}$	$(R_{420} - R_{445})^{-1} - (R_{690} - R_{445})^{-1}$	Zhou et al. (2022)
	$mFI_{415,690}$	$(R_{415} - R_{445})^{-1} - (R_{690} - R_{445})^{-1}$	Modified from Zhou et al. (2022)
$mFI_{\lambda1,\lambda2,800}$	$mFI_{420,710,800}$	$[(R_{420} - R_{445})^{-1} - (R_{710} - R_{445})^{-1}]$ x $(R_{800} - R_{445})^{-1}$	Zhou et al. (2022) modified from Gitelson et al. (2017)
	$mFI_{415,710,800}$	$[(R_{415} - R_{445})^{-1} - (R_{710} - R_{445})^{-1}]$ x $(R_{800} - R_{445})^{-1}$	Modified from Zhou et al. (2022)
	$mFI_{420,690,800}$	$[(R_{420} - R_{445})^{-1} - (R_{690} - R_{445})^{-1}]$ x $(R_{690} - R_{445})^{-1}$	Zhou et al. (2022)
	$mFI_{415,690,800}$	$[(R_{415} - R_{445})^{-1} - (R_{690} - R_{445})^{-1}]$ x $(R_{690} - R_{445})^{-1}$	Modified from Zhou et al. (2022)
$DDblue_{\lambda1,\lambda2}$	$DDblue_{421,733}$	$(R_{445} - R_{421}) / (R_{455} - R_{433})$	Zhou et al. (2022)
$DDnir_{\lambda1,\lambda2}$	$DDnir_{434,421}$	$(R_{800} - R_{434}) / (R_{800} - R_{421})$	Zhou et al. (2022)

4.2.6. Model validation and accuracy assessment

A simple linear regression model was used to find the relationship between vegetation indices and flavonoid content. It was also used to observe the relationship between phenolic compounds and LAI and phenolic compounds with soil moisture content. The accuracy of the flavonoid-vegetation indices was assessed using the Normalized Root Mean Square Error (NRMSE, %) and the coefficient of determination (R^2), based on a dataset of 190 samples.

4.3. Results

4.3.1. Changes in phenolic content post-fire occurrence

To locate areas affected by fire between 2010 and 2024, we used satellite data from Sentinel-2 (2015 - 2021) and Landsat (2010). The analysis revealed fire events in January 2010, March 2015, February 2016, April 2016, and May 2021, with the burnt areas visible in Figure 4.1. Furthermore, we identified one area, labelled “never_exp”, that remained unburnt throughout the study period (Figure 4.1b). In areas that experienced fire, the most fire events occurred in January 2010 and in February and April of 2016, covering large areas. These months coincide with the dry period in the GCFR, which typically extends from late spring to early autumn. During this time, the region experiences its lowest rainfall, creating conditions conducive to fire outbreaks. This is evidenced by the high frequency of fire events recorded during these months.

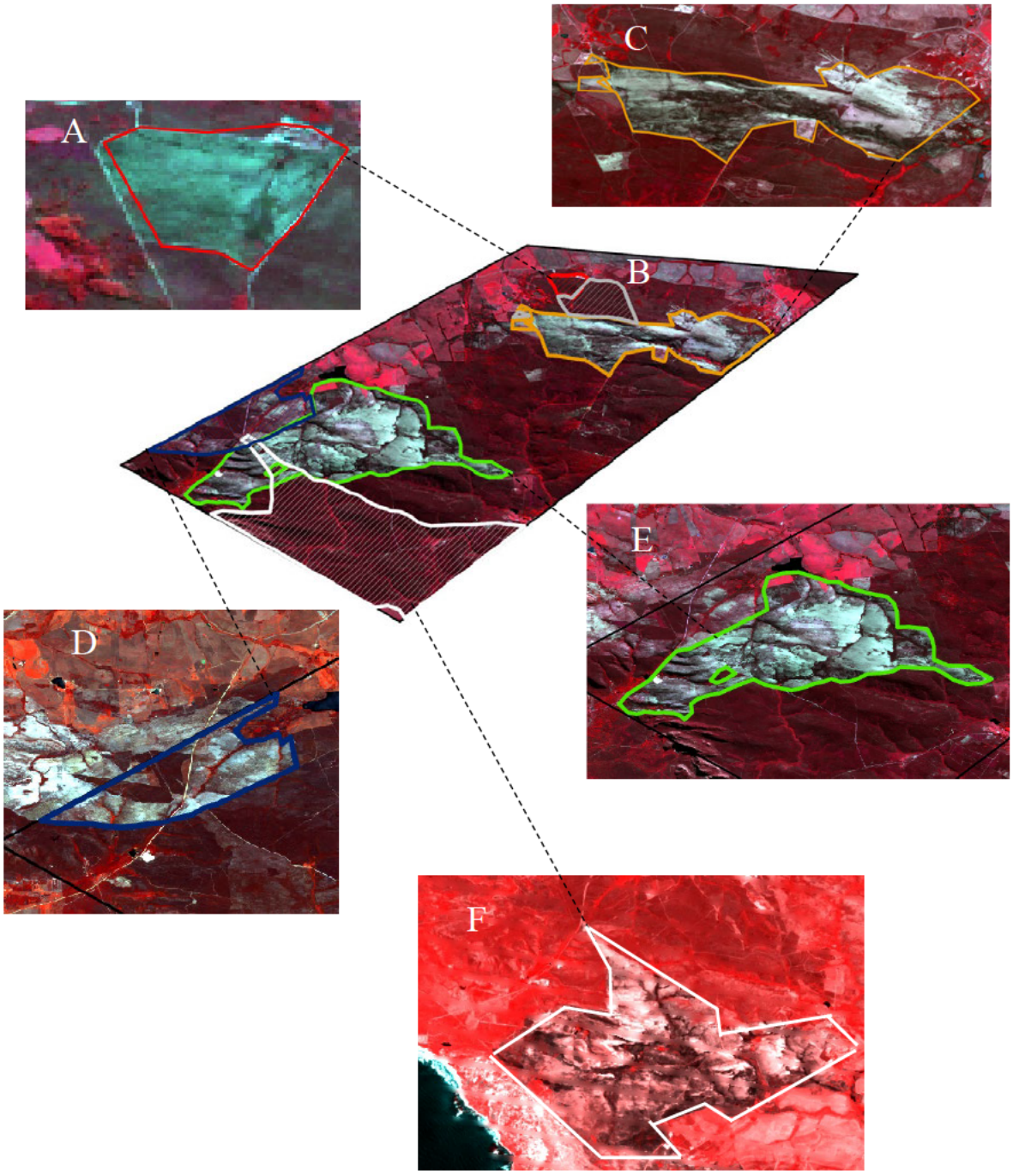


Figure 4.1. Location of burnt areas between 2010 and 2024 using false colour composite: A = 2021 (May), B = Never_exp, C = 2016 (February), D = 2015 (March), E = 2016 (April), F = 2010 (January).

Figure 4.2 shows the vertical distribution of total phenol and flavonoid concentrations after post-fire events. The data display similar trends over the years following the fires. The highest concentrations were recorded in May 2021, followed by April 2016. In contrast, the lowest

concentrations were observed in the area that never experienced fire within this period ("never_exp").

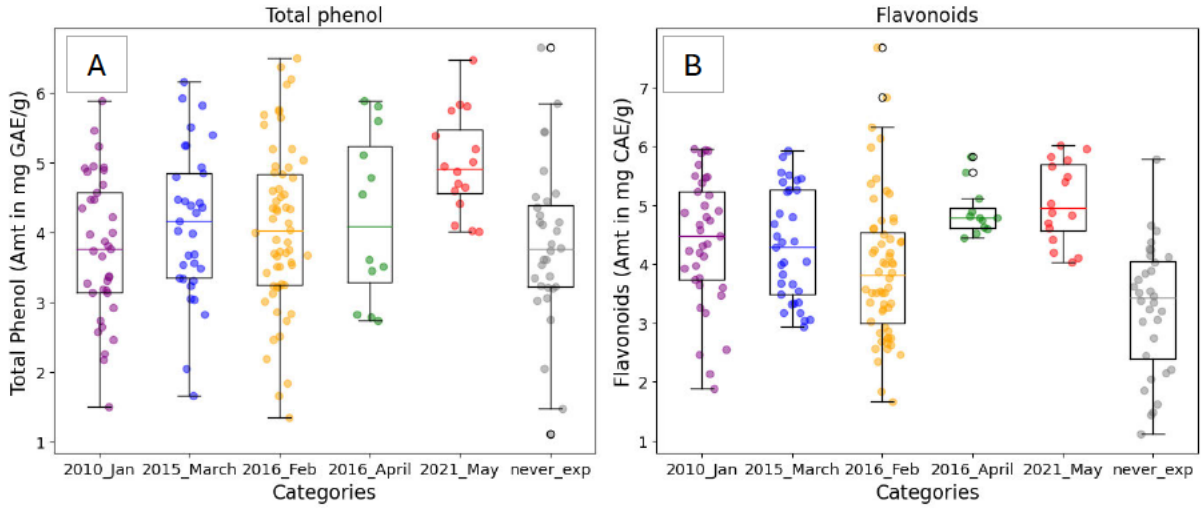


Figure 4.2. Variation in (A) total phenols and (B) flavonoids in five areas that had fires and areas that have not experienced fires from 2010 to 2024.

4.3.2. Evaluation of leaf spectra

4.3.2.1. Comparison of the spectral characteristics

We observed notable differences in mean reflectance and spectral variation (%) between areas previously affected by fire and those unaffected (Figure 4.3). The results show that the highest mean reflectance was recorded in January 2010 and February 2016. Conversely, March 2015 exhibited the lowest mean reflectance. In terms of spectral variation (coefficient of variation %), the highest levels were observed in April 2016 and February 2016, while the lowest levels were found in January 2010 and in the "never_exp" areas.

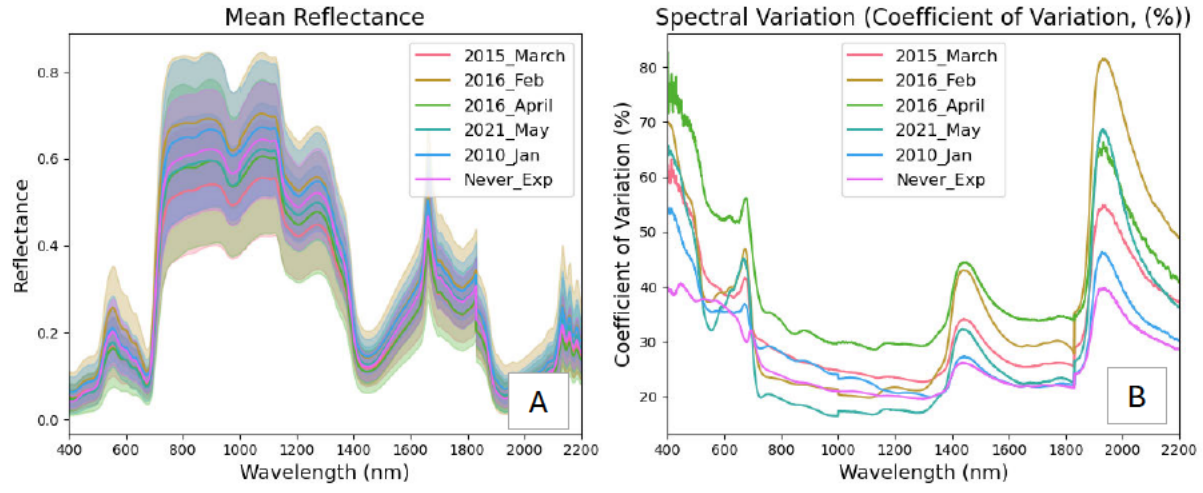
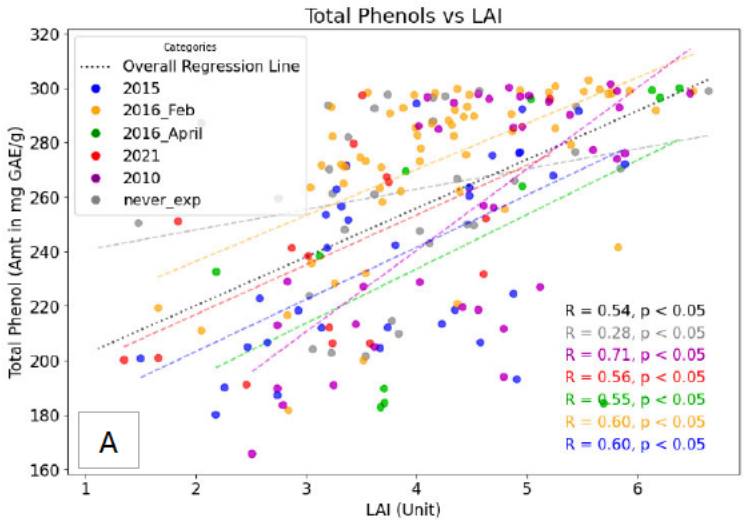


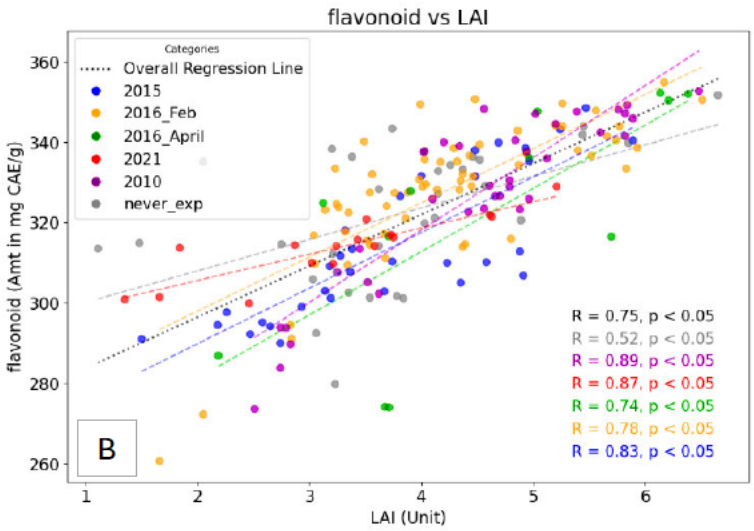
Figure 4.3. Comparing the (A) average reflectance and (B) spectral coefficient of variation of five areas that experienced fire with areas that had not experienced fire during this period.

4.3.2.2. *Influence of Leaf Area Index and Soil Moisture on Phenolic Compounds in Plants*

Figure 4.4 illustrates the relationships between phenolic compounds and LAI and phenolic compounds with soil moisture content. The data reveal a linear relationship between both phenolic compounds and LAI, indicating that as LAI increases, the concentration of phenolic compounds also tends to rise. Specifically, flavonoids produced higher correlation coefficients of r = 0.75 than phenolics with r = 0.5. For both phenolic compounds, areas that experienced fire in 2010 showed the highest correlation coefficient, r = 0.71 for total phenols and r = 0.89 for flavonoids. Conversely, regions that remained unburned from 2010 to 2024 exhibited much lower correlations of r = 0.28 and r = 0.52, respectively.

Regarding soil moisture, due to a limited number of samples, the study did not classify samples based on post-fire occurrence. Nonetheless, the observed trend demonstrates an inverse correlation between soil moisture levels and phenolic compound concentrations, indicating that as soil moisture decreases, phenolic content increase. This pattern aligns with plant responses to water stress, wherein phenolic compounds often serve as protective agents against oxidative stress during drought conditions.





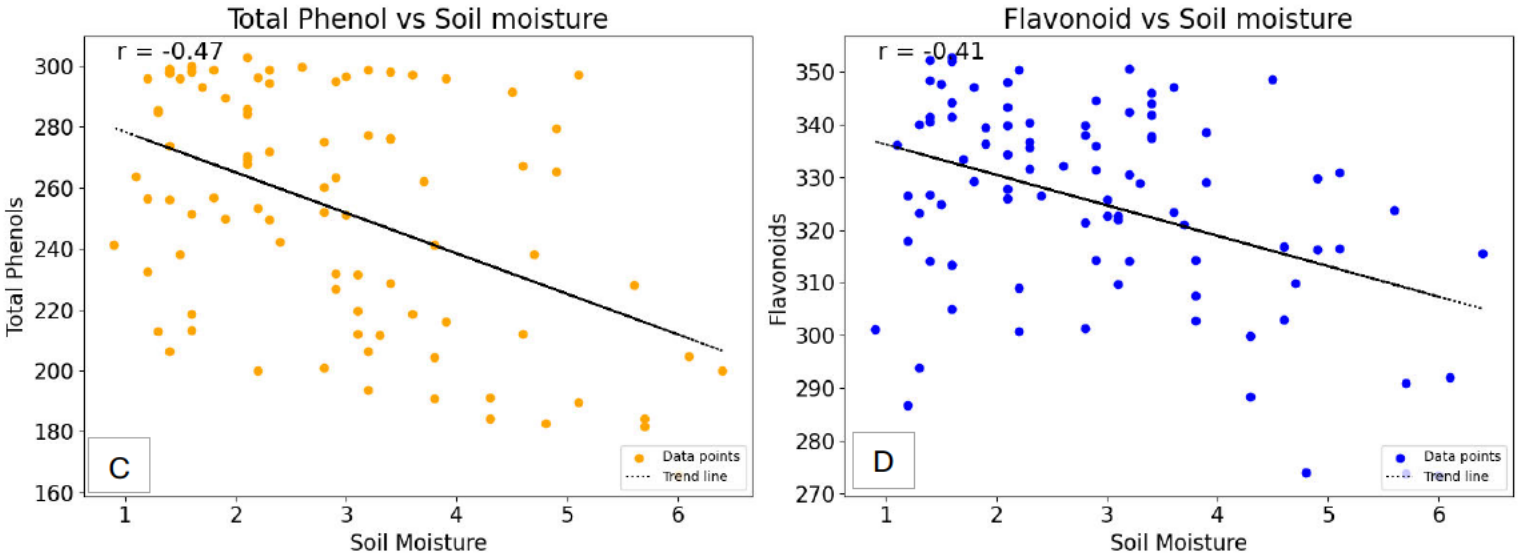


Figure 4.4. The relationships of LAI and phenolic compounds (A-B) and relationships of soil moisture and phenolic compounds (C-D).

4.3.3. Phenolic absorption features

By analysing the average continuum-removed spectra, we determined that the lowest spectral pit occurs at 1660 nm for total phenols and 415 nm for flavonoids (Figure 4.5). The depth of these spectral features demonstrated a positive correlation to the amounts of total phenol and flavonoid in our dataset, yielding correlation coefficients of $r = 0.59$ ($p < 0.05$) for total phenol and $r = 0.47$ ($p < 0.05$) for flavonoid (Figure 4.6). Notably, the correlation was stronger for total phenols than for flavonoids. Further examination of the spectra revealed a prominent peak in the 1620-1690 nm region, which is critical for identifying phenolic compounds, specifically total phenols, as noted by Xie et al. (2024). Additionally, existing literature suggests that flavonoids possess characteristic absorption bands in the 410-430 nm range (Gitelson et al., 2017). However, in this region (410 – 430 nm), chlorophyll absorbs light. Therefore, to minimize the effect of Chlorophyll, we tested and modified the existing flavonoid indices.

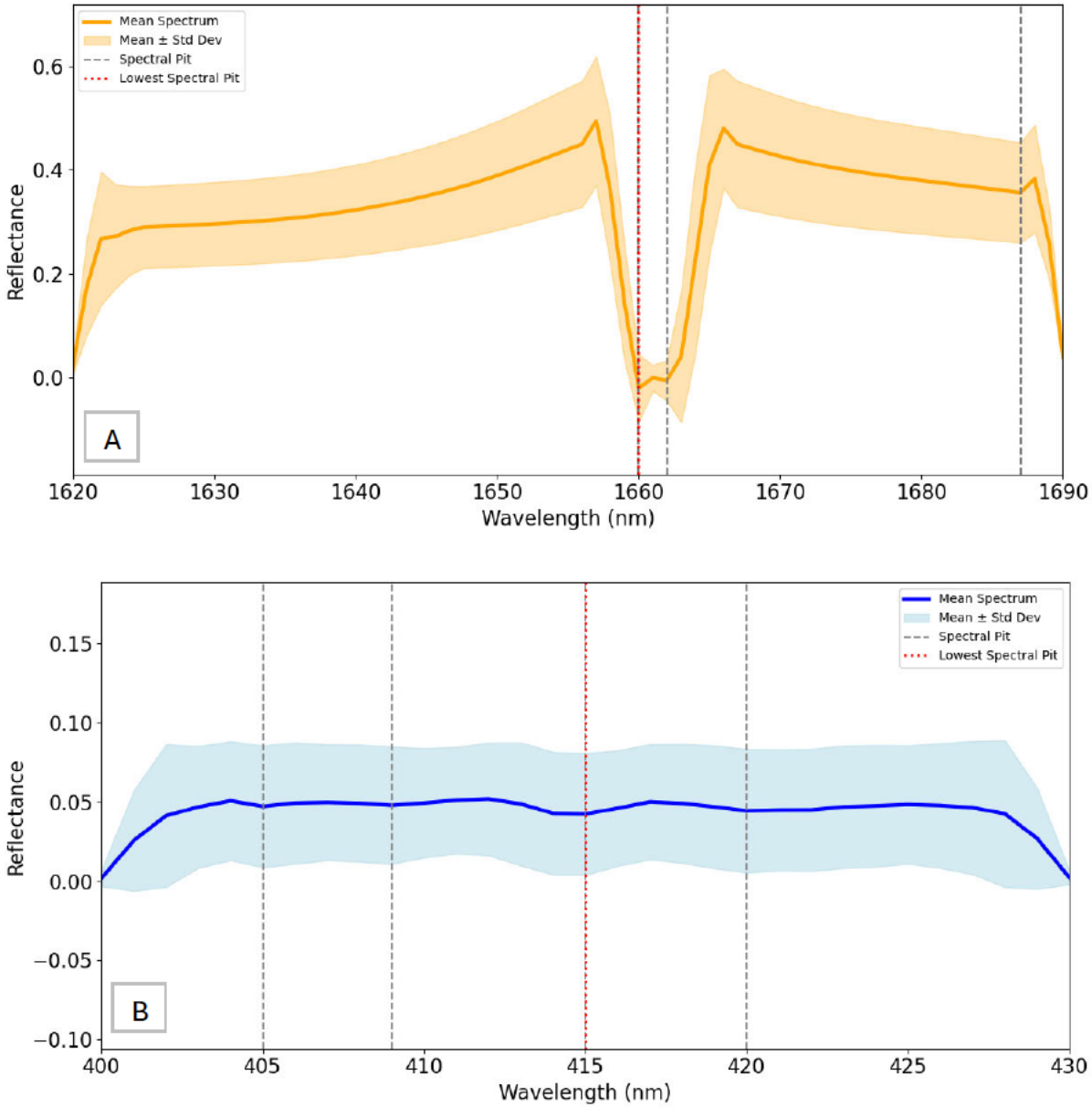


Figure 4.5. Continuum-removed spectral features for the hotspot region: (A) total phenol in the range of 1650–1670 nm, and (B) flavonoid in the range of 400–430 nm. A red dotted line highlights the lowest absorption.

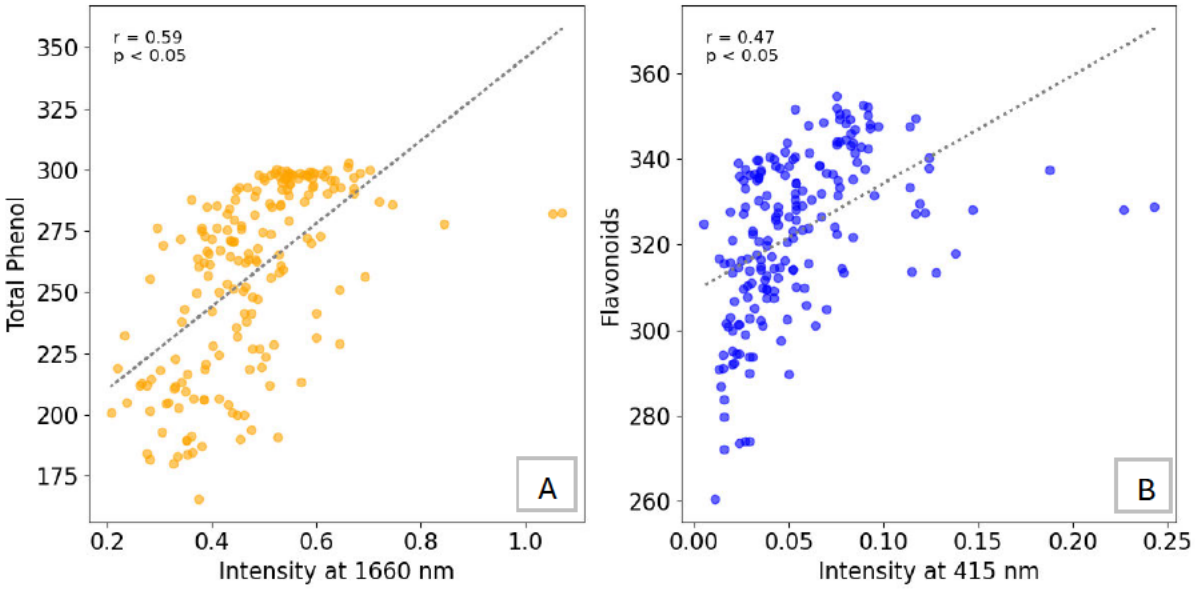


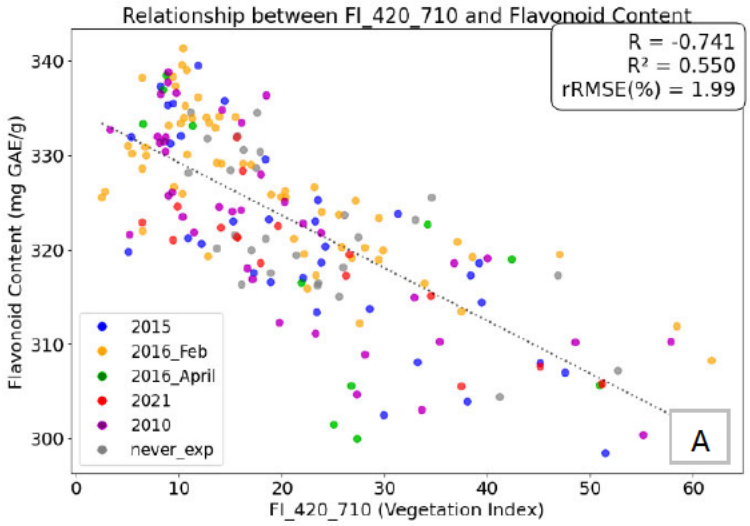
Figure 4.6. The correlation between total phenolics concentration and the spectral feature at 1660 nm (A) and between flavonoids concentration and the spectral feature at 415 nm (B).

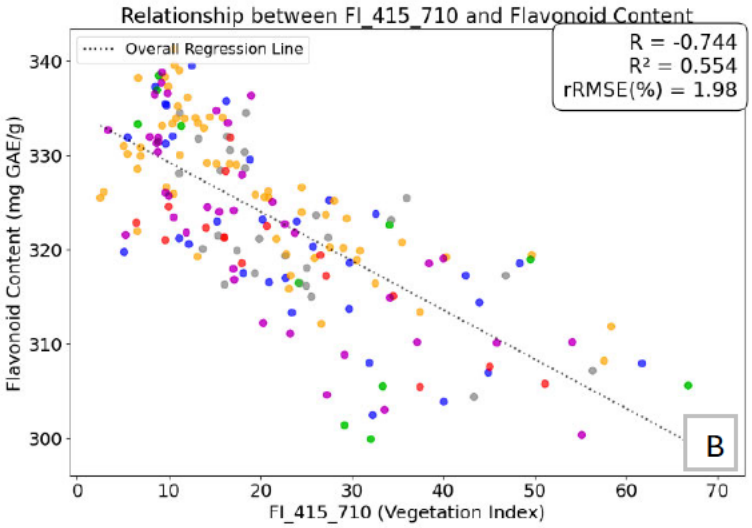
4.3.4. Relationships between flavonoid and vegetation Indices

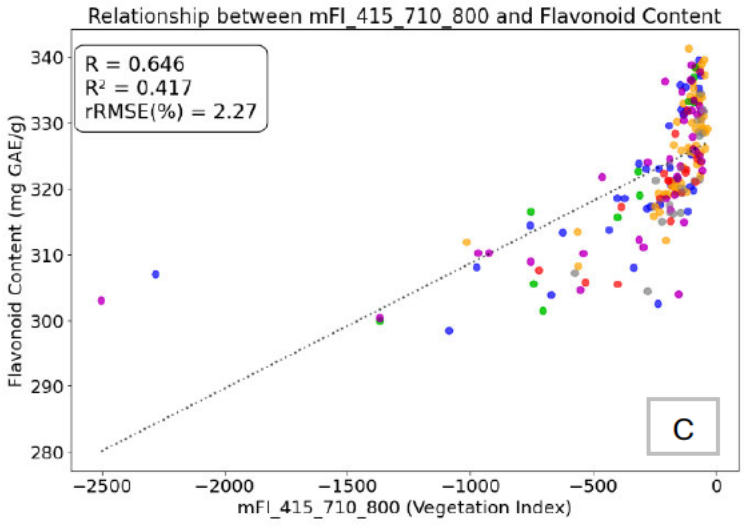
Table 4.3 shows the relationship between flavonoids and spectral indices. We evaluated existing spectral indices from Gitelson et al. (2017) and Zhou et al. (2022) and enhanced these indices by incorporating a spectral feature at 415 nm. Generally spectral indices showed weak correlations with flavonoid levels, with R^2 values below 0.40. However, several indices, including FI415,710, FI420,710, mFI415,710, 800, and mFI415,690, 800, exhibited improved correlations (Figure 4.7). Notably, FI415,710 and FI420,710 achieved the highest correlations, with an R^2 of 0.55 and an rRMSE of approximately 1.98-1.99%. Interestingly, the indices modified in this study, specifically those incorporating the 415 nm wavelength, outperformed the original indices from Gitelson et al. (2017) and Zhou et al. (2022), leading to more accurate flavonoid predictions. This highlights the potential of the 415 nm spectral region in estimating plant flavonoid content.

Table 4.3: Accuracy of various flavonoids spectral indices.

Group	Index	CV-R2	RMSE	rRMSE (%)
$FI_{\lambda1,\lambda2}$	$FI_{420,710}$	**0.55**	**6.94**	**1.99%**
	$FI_{415,710}$	**0.55**	**6.89**	**1.98%**
$FI_{\lambda1,\lambda2,800}$	$FI_{420,710,800}$	0.21	8.71	2.70%
	$FI_{415,710,800}$	0.22	8.66	2.68%
	FRI	0.27	8.35	2.59%
$mFI_{\lambda1,\lambda2}$	$mFI_{420,710}$	0.14	9.01	2.79%
	$mFI_{415,710}$	0.33	8.02	2.49%
	$mFI_{420,690}$	0.14	8.98	2.51%
	$mFI_{415,690}$	0.33	7.99	2.48%
$mFI_{\lambda1,\lambda2,800}$	$mFI_{420,710,800}$	0.15	8.96	2.78%
	$mFI_{415,710,800}$	**0.42**	**7.49**	**2.27%**
	$mFI_{420,690,800}$	0.16	8.915	2.76%
	$mFI_{415,690,800}$	**0.42**	**7.47**	**2.27%**
$DDblue_{\lambda1,\lambda2}$	$DDblue_{421,733}$	0.38	7.69	2.39%
$DDnir_{\lambda1,\lambda2}$	$DDnir_{434,421}$	0.33	8.03	2.49%







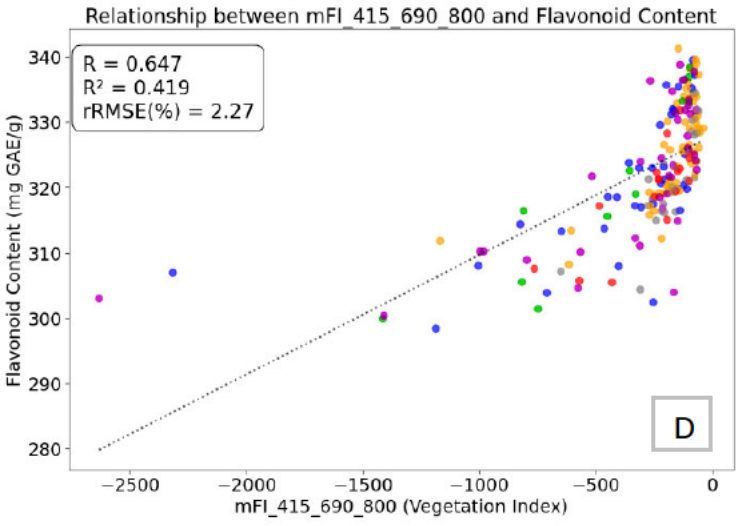


Figure 4.7. Top four best-performing flavonoid spectral indices.

4.4. Discussion

Spectral analyses conducted on 190 leaf samples revealed significant variability in the concentrations of total phenols and flavonoids based on post-fire events. Notably, distinct spectral features were observed at 1660 nm corresponding to total phenols and at 415 nm associated with flavonoids. By employing leaf spectroscopy, we effectively captured this variability.

4.4.1. Changes in the concentration of phenolic compounds

Higher concentrations of phenolic compounds in areas that experienced recent fire events, specifically in 2021 (May), 2016 (February and April), and 2015 (March) compared to older (2010), or areas that have not experience fire within this period denoted as “never_exp” and 2010). This could be due to newly sprouted plants work to establish themselves and defend against stress factors. Fires create a nutrient-rich environment through the decomposition of organic material and the release of nutrients previously locked in dead biomass (Stock and Allsopp, 1992; Mitchell et al., 1986). This availability can further support heightened growth rates and associated increases in secondary metabolite production. This is in line with Zhou et al. (2022), who found that younger Ginkgo trees are significantly higher in flavonoids than old Ginkgo trees.

In contrast, older vegetation may indicate plants that are experiencing nutrient stress, as fynbos is a fire-driven ecosystem that typically burns every 10 to 15 years (Kraaij and van Wilgen, 2014). The period from 2010 to 2024 is significant; vegetation that has not experienced fire during this time may have reached full maturity, leading to senescence. These plants could potentially experience fire at any point within this timeframe. As plants age, their ability to produce metabolic compounds may diminish, particularly in the fynbos, where they compete for limited resources over extended periods. According to Qaderi et al. (2023), new leaves typically accumulate metabolites more rapidly than older leaves.

Furthermore, observations from the current study indicate notable differences in mean and spectral variation (%) in spectroscopy reflectance in different years (Figure 4.3). In terms of the mean reflectance, 2016 (February), 2010 (January), and “never_exp” areas exhibited the highest mean reflectance (Figure 4.3a). This could be attributed to differences in vegetation density and health, with more recently burned areas typically showing lush, green growth that reflects more light (Smit et al., 2024; De Klerk, 2008). On the other hand, older or well-

established plants ("never_exp" and 2010) have thicker, more lignified leaves with a higher proportion of structural components like cellulose and lignin, which can increase the overall reflectance (Mediavilla et al., 2011; Chen et al., 2022). This is in line with Zhou et al. (2022), who indicated that the reflectance of mature *Ginkgo* tree species is greater than that of younger specimens when estimating flavonoid concentrations in the species.

On the other hand, for spectral variation (%), unlike mean reflectance, recently burnt areas such as 2016 (February and April) and 2015 show the highest spectral variation (Figure 4.3b). In contrast, "never_exp" areas and 2010 displayed the lowest spectral variation for phenolic compounds. This lower variation is likely due to the presence of older, more mature vegetation, which tends to have a more uniform structure and composition, resulting in less spectral variability (Mediavilla et al., 2011).

Furthermore, Figure 4.4 reveals a linear relationship between phenolic compounds and LAI. This finding aligns with the work of Zhou et al. (2023), who utilized the SAIL model with hyperspectral data to evaluate the flavonoid content in *Ginkgo* plantations. These studies suggest that phenolic compounds can be accurately estimated by measuring LAI. LAI measures the total surface area of leaves, which helps plants capture more light for photosynthesis. This increased light exposure encourages the production of flavonoids (Ni et al., 2020). In contrast, an inverse correlation between soil moisture and phenolic compounds. This finding aligns with the understanding that plants often produce secondary metabolites, such as phenolics, as a response to stressful conditions (Ashraf et al., 2018; Bartwal et al., 2013). When soil moisture is limited, plants may experience water stress. Hence, in response, they may elevate their production of phenolic compounds (Varela et al., 2016; Shen et al., 2022; Naczk et al., 2004; Reginato et al., 2014).

4.4.2. Absorption features of phenolics

The lowest spectral pit for total phenol occurs at 1660 nm, with a correlation (r) of 59% between total phenolics and 1660 nm (Figure 4.6a). This finding aligns with existing literature, which highlights the significance of the phenolic absorption feature around 1660 nm in relation to other absorption characteristics at 1460, 1930, 2140, and 2170 nm. These other features are closely associated with major plant biochemical constituents such as lignin, cellulose, starch, and protein (Xie et al., 2024; Kokaly and Skidmore, 2015). For instance, Xie et al. (2024) observed that the phenolic absorption features occur at approximately 1653 nm for Scots Pine, 1667 nm for European Beech, 1664 nm for English Oak, and 1666 nm for Norway Spruce.

Such variations underscore that while phenolic absorption peaks may cluster around 1660 nm, they are not universally fixed. Furthermore, while the study identifies significant spectral features associated with phenolics, it is essential to acknowledge the variability of the absorption bands, as well as the potential confounding factors such as water content, dry matter content, cellulose, and lignin, which may influence spectral interpretations.

In contrast, the lowest spectral pit for flavonoids occurs around 415 nm, with (r) of 47% between flavonoid and 415 nm (Figure 4.6b). This study aligns with Gitelson et al. (2017), who identified the spectral range of 410–430 nm as important for detecting flavonoids. However, Gitelson et al. (2017) also noted that chlorophylls contribute significantly to signals in this range. To reduce the influence of chlorophyll and improve the detection of flavonoids, many researchers have developed special indices, e.g., Gitelson et al. (2017), Zhou et al. (2022), and Zhou et al. (2023). In this study, we tested and modified these spectral indices by focusing on the 415 nm wavelength, which we found to be sensitive to flavonoids.

4.4.3. Flavonoids Indices

$FI_{420, 710}$ and $FI_{415, 710}$, demonstrated the highest accuracy among all the tested indices with R^2 values of 0.55. This demonstrates a moderate but significant correlation between these indices and flavonoid content. This was followed by the mFI-type indices, such as $mFI_{415, 710, 800}$ and $mFI_{415, 710, 800}$, which also performed well. Meanwhile, the spectral indices proposed by Zhou et al. (2022), which replaced 710 nm with 690 nm, did not show improvements in their results, as they produced similar levels of accuracy. This could be due to the fact that both 710 nm and 690 nm are close in the red-edge spectral region, and both wavelengths are strongly related to chlorophyll content (Zhou et al., 2022).

The incorporation of 415 nm, in particular, markedly enhanced the indices' predictive power, indicating its sensitivity to flavonoids. However, the moderate R^2 (~0.55) points to a fair level of correlation, which also implies the presence of potential noise sources or other biological and environmental factors affecting spectral measurements. Future efforts could focus on reducing this noise to improve predictive accuracy further.

In our analysis, we found that only the indices $FI_{420, 710}$ and $FI_{415, 710}$ exhibited predictive power with R^2 values exceeding 0.50, while the remaining indices fell below this threshold. Such low R^2 values indicate a lack of predictive power, reflecting the challenges in accurately correlating leaf reflectance data with the underlying physiological and chemical characteristics of leaves. The spectral measurements were conducted in a controlled environment, specifically a dark

room, which aimed to minimize external light interference and enhance measurement consistency. However, several factors still contributed to the low accuracy. Firstly, the structural characteristics of Fynbos vegetation, particularly the presence of sclerophyllous plants with tough, waxy leaves, created significant interference in spectral measurements.

The limitations of the spectral indices themselves may also play a role; indices developed for other ecosystems might not be fully optimized for the unique characteristics of Fynbos vegetation, where specific leaf traits, such as thickness and composition, differ from those found in other environments. This variability can lead to inconsistent patterns in leaf reflectance, complicating efforts to establish robust correlations.

4.4.4. Challenges and prospects

This study showed that leaf spectroscopy has strong potential for studying the changes in phenolic compounds over time after a fire in the Fynbos Biome. We also tested and adjusted different flavonoid indices to improve accuracy. While our research primarily emphasizes ecosystem functioning, it acknowledges the importance of functional diversity as a key determinant of overall ecosystem performance (Cadotte et al., 2011).

However, this focus presents certain limitations, particularly concerning the lack of detailed information on species-level contributions, as well as the age and physiological state of the analyzed leaves. These factors can profoundly affect leaf reflectance and chemical composition, potentially constraining the interpretative power of our findings. Therefore, future studies should examine how specific species and their varying developmental stages influence leaf reflectance patterns and phenolic concentrations. By integrating these aspects, researchers can develop a more comprehensive understanding of the interplay between species diversity, leaf characteristics, and ecosystem functioning, ultimately enriching our knowledge of dynamics within the Fynbos ecosystem.

Furthermore, the sample size across the chronological sequence is highly unbalanced, with 33 points collected in March 2015, 60 points in February 2016, 12 points in April 2016, 16 points in May 2021, and 37 points in January 2010, along with 32 points from the "never-exp" areas. This disparity presents a limitation in our analysis, comparing datasets of significantly different sizes such as 60 samples against only 12 carries statistical risks. Specifically, the variance in the smaller group may be artificially low due to under-sampling, potentially leading to misleading conclusions. Future studies should aim for more balanced sample sizes to enhance the robustness of comparative analyses.

Although this study highlights that phenolic compounds exhibit sensitivity around 1660 nm, it's important to note that other sensors, particularly those mounted on aircraft or satellites, may not capture this absorption feature effectively. Many of these sensors typically operate with spectral resolutions ranging from 3 to 10 nm, which may not sufficiently resolve the detailed features associated with phenolic absorption. For example, according to Xie et al. (2024), specific wavebands identified by spaceborne sensors, such as 1656 and 1666 nm on the PRISMA satellite (Cogliati et al., 2021), as well as 1653 and 1664 nm on EnMAP (Guanter et al., 2015), are promising candidates for the future operational retrieval of phenolic compounds at the satellite level.

In addition, hyperspectral imaging from aircraft or satellites is different from measuring the reflectance of leaves in controlled settings, as there are several challenges that can affect how we estimate traits. These challenges include the structure of the tree canopy, atmospheric moisture, background light, and varied angles of sunlight and sensors (Darvishzadeh et al., 2011). Additionally, the quality of data in the SWIR range is often lower than in the visible and VNIR ranges, making it difficult to generate accurate information. Therefore, it’s important to consider these factors when mapping phenolic compounds using data collected from the air or space.

In the future, it is essential to explore the relationship between LAI and flavonoids. Although Zhou et al. (2023) assessed this relationship, their study relied solely on the 4-SAIL model, which does not account for other biochemical variables, such as chlorophyll. Therefore, employing models such as PROSPECT-D or PROSPECT-PRO, in conjunction with the 4-SAIL model to create a comprehensive PROSAIL model, would facilitate a more thorough investigation of flavonoids in relation to chlorophyll and LAI. This approach would enhance our understanding of how LAI and chlorophyll influence flavonoid concentrations. Additionally, it is necessary to evaluate the existing flavonoid vegetation indices developed by Gitelson et al. (2017), Zhou et al. (2022), and Zhou et al. (2023) using data from various environmental settings. Such testing will provide insights into the applicability and robustness of these indices across different ecological contexts.

4.5. Conclusions

This study revealed that areas that experienced recent fires, specifically in 2021 (May), 2015 (March), and during February and April of 2016 exhibited high levels of phenolic compounds. Conversely, locations that had not recently experienced fire and areas that have not experienced fire in the past decade demonstrated significantly lower concentrations. These findings underscore the critical role of fire in the rejuvenation of the fynbos ecosystem.

We identified a specific wavelength at 1660 nm that is important for detecting phenolic compounds, along with another key wavelength at 415 nm for flavonoids. However, because chlorophyll also absorbs light in this region, we tested and adjusted various spectral indices for flavonoids to improve their sensitivity. Our results showed that FI-type indices, including $FI_{420, 710}$ and $FI_{415, 710}$, achieved the highest accuracy among all the indices tested. Followingly, the mFI-type indices, such as $mFI_{415, 710, 800}$ and $mFI_{415, 710, 800}$, also performed well. These findings add to existing research by emphasizing the role of fire and PSMs in ecosystem health and function. Additionally, they highlight the importance of spectral indices in improving the detection and measurement of flavonoids from spectral data.

4.6. Chapter summary

This chapter used leaf spectroscopy to estimate the levels of total phenols and flavonoids in relation to fire history. The results show that both fire disturbance and the presence of phenolic compounds are vital for the survival and resilience of the Fynbos biome. Specifically, areas that experienced recent fires in 2015, 2016, and 2021 displayed higher levels of phenolics, while regions that have not recently experienced fire (2010) or had no recent fire had significantly lower concentrations. Using the continuum removal technique, we identified that the strongest spectral signals for total phenols appeared at around 1660 nm, whereas flavonoids showed prominent signals near 415 nm. Based on these insights, we modified existing flavonoid indices to focus on the 415 nm wavelength and compared their performance to other indices. Notably, the indices $FI_{415,710}$ and $FI_{420,710}$ showed the highest correlations with flavonoid content, highlighting the importance of subtle spectral features around 415 and 420 nm for accurately estimating flavonoids. Overall, these findings emphasize that fire and secondary metabolites play crucial roles in ecosystem health and function. Building on this, the next chapter, upscaling to canopy level, explores whether including secondary metabolites in the PROSAIL model can improve the estimation of the PAI in a phenolic-rich landscape.

Chapter Five:

Inclusion of secondary metabolites in PROSAIL model improves remote sensing of Plant Area Index in a phenolic-rich landscape

This chapter is based on:

Sibiya, B., Cho, M., Mutanga, O., Odindi, J., Masemola, C., and Wessel, B., 2025. Inclusion of secondary metabolites in PROSAIL RTM improves remote sensing of PAI in a phenolic-rich landscape. *Journal of Photogrammetry, Remote Sensing and Geoinformation Science,* Under Review, Manuscript ID: ef090ca2-acdb-439d-b8b7-dd5f87f25029

Abstract:

Over recent decades, various PROSAIL models have been developed to improve remote sensing of plant canopies. PROSAIL-4 is one of the early versions and have been widely used. However, in a phenolic-rich landscape, the capability of PROSAIL-4 might be limited due to lack of inclusion of secondary metabolites such as carotenoid and anthocyanins. To address this, PROSAIL-D was introduced, incorporating additional parameters accounting for plant secondary metabolites (PSM) such as carotenoids and anthocyanins. Using field spectral data and AVIRIS-NG sensors, this study tested this hypothesis in the Greater Cape Floristic Region (GCFR) of South Africa; a rich area in secondary metabolites. The results demonstrated that PROSAIL-D consistently yielded higher accuracy, with an R^2 of 0.78 and NRMSE of 3.99% for field spectral data, and an R^2 of 0.72 with NRMSE of 3.86% for AVIRIS-NG data. In contrast, the PROSAIL-4 model achieved R^2 values of 0.68 (NRMSE = 5.88%) for field spectral data and 0.51 (NRMSE = 8.13%) for AVIRIS-NG. Overall, the field spectral data sensor data outperformed AVIRIS-NG in estimating biophysical variables using the PROSAIL model. These findings highlight the significance of incorporating PSM into model simulations, especially in ecosystems like the GCFR, where plants are subjected to harsh and stressful environmental conditions.

Keywords: PROSAIL model, secondary metabolites, AVIRIS-NG, field spectral data, Plant Area Index (PAI), Greater Cape Floristic Region (GCFR).

5.1. Introduction

Biophysical and biochemical properties are essential plant attributes valuable for ecosystem functioning and plant physiology. These characteristics serve as valuable indicators for monitoring alterations in vegetation cover, including increases or decreases in tree canopy. Such changes can be early warning signals for ecosystem variations (Cho et al., 2014a). Consequently, given their critical role in maintaining ecosystem functions and their utility in assessing ecological health, it is essential to systematically map and observe plant attributes over time (Tsele and Ramoelo, 2024; Skidmore et al., 2021). Methods for quantifying plant attributes using remotely sensed data can be broadly classified into empirical and physical approaches (Verrelst et al., 2015). The empirical model is used to find the correlation between plant traits such as LAI and remotely sensed data (Ali et al., 2016; Darvishzadeh et al., 2008). Studies utilize spectral bands (Skidmore et al., 2010; Ramoelo et al., 2013), as well as enhanced spectral bands, to develop vegetation indices (Zhou et al., 2022; Tsele et al., 2022) or image texture (Mngadi et al., 2022; Sibiya et al., 2022; Mthembu et al., 2024). Over the past years, these empirical methods have gained recognition for their efficacy in producing reliable results (Verrelst et al., 2019). However, the empirical models require calibration for each dataset, which makes them site-specific and limits their transferability (Li et al., 2018; Sibiya et al., 2025a). Physical models, such as RTMs, use physical principles to describe how sunlight interacts with plants (Jacquemoud et al., 2009; Jacquemoud and Baret, 1990; Verrelst et al., 2015). RTMs simulate how solar radiation interacts with plant canopies, accounting for leaf structure, biochemical properties, canopy architecture, and soil background. Over the years, many RTMs have been developed for various applications in remote sensing (Widlowski et al., 2015). Among these, the PROSAIL model has become the most widely used (Sibiya et al., 2025b; Berger et al., 2018; Jacquemoud et al., 2009).

Over the past decades, various versions of PROSAIL have been developed to better understand plant canopies using satellite data (Jacquemoud et al., 2009). One of the earliest and most widely used versions is PROSAIL-4, which incorporates parameters such as chlorophyll content, water content, and dry matter (Féret et al., 2008). These parameters are instrumental in assessing general plant health and estimating biomass, and several studies have shown that PROSAIL-4 can often produce accurate and reliable results (Darvishzadeh et al., 2008; Si et al., 2012; Duan et al., 2014; Sinha et al., 2020; Atzberger et al., 2003). However, PROSAIL-4 might not perform well in an environment characterized by an abundance of PSM. To address the shortcomings of PROSAIL-4, a new version called PROSAIL-D has been developed (Féret

et al., 2017), which includes parameters related to PSM, such as carotenoids and anthocyanins, making it potentially more accurate in complex environments where these compounds are abundant (Lu et al., 2021). Therefore, this hypothesis was tested in the GCFR of South Africa, which is renowned for its exceptional biodiversity (Cardoso et al., 2025) and is very rich in secondary metabolites (Kumar et al., 2020; Nandave et al., 2009; Van Staden, 1966).

The GCFR is one of the world's most species-rich and ecologically complex areas, with nearly 10,000 plant species exhibiting a remarkable diversity of colours and is very rich in phenolic compounds (City of Cape Town Biodiversity Report, 2018; van der Merwe et al., 1994). The region's unique environment, however, is characterized by harsh conditions such as limited soil nutrients, intense heat, and drought, which impose significant stress on its vegetation (Magadlela et al., 2015; Bengtsson et al., 2011). These conditions necessitate sophisticated plant adaptations, including the synthesis and regulation of protective pigments such as carotenoids and anthocyanins, which are crucial for survival and ecosystem functioning in such an environment (Manjunath et al., 2016).

Carotenoids play a vital role in plant resilience by acting as antioxidants that protect chloroplasts from oxidative stress caused by high light intensity and temperature extremes (Blackburn, 2007). Their ability to absorb excess light energy and transfer it safely reduces cellular damage, thus maintaining photosynthetic efficiency in stressful conditions (Britton et al., 2004; Zhu et al., 2010). Additionally, carotenoids contribute to the attractant functions during chromoplast development, facilitating interactions with pollinators and seed dispersers necessary for plant reproduction and genetic diversity within the GCFR.

Anthocyanins, on the other hand, provide protective and ecological advantages by serving as antioxidants and visual signals (Manjunath et al., 2016). Their colourful pigmentation not only attracts pollinators like bees and butterflies but also acts as a sunshade, reducing excessive light absorption by leaves, thereby mitigating oxidative damage (Khoo et al., 2017; Gould, 2004). In the context of the GCFR’s diverse floral phenology, the variation in anthocyanin expression influences reproductive timing and pollinator interactions, essential for maintaining the region's rich biodiversity.

Therefore, the inclusion of carotenoids and anthocyanins in PROSAIL-D enhances its capability to differentiate vegetation based on biochemical and structural traits, especially in ecosystems highly enriched with secondary metabolites like the GCFR. Testing this advanced model in such a complex environment not only advances remote sensing methodologies but

also improves our understanding of plant adaptive mechanisms at the landscape and ecosystem levels. Over the past decade, studies have tested PROSAIL-D in various environmental settings, including maple trees, grasslands, and mangroves, demonstrating its versatility and potential for broader ecological studies (Feret et al., 2017; Lu et al., 2021; Zhao et al., 2023). Despite these advancements, there is still a significant knowledge gap regarding its performance in highly diverse, phenolic-rich environments like the GCFR, where secondary metabolites such as carotenoids and anthocyanins play crucial roles in plant adaptation and ecological interactions.

Therefore, this study aimed to evaluate whether incorporating secondary metabolites, specifically carotenoids and anthocyanins, into the PROSAIL enhances the accuracy of remote sensing-based estimates of the Plant Area Index (PAI) in the phenolic-rich landscape of the GCFR. The word PAI was used because the vegetation in the GCFR differs from typical plant structures that primarily rely on leaves for photosynthesis.

PAI is a measure of the total green biomass within a given area, including both foliar and non-foliar (such as stems) green tissues that actively participate in photosynthesis (Campos-Taberner et al., 2016). Unlike LAI, which specifically quantifies the surface area of leaves (López-Serrano et al., 2000), PAI provides a more comprehensive assessment of the total photosynthetically active plant biomass, especially in environments where plants have evolved to perform photosynthesis in stems or other green structures (Holst et al., 2004). In the GCFR, many plant species do not have traditional leaves; instead, they possess green, photosynthetic stems that perform the majority of photosynthesis. These stem-like species have evolved to allocate their resources differently, emphasizing stem-based photosynthesis as their primary means of energy capture, rather than relying on leaves as seen in most r plant communities.

5.2. Methodology

5.2.1. Sampling

Prior to conducting the field surveys, Sentinel-2 satellite imagery was used to identify and target representative areas within the study sites. This helped in the strategic placement of sample plots, ensuring that the selected areas accurately reflected the dominant vegetation types present. The fieldwork was conducted in November 2023, utilizing a purposive sampling approach aimed at capturing the most representative vegetation patches.

During the field surveys, a total of 35 square plots, each measuring 10 by 10 meters, were established. To set up each plot, two 10-meter-long tape measures were positioned perpendicular to each other, forming a cross that marked the centre of the plot. At this central point, coordinates were recorded using a high resolution submeter (50 cm) accurate handheld Trimble Global Navigation Satellite System (GNSS) device. As illustrated in Figure 5.1, the sampling protocol involved subdividing each plot into four subplots of 5 m × 5 m.

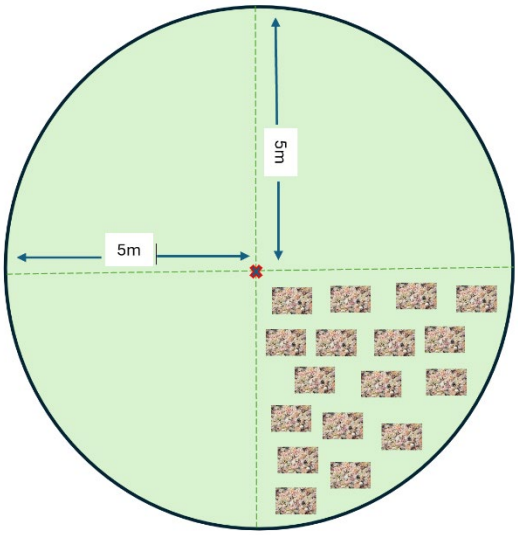


Figure 5.1. A depiction of the sampling protocol.

Within each subplot, five spectral measurements were taken at the top of the canopy using an field spectral data, and readings averaged to produce one spectrum per subplot, resulting in a total of four spectra per plot (a total of 140 spectra across all plots). After applying quality control, 136 spectra were retained for analysis (Figure 5.2). Following spectral data collection, LAI/PAI was measured in the field using a portable LI-COR LAI-2200C Plant Canopy Analyzer. For each plot, canopy readings were obtained using the instrument's standard above (A) and below (B) procedure, where A denotes a sky reference reading above the canopy and B denotes readings below the canopy (Li-Cor, 1992). Specifically, an A–B sequence was used with an effective zenith angle of 38° and a 270° view cap to capture sky radiance under representative conditions, and this sequence was repeated consistently across the sampling area. For each subplot, we took one above-canopy A reading and then took four below-canopy B readings within the same plot to capture small-scale variability in canopy structure; these four below-canopy readings were then averaged to produce a single representative PAI value for that subplot.

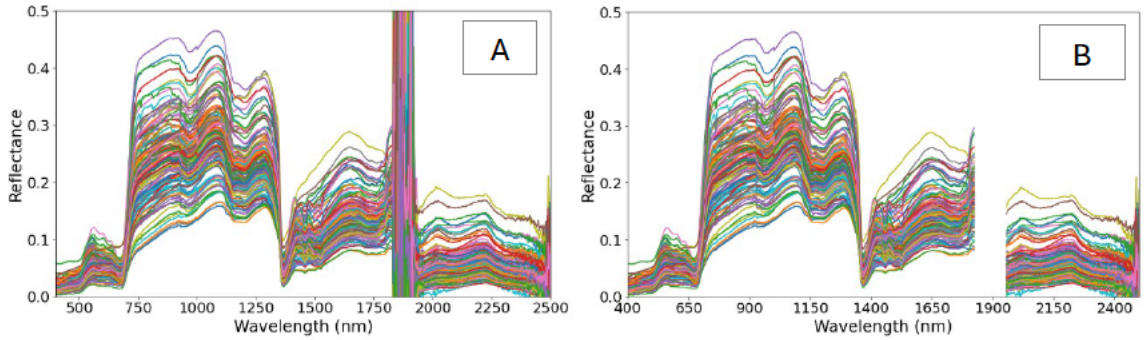


Figure 5.2. field spectral data spectral reflectance (n = 136). Due to high noise regions from 1824 to 1954 nm were excluded.

In addition to spectral and structural measurements, leaf samples were randomly harvested from various canopy branches within each subplot. Leaves were carefully collected from different positions on the canopy, such as the top, middle, and lower branches, to ensure a representative sample. In total 17922.2g of leaves were harvested from dominant species and were placed in zipped plastic bags and then preserved in a cooler box with ice. These leaves were placed in zip-lock bags and stored immediately in a cooler with ice to preserve their moisture content and prevent any degradation. Later, the leaves' fresh weight (W_f) was measured immediately using an electronic scale with a precision of 0.01 g. The leaves were then oven-dried at 80 °C. Thereafter, the dry weight (W_d) was measured. The measured field parameters, including PAI, water (equestion 2), and dry matter content (equestion 1), were later incorporated into the PROSAIL model parameter ranges (Table 5.1).

$$C_m = \frac{W_d}{PAI} \qquad (1)$$

$$\text{EWT} = \frac{W_f - W_d}{PAI} \qquad (2)$$

5.2.2. Hyperspectral images

In October and November 2023, AVIRIS-NG hyperspectral images were acquired over the CFR in South Africa. This project was part of a combined effort called NASA's BioSCape, which involved fieldwork and airborne surveys to study the diverse ecosystems in the GCFR. We used the AVIRIS-NG reflectance mosaic, available via the Oak Ridge National Laboratory Distribution Active Archive Centre (ORNL DAAC) (https://popo.jpl.nasa.gov/pub/bioscape/mosaics/ang/v02/rfl/) (Brodrick et al., 2025). The

image contains data in 425 bands (Figure 5.3a), covering a range from visible light to shortwave infrared, specifically from 377 nm to 2500 nm. However, some bands contain noise that was removed (Figure 5.3b). After removing the noisy bands, we kept 373 bands, which provide detailed information about the vegetation.

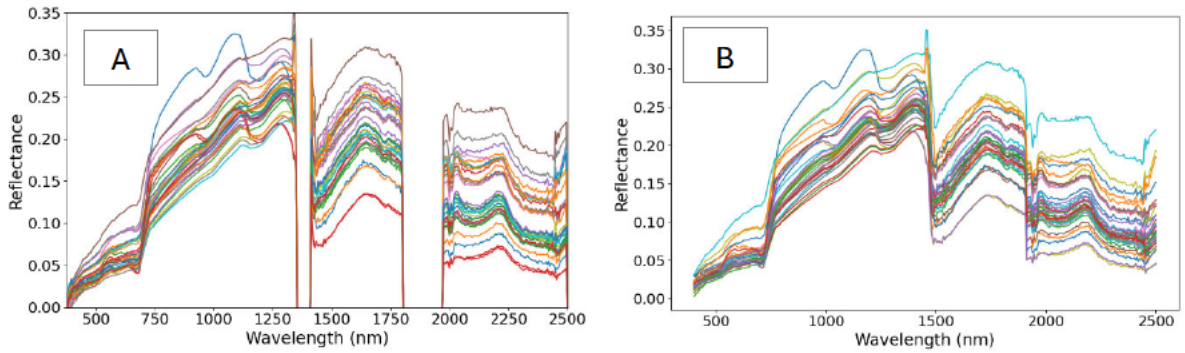


Figure 5.3. The canopy reflectance spectra obtained from AVIRIS-NG data. Specific spectral regions, 377 to 397 nm, 1364 to 1409 nm, and 1805 to 1970 nm, have been omitted due to high levels of noise.

5.2.3. PROSAIL calibration

This study employed the PROSAIL to simulate PAI. PROSAIL combines the principles of SAIL and PROSPECT (a leaf optical property model), allowing for robust canopy reflectance simulation across a broad range of spectral wavelengths (Jacquemoud et al., 2009). The PROSAIL-4 and PROSAIL-D models were parameterized using the data shown in Table 5.1 and utilized in two ways: forward and backward mode.

Table 5.1: Ranges for input parameters used for generating look-up table (LUT) using the PROSAIL model. The model parameter LAI represents PAI.

Parameter	Abb.	Unit	Range or fixed value
Leaf			
Leaf structural parameter	N	No dimension	1 – 1.9
Chlorophyll	C_{ab}	µg cm^{-2}	3 - 70
Carotenoids	C_{ar}	µg cm^{-2}	2 - 25
Brown pigment	C_{brown}	-	0 - 1
Equivalent water thickness	EWT	µg cm^{-2}	0.0012 - 0.052
Dry matter content	C_m	µg cm^{-2}	0.0041 - 0.044
Anthocyanin	C_{anth}	µg cm^{-2}	0 – 40
Canopy			
Leaf area index	LAI	m^2 m^{-2}	0.5 - 5
Hot spot size parameter	Hspot	-	0.05 – 0.1
Sun zenith angle	Tts	Deg	0°
Sensor viewing angle	Tto	Deg	0°
Relative azimuth angle	Phi	Deg	0°
Leaf angle distribution type	Typelidf	-	1
Leaf angle distribution	Lidfa	-	-0.35
Leaf angle distribution	Lidfb	-	0

In the forward mode, PROSPECT-4 was configured using four key parameters: the structural parameter (N), which relates to leaf internal structure and chlorophyll, water and dry matter content (Féret et al., 2008). On the other hand, the PROSPECT-D extends from this model by including three additional parameters: the brown pigment content (cbrown), carotenoid concentration, and anthocyanin concentration (Féret et al., 2017). These additional pigments enable a more detailed simulation of leaf optical properties across the 400 to 2500 nm spectral range. The SAIL model completes the system by simulating canopy reflectance based on structural and optical properties. Key parameters within the SAIL model include LAI, Leaf angle distribution, Hotspot, soil reflection parameter, Solar zenith, Observer zenith, and Relative azimuth angle (Verhoef, 1985).

Before model parameterization, the Global Sensitivity Analysis (GSA) was performed to assess which PROSAIL-D parameters have influenced the retrieval of the PAI model output. This process entails keeping all parameter ranges constant and adjusting one parameter range value at a time, then assessing the variations in the model outputs (Wang et al., 2023). It is also possible to change all parameters concurrently to assess their collective influence on the model output.

The parameter ranges for water content, dry matter, and PAI were measured in the field and were therefore used to constrain the model within the minimum–maximum values observed during the field campaign. For PAI, however, we did not restrict the inversion strictly to the observed field min–max range because PAI can vary within a scene due to differences in canopy structure and viewing geometry, and the retrieval can be sensitive to these effects even when field measurements fall within a limited range. Consequently, we applied a broader, generic LAI range (0.5–5) to ensure that the inversion was not artificially limited by the sampling range of the field campaign and to improve model robustness across the phenological and structural variability expected in the study area.

Furthermore, the ranges of other parameters were set randomly, guided by sensitivity analysis (see Table 5.1). In total, 100,000 random combinations of input parameters, covering plausible ranges for each, were generated to create a comprehensive LUT of simulated reflectance spectra. Figure 5.4 illustrates various spectral combinations produced by PROSAIL-4 and PROSAIL-D. To ensure that the LUT spectra are comparable to those captured by field spectral data (Figure 5.2) and AVIRIS-NG sensors (Figure 5.3), we removed the spectral regions between 1824 and 1954 nm to match the field spectral data spectral range, which does not

extend into this range (Figure 5.4a-b). Additionally, we selected bands within the LUT that correspond to the spectral bands of the AVIRIS-NG sensor. This process resulted in a spectral dataset comprising 373 bands, matching the number of bands in the AVIRIS image (Figure 5.4c-d).

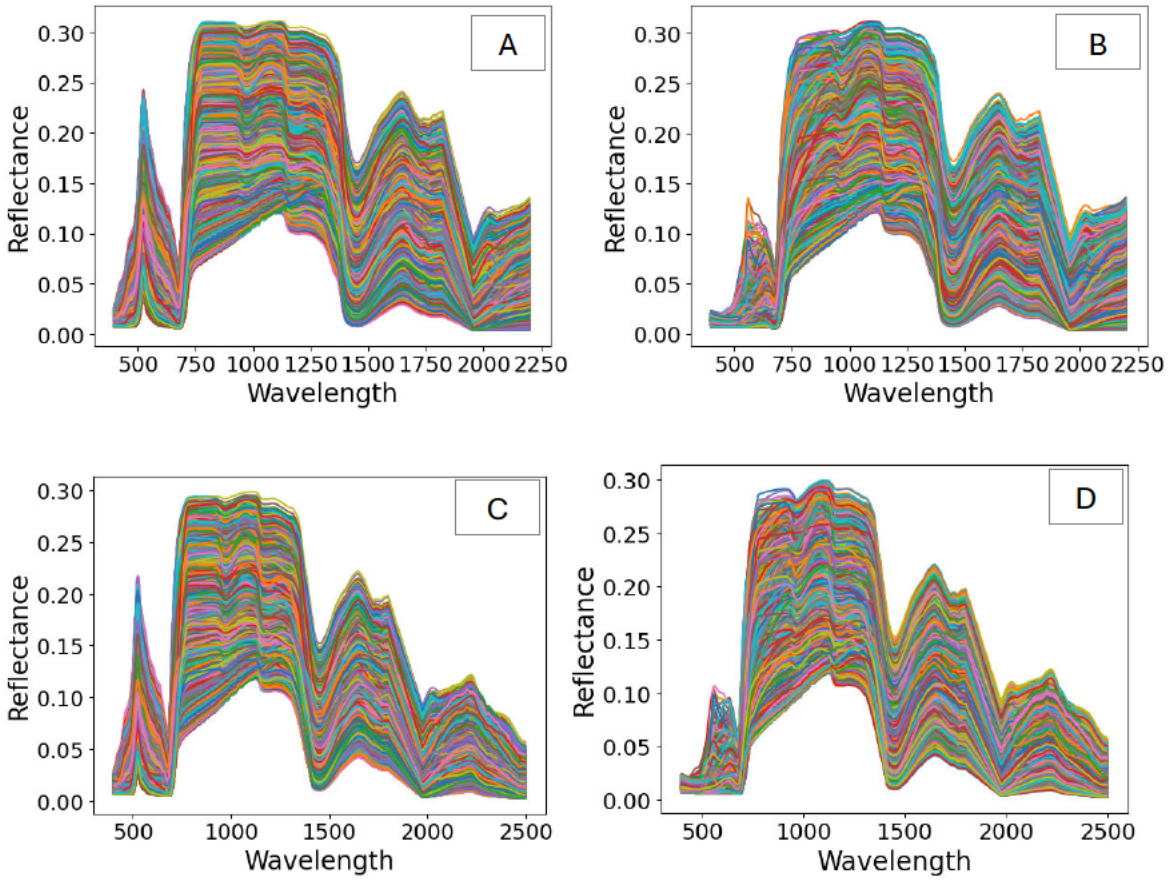


Figure 5.4. Shows 100'000 simulated spectra from PROSAIL-4 (A and C) and (B) PROSAIL-D (B and D), for field spectral data (A and B) and AVIRIS-NG (C and D).

In the backward mode, the measured canopy reflectance spectra from the field spectral data data and AVIRIS-GN were compared against the simulated spectra in the LUT. The RMSE was used as a cost function to iterate 100,000 simulated spectra in the LUT and the best closest spectra to the measured spectra (i.e., field spectral data or AVIRIS data). This process involved calculating the RMSE for each comparison to quantify the degree of similarity between the measured reflectance data and each simulated spectrum. The spectrum exhibiting the lowest RMSE value was identified as the best fit, indicating the closest match to the measured data. To ensure robustness and prevent overfitting, we implemented a 5-fold cross-validation procedure.

Specifically, the set of measured spectra was split into five folds; in each iteration, 70% of the data served as the training set, used to calibrate or identify the optimal spectra, while the remaining 30% served as the testing set to evaluate retrieval accuracy by comparing measured spectra to their best matches. This process was repeated five times, with each fold functioning as the test set once. Figure 5.6 shows an example of the closest match between the measured and simulated spectra from PROSAIL-4 and PROSAIL-D. This was implemented in Python using the PROSAIL package (https://pypi.org/project/prosail/), which allows for both forward simulations and inverse retrievals. Consequently, the LAI associated with this optimal spectrum was retrieved, ensuring that the estimated parameters accurately reflect the actual canopy conditions.

$$\mathrm{RMSE_r} = \frac{\sqrt{\sum_{i-1}^{N}(R_{measured} - R_{lut})^2}}{n} \qquad (3)$$

Table 5.2: The descriptive statistics showing the Maximum (Max), Minimum (Min), Standard Deviation (Std), Range, and Coefficient of Variation (CV).

Plant Attribute	Max	Min	Mean	Std.	Range	CV (%)
LAI	5.32	2.88	4.25	0.74	2.44	17.46%
ETW	0.052	0.0012	0.01	0.009	0.051	76.79%
Dry matter content	0.044	0.0041	0.0097	0.006	0.04	63.26%

5.3. Result

5.3.1. Statistical Analysis

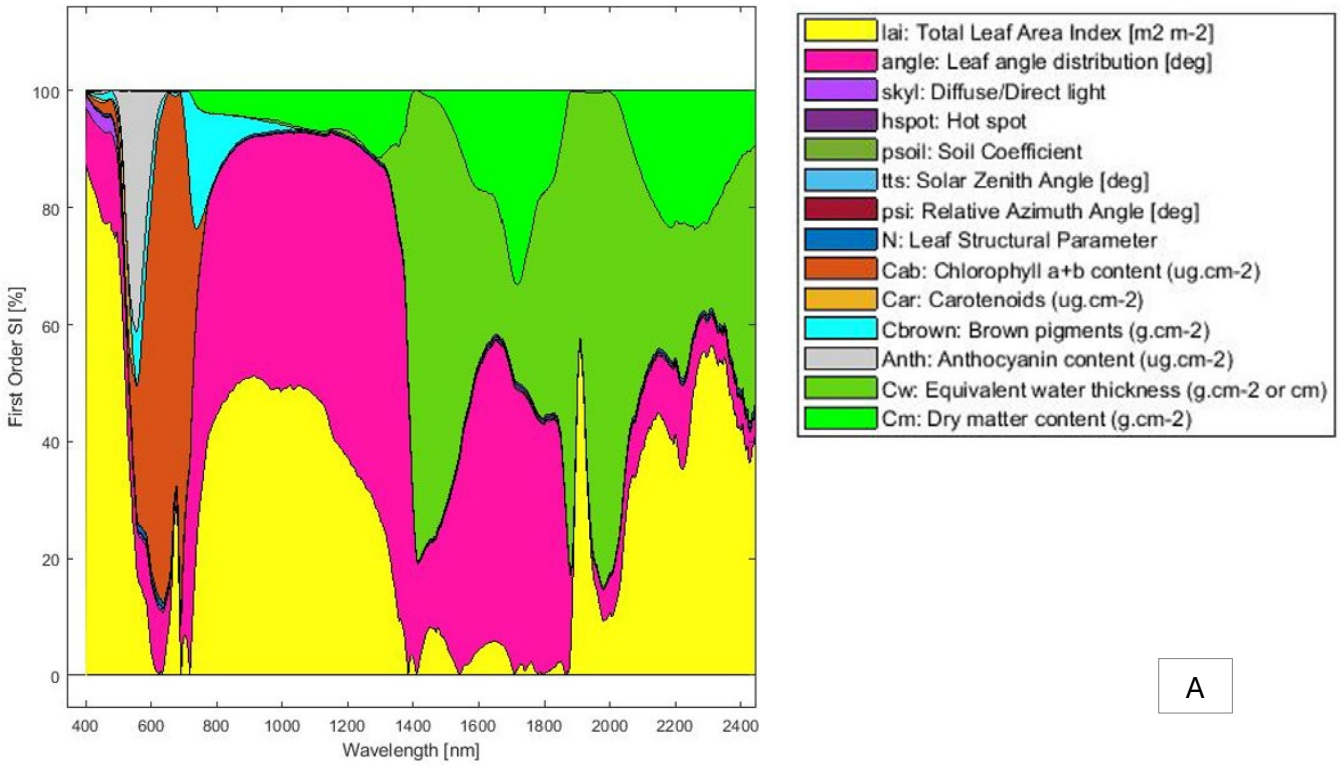
Summary statistics for the three measured plant attributes indicate distinct patterns of central tendency and variability (Table 5.2). PAI values were relatively high and consistent, ranging from 2.88 to 5.32 (mean = 4.25, SD = 0.74, range = 2.44), with a coefficient of variation (CV) of 17.46%, which denotes low to moderate relative variability around the mean. In contrast, ETW exhibited both the largest absolute spread and the greatest relative variability: values ranged from 0.0012 to 0.052 (mean = 0.010, SD = 0.009, range = 0.051) with a CV of 76.79%. This high CV shows that ETW fluctuated substantially among samples and that differences among observations are large relative to the mean value. Dry matter content showed an intermediate pattern: low absolute values (min = 0.0041, max = 0.044) and a mean of 0.0097 with SD ≈ 0.006 (range = 0.040) and a CV of 63.26%, indicating considerable relative variability though slightly less than that of ETW. In summary, these results suggest that

structural leaf properties (PAI) were comparatively stable across the study samples, whereas compositional properties (ETW and dry matter content) varied widely.

5.3.2. Sensitivity analysis of PROSAIL-D parameters

Figure 5.5 illustrates the GSA of both total and first-order sensitivity indices for each PROSAIL-D input parameter across the visible, near-infrared, and shortwave-infrared spectral ranges. The analysis indicates that the PAI has the most significant influence, exhibiting high contributions throughout the entire spectral range. This is similar to Zhao et al. (2023), who find that LAI has a great effect on the whole spectral range. This is followed by the leaf angle distribution, which also shows notable effects across wavelengths.

In the visible region, chlorophyll content has a substantial impact, followed by anthocyanin and brown pigments, whereas carotenoids have a minimal effect. In the shortwave near-infrared region, the reflectance is predominantly influenced by equivalent water thickness and dry matter content. These two parameters show limited contributions in the 400–800 nm spectral region, reflecting their limited absorption within the visible spectrum. This sensitivity analysis is critical for the parameterization and fine-tuning of the PROSAIL-D model, as it highlights how parameter ranges should be adjusted, and which parameters may be held constant to optimize model performance.



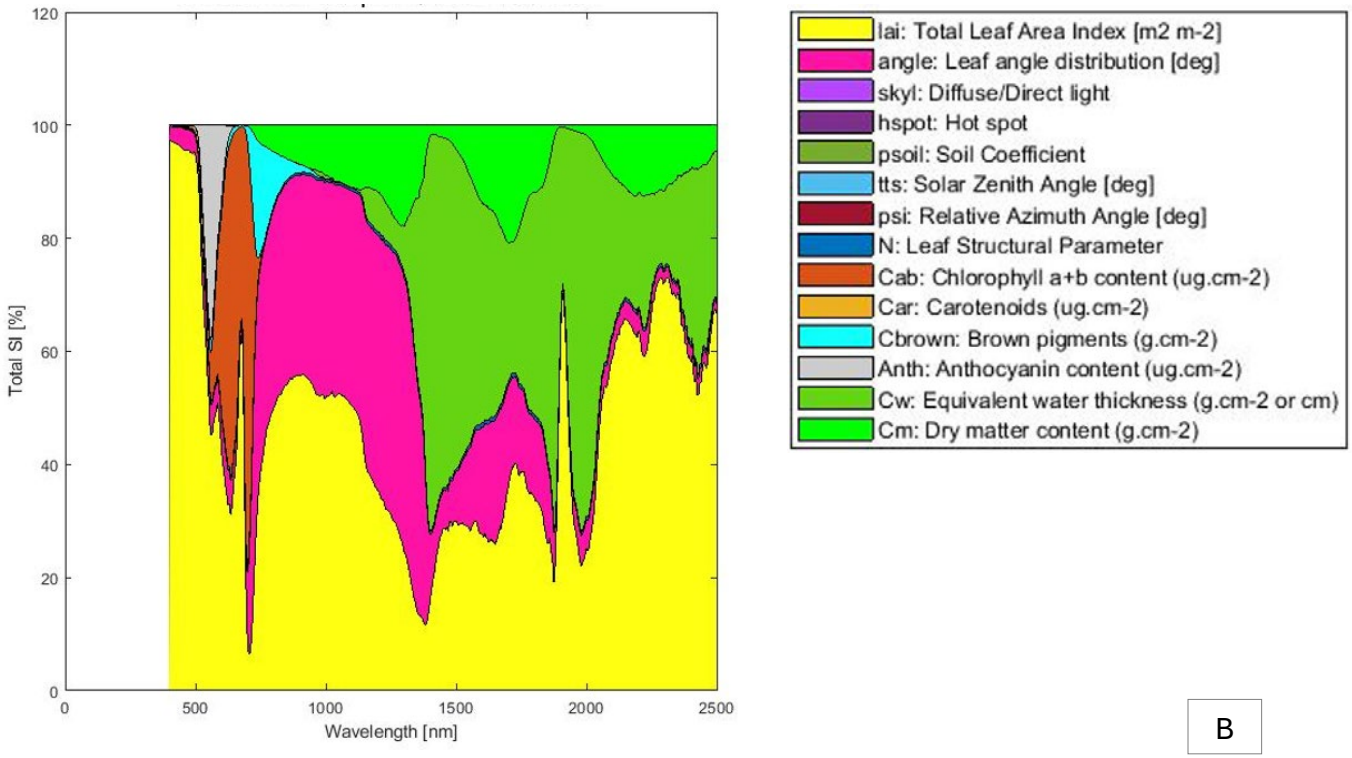


Figure 5.5. First-order Global Sensitivity Index (A) and Total Global Sensitivity Index (B) for PROSAIL-D input parameters.

5.3.3. Results of the PROSAIL models for field spectral data data and AVIRIS-NG image.

This study employed RMSE as the cost function to address the inverse problem; a lower RMSE indicates a closer match. Consequently, the spectral solution that results in the lowest RMSE is considered the most accurate representation of the measured canopy reflectance. Figure 5.6 compares the measured and simulated spectra of the PROSAIL-4 and PROSAIL-D for two datasets: field spectral data and AVIRIS-NG collected within the GCFR. Figure 5.7 displays the RMSE values comparing the simulated with the measured reflectance for field spectral data data and AVIRIS images. Overall, the PROSAIL-D performed better than PROSAIL-4 when applied to field spectral data data and the AVIRIS-NG imagery. As illustrated in Figure 5.6, highlighted by a red circle, there is a huge overlap between simulated and measured spectra when utilizing PROSAIL-4. In contrast, PROSAIL-D exhibits a closer alignment between simulated and observed data, particularly within the visible spectral region.

For the field spectral data data, PROSAIL-D achieved an average RMSE of approximately 0.97, with the highest error being below 2.74. For the AVIRIS-NG images, PROSAIL-D showed an even better performance, with an average RMSE of around 0.49 and the maximum error below 2.69. In comparison, PROSAIL-4 resulted in slightly higher errors. Its average

RMSE for the field spectral data data was about 1.03, with a maximum error below 2.52. For the AVIRIS images, an average RMSE of roughly 0.55 and a maximum below 2.58. These results suggest that PROSAIL-D is more accurate in simulating reflectance data, especially for the AVIRIS-NG images, producing the lowest RMSE, indicating its potential to simulate PAI in the GCFR effectively.

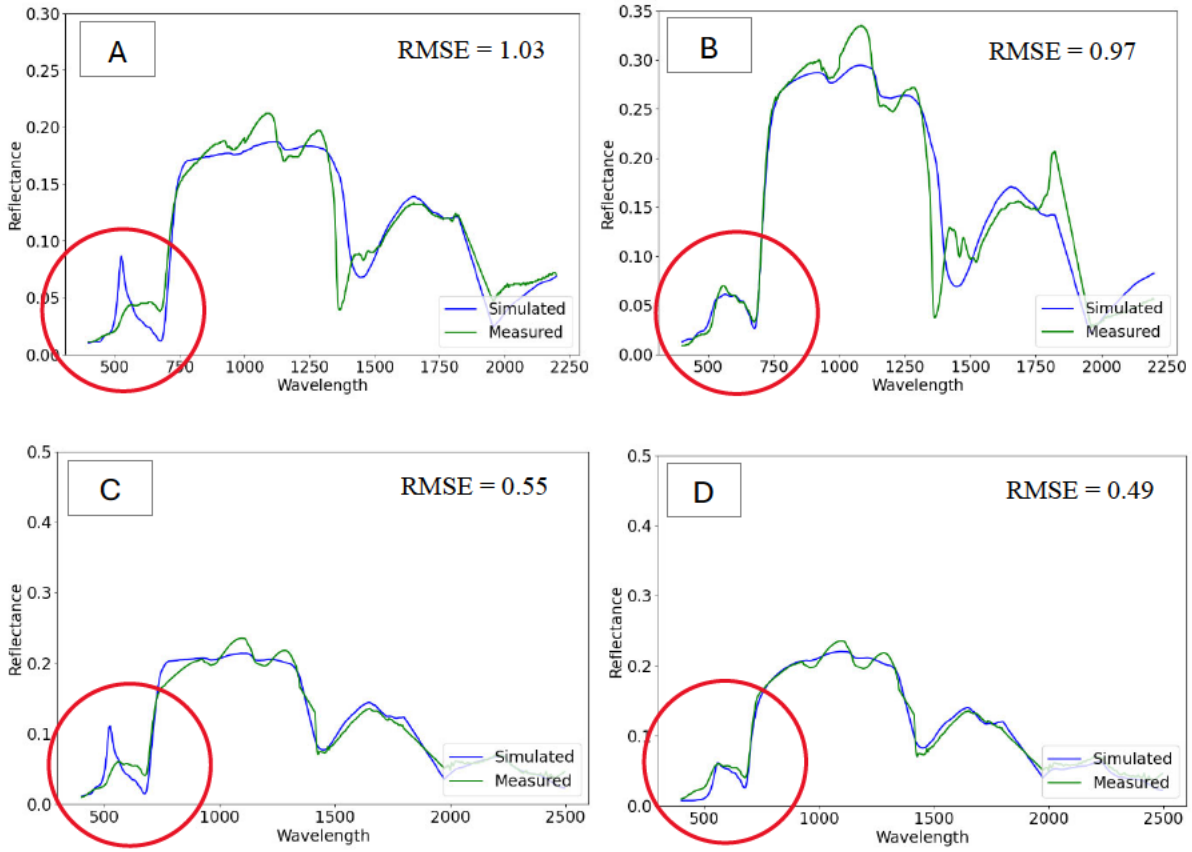


Figure 5.6. An example of the differences between the measured and simulated spectra from PROSAIL-4 (A and C) and (B) PROSAIL-D (B and D), for field spectral data (A and B) and AVIRIS-NG (C and D).

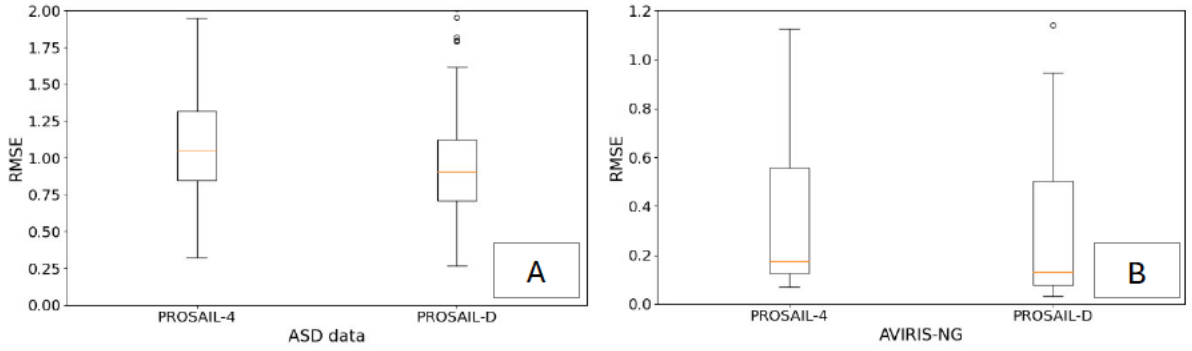
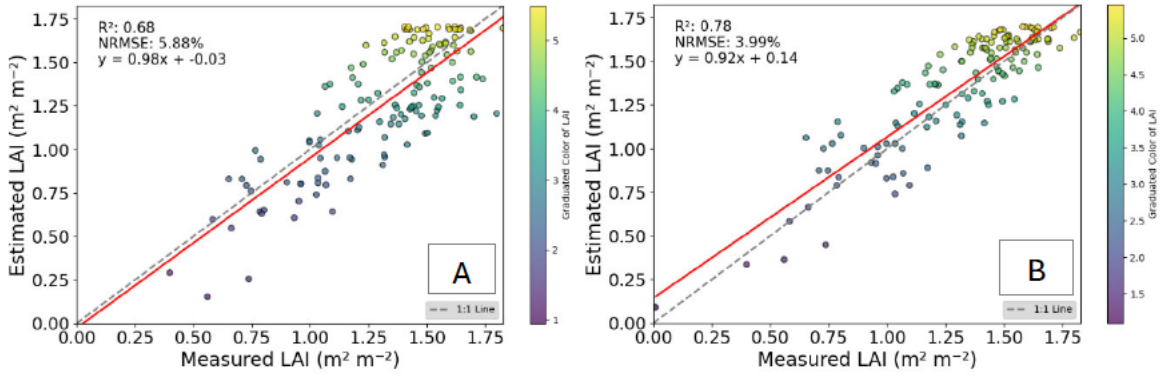


Figure 5.7. RMSE of simulated and measured field spectral data data (A) and AVIRIS-NG (B) with the PROSAIL-4 and PROSAIL-D.

5.3.4. Results of the PROSAIL model

This study estimated PAI using two different PROSAIL models, including PROSAIL-4 and PROSAIL-D, applied to field spectral data and AVIRIS-NG images. For the field spectral data, PROSAIL-D yielded the highest accuracy, with an R^2 value of 0.78, indicating a strong correlation between the estimated and measured PAI (Figure 5.8b). The corresponding NRMSE was 3.99%, reflecting a low percentage error. This suggests that PROSAIL-D was highly effective in estimating PAI from field spectral data. Using AVIRIS-NG images, PROSAIL-D showed slightly lower accuracy but still performed well, with an R^2 of 0.72 and NRMSE of 3.86% (Figure 5.8d). These results demonstrate that PROSAIL-D can reliably estimate PAI from different remote sensing datasets.

In contrast, PROSAIL-4 produced lower accuracy estimates for both datasets. For field spectral data, produced the R^2 of 0.69 with an NRMSE of 5.58% (Figure 5.8a) indicates a moderate correlation. For AVIRIS-NG images, the accuracy was further reduced, with an R^2 of 0.51 and an NRMSE of 8.13%, suggesting less reliable LAI estimates in the GCFR (Figure 5.8c). Furthermore, Figure 5.9 presents the distribution map of the PAI within the GCFR region, demonstrating a more precise representation of PAI distribution when using PROSAIL-D compared to PROSAIL-4. Overall, while PROSAIL-4 demonstrated some capacity to estimate PAI, PROSAIL-D consistently outperformed it across both data sources, providing more accurate and reliable PAI estimations. These findings highlight the potential of PROSAIL-D with secondary metabolites to estimate PAI in the GCFR.



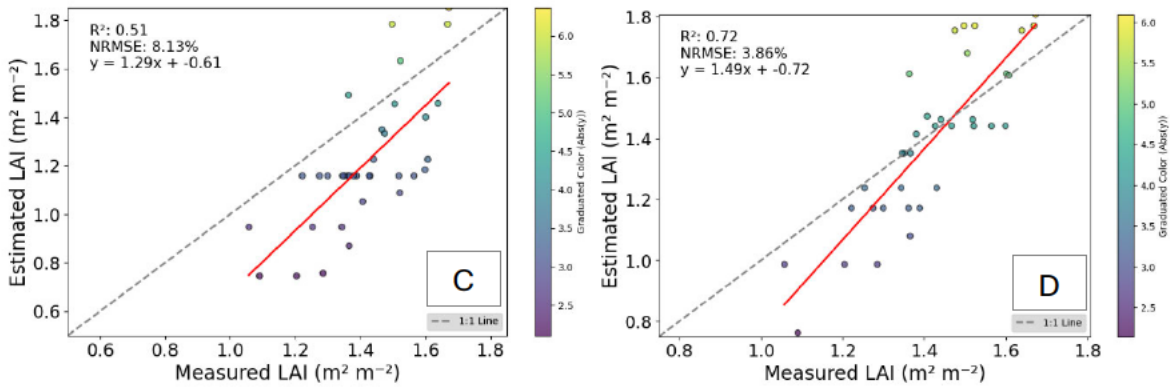
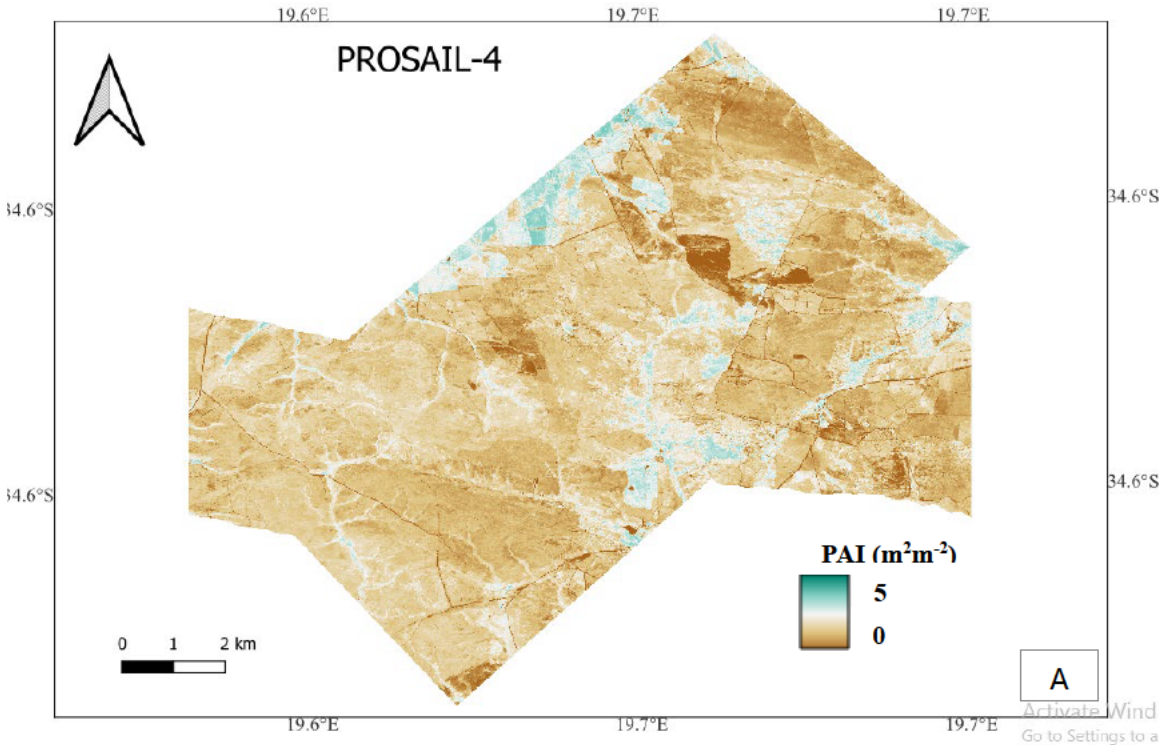


Figure 5.8. Illustrate the relationship between predicted and measured PAI (the model parameter LAI represents PAI). (A) PROSAIL-4 vs measured PAI for field spectral data plots, (B) PROSAIL-D vs measured PAI for field spectral data plots, (C) PROSAIL-4 vs measured PAI for AVIRIS image, and (D) PROSAIL-D vs measured PAI for AVIRIS image.



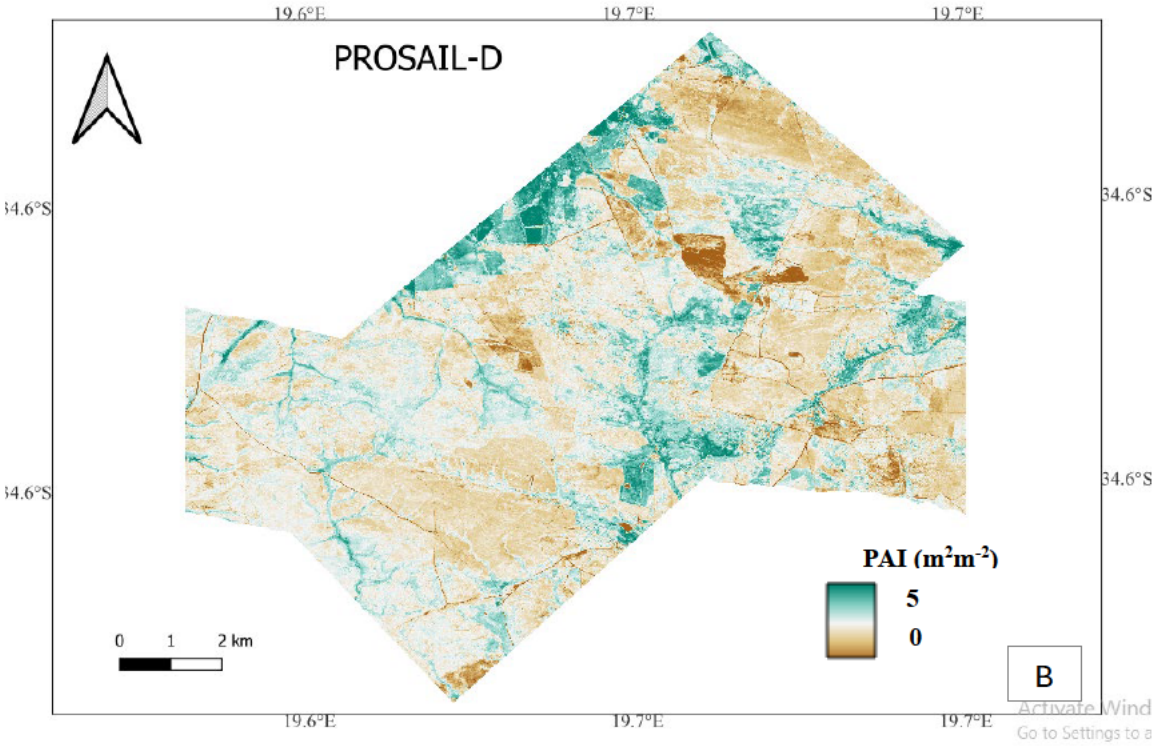


Figure 5.9. PAI maps derived from (A) PROSAIL-4 and (B) PROSAIL-D model.

5.4. Discussion

The results demonstrated that the LUT inversion method applied to the PROSAIL-4 and PROSAIL-D models exhibited significant potential for accurately estimating GCFR PAI. Among the data sources, field spectral data data yielded the highest accuracy, compared to AVIRIS-NG (Figure 5.8). This superior performance of field spectral data data can be attributed to its richer spectral information and the ability to provide real-time spectral measurements captured at close canopy distances (Darvishzadeh et al., 2008a; Tshabalala et al., 2021). The proximity afforded by field spectral data measurements results in high spatial, spectral, and temporal resolution, enabling more effective characterization of the heterogeneity and biochemical complexity within GCFR vegetation (Darvishzadeh et al., 2008b; Cho et al., 2014a; Hansen and Schjoerring, 2003; Mutanga et al., 2004). Furthermore, real-time field spectral data measurements are less susceptible to atmospheric distortions and variations in illumination, factors that frequently impact airborne or satellite data (Mutanga et al., 2009).

Although AVIRIS-NG exhibited slightly lower accuracy compared to field spectral data data, it demonstrated comparable performance. Notably, AVIRIS-NG achieved the lowest average RMSE values of 0.49 and 0.55 for both PROSAIL-D and PROSAIL-4 models when comparing measured and simulated data, respectively (Figure 5.6). Additionally, it also recorded the

lowest NRMSE of 3.86% among all four models tested (Figure 5.8d), indicating high precision in spectral measurements. This sensor provides near-laboratory spectral resolution and low signal-to-noise ratio, critical for accurate PAI estimation in diverse and complex environments like GCFR (Moses et al., 2012; Paramanik et al., 2025; Bhattacharya et al., 2019). Its ability to deliver detailed spectral data from airborne platforms makes it a powerful tool for remote sensing applications, particularly when considering the balance between spatial coverage and spectral fidelity (Hati et al., 2021; Prabhakar et al., 2024). Furthermore, the operational advantages of AVIRIS-NG, such as rapid data acquisition and the capacity to capture high-resolution imagery over large areas, enhance its suitability for ecological monitoring and management (Singh et al., 2020; Verma et al., 2022; Ahmad et al., 2021).

The findings of this study indicate that PROSAIL-D outperformed PROSAIL-4 in estimating PAI, evidenced by higher accuracy with both field spectral data and AVIRIS-NG datasets (Figure 5.8b–d). This enhanced performance is largely driven by PROSAIL-D's ability to incorporate secondary metabolites, such as carotenoids and anthocyanins, which play vital roles in plant physiology and adaptation. Secondary metabolites like carotenoids and anthocyanins absorb light predominantly in the visible spectrum (Huang et al., 2018). PROSAIL-4 does not adequately account for these absorption features, leading to potential underestimations of PAI in biomass-rich areas where anthocyanins are prevalent (Feret et al., 2008). This gap can result in inaccuracies in modeling, as the absorption effects caused by these secondary metabolites overshadow the scattering signals that are essential for PAI estimation.

By integrating the effects of these significant pigments, PROSAIL-D is better equipped to capture the full spectral dynamics associated with plant leaves, allowing for a more accurate representation of PAI across various environmental conditions. Consequently, the results demonstrate the importance of considering both scattering and absorption dynamics in remote sensing models to improve the reliability of vegetation indices and biomass estimations.

Furthermore, this is also shown in Figure 5.6, the spectral simulations using PROSAIL-4 exhibit significant overlap with measured spectra, marked by the red circle, indicating less accuracy. Conversely, PROSPECT-D demonstrates a closer match to observed spectral data, especially within the visible range, underscoring its improved modelling capacity.

In the context of the GCFR, characterized by challenging environmental conditions including limited nutrients like nitrogen and phosphorus (Magadlela et al., 2015; Bengtsson et al., 2011), these pigments are particularly crucial for plant survival and ecosystem function. As a

aforementioned, carotenoids, for instance, contribute to resilience by acting as antioxidants that safeguard chloroplasts from oxidative stress caused by intense light and temperature extremes (Blackburn, 2007). Their ability to absorb excess light energy and safely transfer it reduces cellular damage, thereby maintaining photosynthetic efficiency under stressful conditions (Britton et al., 2004; Zhu et al., 2010). Additionally, carotenoids are involved in chromoplast development and attract pollinators, facilitating plant reproductive success and enhancing genetic diversity within the region. Similarly, anthocyanins offer protective and ecological advantages. They function as antioxidants and serve as visual signals to attract pollinators like bees and butterflies (Manjunath et al., 2016). Moreover, their pigmentation helps shade leaves from excessive sunlight, reducing oxidative damage and improving plant tolerance to high light intensity. In the GCFR's highly diverse floral phenology, variation in anthocyanin expression influences reproductive timing and pollinator interactions, which are essential for sustaining the region's rich biodiversity. Overall, integrating secondary metabolites into spectral models not only improves remote sensing accuracy, but also provides insight into the adaptive strategies of plant species thriving in the extreme condition's characteristic of the GCFR.

In contrast, PROSAIL-4, with its limited parameters and lack of explicit inclusion of traits like carotenoids and anthocyanins (Feret et al., 2017), is less capable of accurately representing the spectral reflectance features specific to the GCFR's vegetation, rendering it less suitable for this environment. These findings align with the study by Cho et al. (2014b), who evaluated LAI estimation across various South African biomes using MODIS satellite data in conjunction with the PROSAIL-4 model. Their results showed reduced accuracy in ecosystems similar to the GCFR, such as the Karoo, emphasizing the importance of including secondary metabolites in radiative transfer models for precise LAI estimation.

The comparison between PROSAIL-D and PROSAIL-4 has demonstrated promising results in simulating biophysical variables within the GCFR using data acquired from field spectral data and AVIRIS-NG sensors. Notably, PROSAIL-D, which incorporates PSM, appears to enhance the model's ability to accurately estimate these variables. However, to fully harness the potential of this approach, further investigation is needed, such as optimizing the selection of spectral regions related to PAI, which could significantly improve model accuracy and robustness (Darvishzadeh et al., 2008).

However, it is essential to acknowledge certain limitations and underlying assumptions associated with this study. First, the PROSAIL model assumes a turbid medium canopy

primarily composed of leaves, but in the GCFR, many plants also have green stems and other structural features that contribute to PAI. This raises questions about how well the model accounts for non-leaf biomass. While the leaf angle distribution parameter was calibrated to handle some structural differences, it may not fully capture the heterogeneity introduced by stems and branches, which could influence spectrally inferred PAI. Additionally, soil reflectance and background reflectance were included as parameters within the model; however, soil spectral measurements were not explicitly collected during this study. This omission could impact the accuracy of the spectral modelling, especially in areas with significant ground cover heterogeneity. Managing ground cover variability remains a challenge, and future work might incorporate soil spectral measurements or multispectral ground data to refine background calibration.

In this study, we acknowledge that the sample size of 35 square plots established during the field survey presents notable limitations. This sample size is statistically borderline for a remote sensing validation study, particularly when data are partitioned for k-fold cross-validation. While our results indicate a commendable R^2 value of 0.78, the reduced sample size may heighten the risk of overfitting the model to specific site characteristics, rather than accurately capturing general biophysical trends. The limited number of plots restricts the statistical power of our analysis, potentially leading to increased variability and uncertainty in our findings, thereby impacting the generalizability of the developed models.

Athough the number of sampling plots is relatively low; our sampling protocol was designed to ensure coverage of the expected range of canopy conditions across the landscape. Plots were intentionally selected to represent contrasting canopy structures and biophysical states, including cases with low PAI and very high PAI. We therefore aimed to capture variability across the land rather than sampling only a narrow portion of the parameter space; for example, the presence of both low and high PAI plots suggests that the observations were not heavily clustered around a single condition. Nonetheless, we recognize that with only 35 plots, we cannot guarantee full representativeness of all combinations of canopy properties, and this may limit the extent to which our findings generalize beyond the sampled scenarios.

Furthemore, these plots represent a snapshot in time, with sampling conducted during the peak spring period (November), thus not accounting for seasonal variations in pigment concentrations, such as anthocyanins and carotenoids, which can significantly influence spectral reflectance. Such phenological changes could impact the generalizability of the

findings throughout the year. Moreover, while the spectral models incorporated secondary metabolites to improve accuracy, actual biochemical quantification of pigment concentrations (e.g., chlorophyll, carotenoids, anthocyanins) was beyond the scope of this study. Direct measurements of leaf pigment levels could serve as valuable validation data for future model refinement, ensuring that the spectral effects associated with plant biochemical traits.

We acknowledge that LAI (and canopy scattering) tends to exert its strongest influence in the NIR. Therefore, future studies could restrict the LAI inversion to the 750–1300 nm range, provided the relevance and added value of the secondary metabolites in the remaining wavelengths have been assessed beforehand. Moreover, future research should explore the application of various structural indices such as the Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), and the Simple Ratio (SR) to assess their efficacy in conjunction with PROSAIL-based modelling. Additionally, hybrid inversion techniques that combine machine learning approaches with PROSAIL modelling could potentially enhance the accuracy and robustness of biophysical parameter retrievals, offering a promising avenue for future investigations. For instance, machine learning could help capture non-linear interactions between parameters or vegetation indices might simplify the detection of pigment effects. Additionally, future studies should explore options of model inversion such as the mean of several solution (5, 10, 20, 50). Literature shows that the selection of single best fitting spectra may not be the optimal strategy of inverting the LUT as this is prone to ill-poseness.

5.5. Conclusion

In this study, both PROSAIL-D (with PSM) and PROSAIL-4 (without PSM) were applied to estimate PAI, in the GCFR. The key findings are as follows:

- The field spectral data sensor data yielded the highest accuracy in estimating PAI compared to AVIRIS-NG data.
- PROSAIL-D consistently outperformed PROSAIL-4 in estimating PAI across both datasets, underscoring the importance of accounting for PSM in model simulations.

To our knowledge, this study is the first to test the inclusion of secondary metabolites in PROSAIL in a phenolic-rich landscape to estimate PAI. The results underscore the PSM in radiative transfer models, which can enhance the fidelity of biophysical property estimations. These insights are valuable for advancing remote sensing applications in vegetation monitoring

and management, particularly in regions characterized by complex vegetation interactions and stress responses.

5.6. Chapter summary

This chapter investigated whether including secondary metabolites in the PROSAIL model improves the estimation of PAI in the phenolic-rich Fynbos biome. We compared the performance of two models: PROSAIL-4, which does not consider secondary metabolites, and PROSAIL-D, a modified version that includes these compounds, which include carotenoids and anthocyanins. The results showed that PROSAIL-D, which incorporates secondary metabolites, more accurately retrieved PAI in the phenolic-rich environments of the GCFR. This highlights the important influence these secondary metabolites have on canopy reflectance and underscores the need to include carotenoids and anthocyanins in radiative transfer models like PROSAIL-D to enhance biophysical parameter estimation in complex, metabolite-rich landscapes. Based on the improved performance of PROSAIL-D, the next chapter will focus on estimating flavonoids using PROSAIL-D inversion applied to imaging spectroscopy data.

Chapter Six:
Estimating flavonoids using radiative transfer model inversion on imaging spectroscopy data

This chapter is based on:

Sibiya, B., Cho, M., Mutanga, O., Odindi, J., Masemola, C. and Van Staden, J., 2025. Estimating flavonoids using radiative transfer model inversion on imaging spectroscopy data. *Frontiers in Remote Sensing*, Under Review, Manuscript ID: 1701367

Abstract:

Assessment of flavonoid content provides essential information on how plants adapt to biotic and abiotic stresses. Flavonoids low concentration between 410 to 430 nm makes their quantification difficult due to the interference of chlorophyll and other pigments. Previous work has made progress in enhancing the detectability of flavonoids by proposing various spectral indices, however, such indices are primarily based on empirical methods. Hence, they often lack transferability across different environments, limiting their broader applicability. In this study, we employed the PROSAIL-D radiative transfer model (RTM) in conjunction with AVIRIS imagery to improve the estimation of flavonoid content through spectral indices. Notably, the spectral index $FI_{417,693}$ demonstrated the highest accuracy, with an R^2 value of 0.63 and an RMSE (mg CAE/g) of 0.24. The second most effective index, $FI_{417,713}$, yielded an R^2 of 0.44 and an RMSE (mg CAE/g) of 0.43. These results underscore the significant potential of integrating AVIRIS data with PROSAIL for non-destructive, remote sensing-based assessment of flavonoid content at the canopy level. This approach offers a promising pathway toward more accurate, transferable, and scalable monitoring of plant biochemical traits, which is crucial for understanding plant stress responses and ecosystem health.

Keywords: Flavonoids, PROSAIL-D model, AVIRIS-NG, Spectral indices.

6.1. Introduction

Flavonoids are located in the outer layers of plants, primarily within the cuticle and epidermis (Gitelson et al., 2017). They play a crucial role in helping plants protect themselves from various stress factors, such as UV rays, water scarcity, nutrient deficiencies, and diseases caused by pathogens and insects (Tattini et al., 2004; Gui-bin et al., 2013; Agati et al., 2016; Treutter, 2006). By absorbing UV radiation, flavonoids shield plant tissues from damage, thereby enhancing their survival and resilience. In addition to their significance in the plant kingdom, flavonoids offer numerous health benefits for humans. They are known for their potential anti-diabetic effects, contributing to improved insulin sensitivity and glucose metabolism (Zhou et al., 2022, 2023).

Moreover, flavonoids provide protection against cardiovascular diseases by promoting endothelial function and reducing inflammation, as well as exhibiting anti-cancer properties through mechanisms such as apoptosis induction and inhibition of tumor growth (Wu et al., 2020; Williamson et al., 2000; Tshabalala et al., 2021). Furthermore, flavonoids possess distinct spectral features that facilitate their non-destructive detection using remote sensing technologies. For example, flavonoids strongly absorb shortwave blue light (410-430 nm), making them detectable in various environmental monitoring applications (Gitelson et al., 2017).

However, in the spectral region of 410-430 nm, various plant pigments, such as chlorophylls, carotenoids, and anthocyanins, also absorb light, which affects the detection of flavonoids (Gitelson et al., 2017). In order to enhance the sensitivity of flavonoid content in this region and limit the influence of other pigments, specifically the most abundant chlorophyll. Over the years, various spectral indices have been developed (e.g., Merzlyak et al., 2005; Gitelson et al., 2017; Zhou et al., 2022). Although these spectral indices have shown promising results, they are primarily based on empirical methods, meaning they are tested and calibrated on the specific datasets, which may limit their applicability across different environments or plant species, consequently limiting their transferability (Verrelst et al., 2015; Sibiya et al., 2025b; Berger et al., 2018).

To date, most flavonoid remote sensing studies have been empirical (Gitelson et al., 2017; Zhou et al., 2022; Tshabalala et al., 2021); the potential of physical RTM inversion approaches for estimating flavonoids remains under-explored. RTMs are generally considered more robust across different seasons, sites, and sensor types, as they simulate how light interacts with plant

canopies (Verrelst et al., 2019; Darvishzadeh et al., 2008a; Cho et al., 2014a). Zhou et al. (2023) made initial progress to using SAIL to estimate flavonoids, but canopy-scale applications are lacking. This study addresses this gap by developing a physically grounded inversion approach to estimate flavonoids across diverse canopy conditions. Therefore, the current study evaluated the performance of flavonoid-related spectral indices using PROSAIL-D and AVIRIS-NG hyperspectral imagery in the GCFR. This is the first study to map flavonoid content at the canopy level using an airborne imaging spectrometer (AVIRIS-NG) coupled with a physical radiative transfer model.

6.2. Materials and Methods

6.2.1. Data collection and chemical analysis

A total of 35 samples were collected during the field campaigns in November 2023, utilizing a purposive sampling approach guided by Sentinel-2 imagery to identify representative vegetation patches. At each plot, two 10-meter tapes were laid perpendicular to one another at the plot's centre, and the GPS coordinates recorded to precisely locate each sampling point (Figure 6.1). Field measurements were closely paired with image pixels by ensuring that the plots fell within specific Sentinel-2-pixel boundaries. Within each plot, LAI was measured using a portable LAI-2200 analyser, providing a rapid estimation of canopy density. Additionally, leaf samples were randomly harvested from various canopy levels, top, middle, and lower branches to capture intra-canopy variability. These samples were weighed fresh, oven-dried at 80 °C, and dry weights were measured.

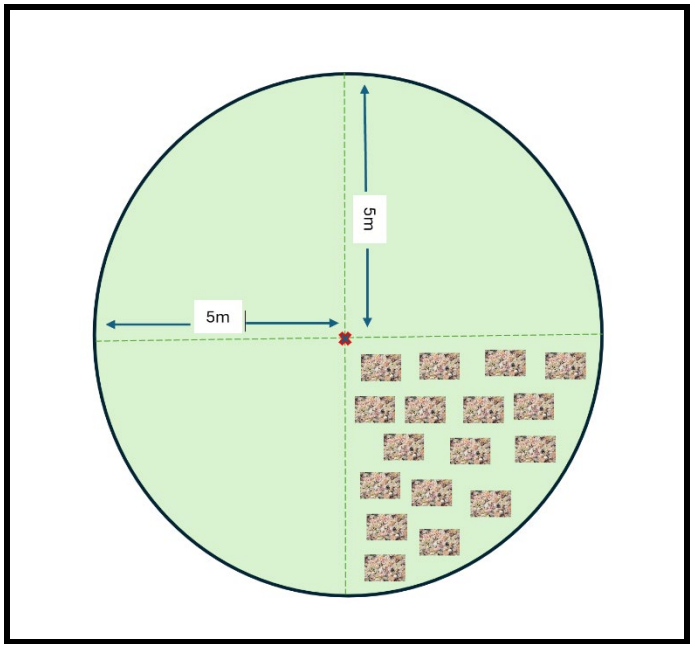


Figure 6.1. Depiction of the sampling protocol.

The dried samples were ground using a coffee grinder, and for extraction, 0.5 g of leaf powder was mixed with 5 mL of a 50% water and 50% methanol solution, sonicated for 20 min and centrifuged for another 20 min. The supernatants obtained were utilized for flavonoid quantification through aluminium chloride colorimetric assays, adhering to protocols established by Markar (1999) and further detailed by Vambe et al. (2024). The assay procedure entailed mixing 250 µL of the extract with 1 mL of distilled water, followed by sequential addition of 75 µL of sodium nitrite ($NaNO_2$), 75 µL of aluminium chloride ($AlCl_3$), and 0.5 mL of sodium hydroxide (NaOH) in a test tube placed in a rack. Catechin at a concentration of 0.1 mg/mL was employed as the calibration standard. Absorbance readings were conducted at 725 nm using a spectrophotometer. Each assay was performed in triplicate to ensure analytical precision, and the quantitative results expressed as the mean value with associated standard error.

6.2.2. AVIRIS data and Flavonoid Indices

In this study, we utilized the processed AVIRIS-NG hyperspectral image captured on November 9, 2023, which is accessible via the Oak Ridge National Laboratory Distributed Active Archive Centre (ORNL DAAC) at the following link: https://popo.jpl.nasa.gov/pub/bioscape/mosaics/ang/v02/rfl/ (Brodrick et al., 2025). The AVIRIS-NG spectrometer records spectral reflectance across a broad wavelength range from 380 to 2510 nm, with a spectral resolution of 5 nm. The original dataset comprises 425 spectral bands, covering regions from visible light to the shortwave infrared. During pre-processing,

bands contaminated with noise were identified and removed (Figure 6.2). After excluding the noisy bands, 373 high-quality spectral bands remained, providing comprehensive and detailed information about vegetation characteristics critical for subsequent analyses.

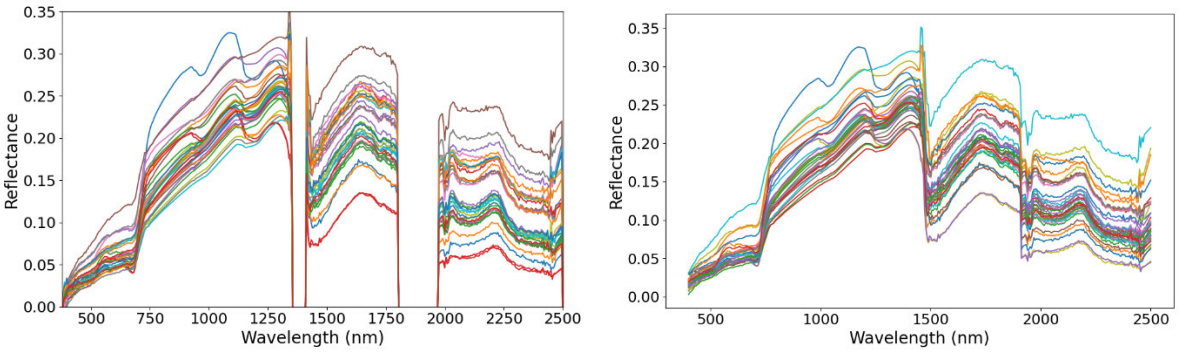


Figure 6.2. Example of AVIRIS-NG spectrum (a) Raw; (b) After excluding the noisy bands.

Previous studies identified the 420 nm band to be sensitive to flavonoids (Zhou et al., 2022; 2023). However, AVIRIS-NG has a spectral sampling of ~5nm and it doesn't have a band center at exactly 420.0 nm. AVIRIS-NG bands are centred at 417 nm and 422 nm, requiring an adjustment of the index originally designed for 420 nm (Table 6.1).

Table 6.1: Flavonoids-related indices.

Index	Formulation	Reference
$FI_{412,688}$	$R\frac{-1}{412} - R\frac{-1}{688}$	Modified from Zhou et al. (2022).
$FI_{412,693}$	$R\frac{-1}{412} - R\frac{-1}{693}$	Modified from Zhou et al. (2022).
$FI_{412,708}$	$R\frac{-1}{412} - R\frac{-1}{708}$	Modified from Gitelson et al. (2017
$FI_{412,713}$	$R\frac{-1}{412} - R\frac{-1}{713}$	Modified from Gitelson et al. (2017)
$FI_{417,688}$	$R\frac{-1}{417} - R\frac{-1}{688}$	Modified from Zhou et al. (2022).
$FI_{417,693}$	$R\frac{-1}{417} - R\frac{-1}{693}$	Modified from Zhou et al. (2022).
$FI_{417,708}$	$R\frac{-1}{417} - R\frac{-1}{708}$	Modified from Gitelson et al. (2017).
$FI_{417,713}$	$R\frac{-1}{417} - R\frac{-1}{713}$	Modified from Gitelson et al. (2017).
$FI_{422,688}$	$R\frac{-1}{422} - R\frac{-1}{688}$	Modified from Zhou et al. (2022).
$FI_{422,693}$	$R\frac{-1}{422} - R\frac{-1}{693}$	Modified from Zhou et al. (2022).
$FI_{422,708}$	$R\frac{-1}{422} - R\frac{-1}{708}$	Modified from Gitelson et al. (2017).
$FI_{422,713}$	$R\frac{-1}{422} - R\frac{-1}{713}$	Modified from Gitelson et al. (2017).

$FI_{427,688}$	$R\frac{-1}{427} - R\frac{-1}{688}$	Modified from Zhou et al. (2022).
$FI_{427,693}$	$R\frac{-1}{427} - R\frac{-1}{693}$	Modified from Zhou et al. (2022).
$FI_{427,708}$	$R\frac{-1}{427} - R\frac{-1}{708}$	Modified from Gitelson et al. (2017).
$FI_{427,713}$	$R\frac{-1}{427} - R\frac{-1}{713}$	Modified from Gitelson et al. (2017).
$FRI_{412,462}$	$\frac{-1}{412} - R\frac{-1}{462}$	Modified from Merzlyak et al. (2005).

6.2.3. PROSAIL-D model

In this study, the PROSAIL-D model was employed, which integrates the PROSPECT model (Jacquemoud and Baret, 1990) and the SAIL model (Verhoef, 1985) to form a comprehensive framework for analysing leaf and canopy reflectance. Both PROSPECT and SAIL are designed to be relatively straightforward, requiring a limited number of input parameters while maintaining an efficient computation time, thereby facilitating the estimation of both leaf and canopy parameters through model inversion.

The PROSPECT-D model was configured with seven input variables: the leaf structure parameter, chlorophyll content, carotenoids, brown pigments, equivalent water thickness, leaf mass per area, and anthocyanins (Jacquemoud et al., 1990; Féret et al., 2017). The leaf optical properties, specifically leaf reflectance and transmittance, as simulated by the PROSPECT model, were subsequently integrated into the SAIL model.

The SAIL model was utilized to simulate canopy reflectance, incorporating both structural and optical properties. Key parameters include LAI, leaf angle distribution, the hotspot effect, background reflectance (soil reflectance) for each wavelength, solar zenith angle, observer zenith angle, and relative azimuth angle (Verhoef, 1985). A soil brightness parameter (psoil) was implemented to account for variations in soil brightness induced by moisture content and surface roughness (Atzberger et al., 2003; Darvishzadeh et al., 2008). Field measurements were conducted to obtain values for water content, dry matter, and LAI.

Furthermore, a sensitivity analysis was performed focusing on five parameters (N, Cab, Car, Canth, hspot), and their respective ranges were established. All input parameters for the PROSAIL model are detailed in Table 6.2. To generate the look-up table (LUT), the PROSAIL model was executed in forward mode to simulate canopy reflectance across a variety of parameter combinations, with simulations conducted for 1,000, 5,000, 10,000, 50,000, and 100,000 iterations. An LUT comprising 1,000 parameter combinations was determined to

provide an effective balance between computational resource demands and the accuracy of canopy variable estimations. Figure 6.3 shows the example of the simulated spectra using PROSAIL-LUT based on parameter ranges.

Table 6.2: Forward mode ranges used in PROSAIL-D.

Parameter	Abb.	Unit	Range or fixed value
Leaf			
Leaf structural parameter	N	No dimension	1 – 1.9
Chlorophyll	C_{ab}	µg cm^{-2}	3 - 70
Carotenoids	C_{ar}	µg cm^{-2}	2 - 25
Brown pigment	C_{brown}	-	0 – 1
Equivalent water thickness	EWT	µg cm^{-2}	0.0012 – 0.02
Dry matter content	C_m	µg cm^{-2}	0.0041 – 0.016
Anthocyanin	C_{anth}	µg cm^{-2}	0 – 40
Canopy			
Leaf area index	LAI	m^2 m^{-2}	0.5 - 5
Hot spot size parameter	hspot	-	0.05 – 0.1
Soil brightness parameter	scale	No dimension	0.5–1.5
Sun zenith angle	Tts	Deg	Fixed per pixel
Sensor viewing angle	Tto	Deg	Fixed per pixel
Relative azimuth angle	Phi	Deg	Fixed per pixel
Leaf angle distribution type	typelidf	-	1
Leaf angle distribution	Lidfa	-	-0.35
Leaf angle distribution	Lidfb	-	0

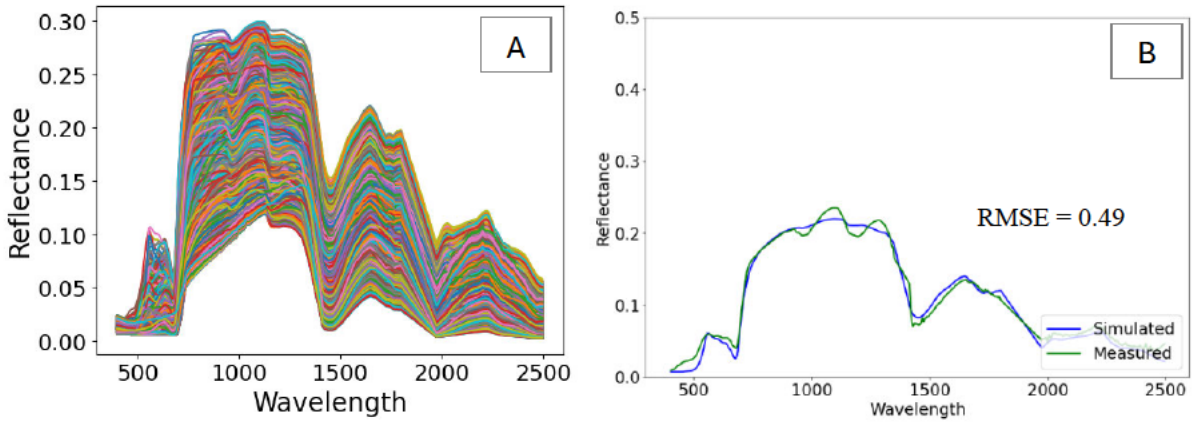


Figure 6.3. The simulated spectra using PROSAIL-LUT based on parameters ranges (A) and example of measured vs. PROSAIL-simulated spectra (B).

For the selection of the optimal solution to the inverse problem, the LUT was ranked according to the cost function defined by the RMSE between the measured reflectance ($R_{measured}$) and the simulated reflectance (R_{lut}) derived from the LUT (Combal et al., 2002). The inversion process employed vegetation indices rather than utilizing the complete spectral range of 400 to 2500 nm, as detailed in Table 1. This process commenced with the calculation of vegetation indices using specific combinations of AVIRIS spectral bands, instrumental in estimating various plant

parameters, including flavonoid content. The indices were subsequently compared against those generated from the LUT encompassing 1,000 simulated spectra produced by the PROSAIL-D model, thereby serving as a crucial reference for validating the measured data.

The quality of the correspondence between the measured AVIRIS-derived indices and those obtained from LUT simulations was evaluated using RMSE as the cost function (eq.1):

$$\text{RMSE} = \frac{\sqrt{\sum_{i-1}^{N}(R_{measured} - R_{lut})^2}}{N} \qquad \text{eq.1}$$

Given that the PROSAIL-D model does not explicitly incorporate flavonoid content as an input parameter, LAI was employed as a proxy. This approach was substantiated by observing a positive correlation between LAI and flavonoid content, where the correlation coefficient was determined to be r=0.61 ($p < 0.05$), indicating statistical significance (Figure 6.4).

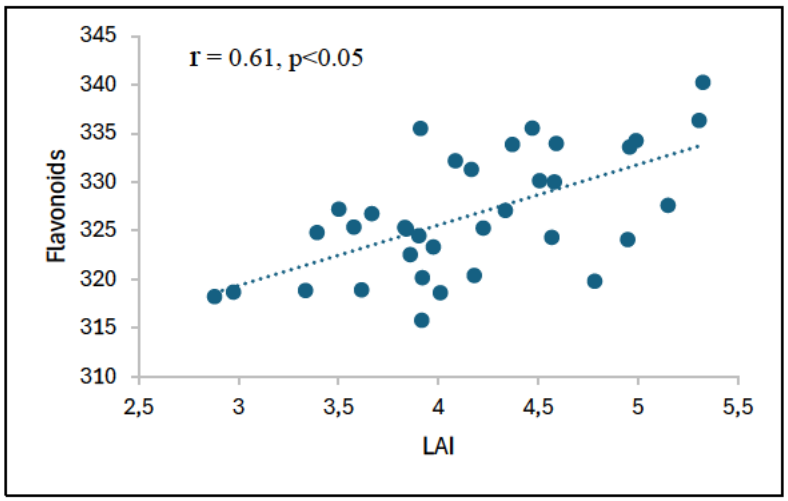


Figure 6.4. The relationships between field LAI and flavonoids.

This iterative process was executed utilizing Python in conjunction with the PROSAIL package (https://pypi.org/project/prosail/) (Gómez-Dans, 2019), which facilitates both forward spectral simulations and inverse retrievals, thereby enabling effective estimation of vegetation and canopy parameters from remote sensing data. Finally, the linear relationship between simulated LAI and flavonoid content was modelled using cross validation CV-R2 (eq.2) and RMSE (eq.3), providing a quantifiable assessment of the extent to which the simulated data elucidated the observed results.

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

$$\text{RMSE}_r = \frac{\sqrt{\sum_{i-1}^{N}(Y_{measured} - Y_{predicted})^2}}{N} \quad \text{eq.3}$$

6.3. Results

6.3.1. Result of PROSAIL-D

The RMSE was used to compare the simulated indices from the LUT with the measured indices from AVIRIS. The index with the lowest RMSE was considered the best match and most accurate in showing the real optical properties of the canopy.

Figure 6.5 presents the average RMSE values between simulated and measured indices. Among the indices tested, $FI_{417,693}$ and $FI_{422,693}$ achieved the lowest average RMSEs, both around 0.002, with maximum errors remaining below 0.015 and 0.012, respectively. Following closely, $FI_{427,688}$ exhibited an average RMSE of approximately 0.003, with its maximum error remaining below 0.071, suggesting reliable performance with slightly more variance. In contrast, $FI_{427,713}$ showed significantly higher errors, with an average RMSE of 3.01 and a maximum error reaching up to 31.85. The maximum observed error of 31.85, appears to be influenced by a few high-value outliers. These outliers may result from anomalous spectral responses, measurement uncertainties, or heterogeneity in the sampled plots that are not fully captured by the current model.

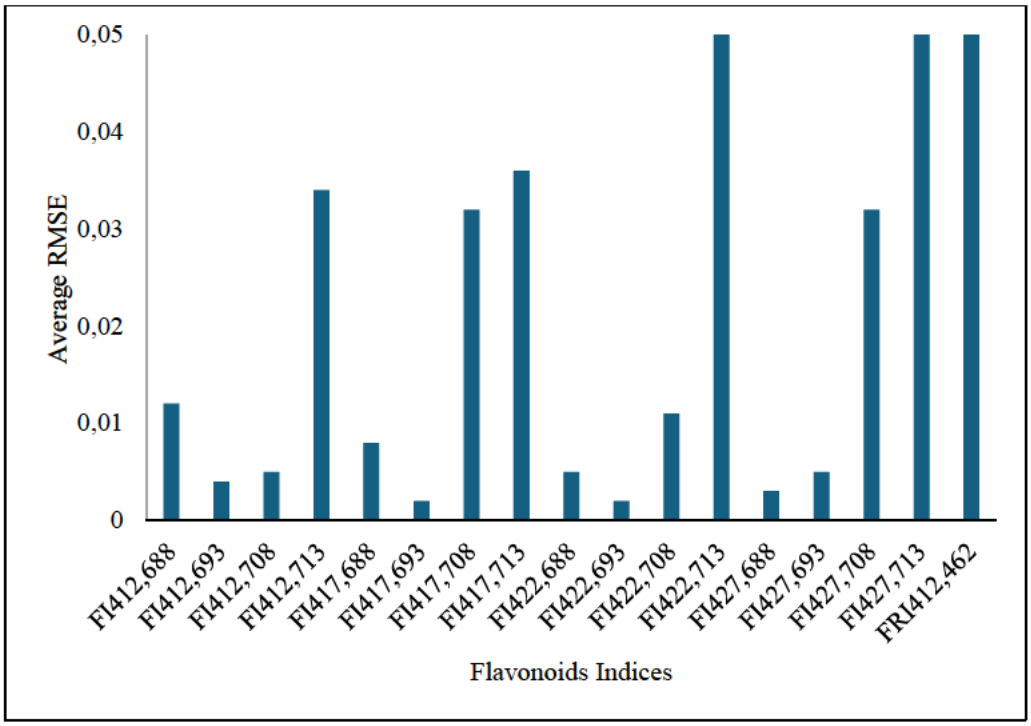


Figure 6.5. Average spectral RMSE for the various flavonoid indices tested.

6.3.2. Performance of Flavonoids indices

Table 6.3 illustrates the relationships between the simulated LAI and the measured flavonoid content, which was derived using various flavonoid indices. Given the limited number of field samples (n = 35), the models were validated using a k-fold cross-validation (e.g., k=5) approach to evaluate the robustness of the spectral indices in predicting flavonoid content. Figure 6.6 highlights the top three performing spectral indices. Among these, $FI_{417,693}$ demonstrated the highest accuracy, achieving an R^2 value of 0.63, RMSE of 0.24, and $p < 0.05$. Following this, $FI_{417,713}$ exhibited an R^2 of 0.44, an RMSE of 0.43, and $p < 0.05$. The third-best performer, $FI_{412,713}$, showed an R^2 of 0.40 and an RMSE of 0.62.

Conversely, $FI_{422,688}$ displayed the lowest predictive accuracy among the indices, with a R^2 of 0.002 and an RMSE of 1.56. Figure 6.7 presents a spatial distribution map of flavonoid content within the GCFR region, generated using the $FI_{417,693}$ spectral index derived from the PROSAIL-D model, illustrating its efficacy in delineating flavonoid variability across the landscape.

Table 6.3: Relationships between simulated LAI and measured Flavonoids derived from Flavonoids indices.

Index	R2	RMSE (mg/g)	rRMSE (%)	p-values
$FI_{412,688}$	0.16	1.50	50.69	$P < 0.05$
$FI_{412,693}$	0.27	1.23	45.07	$P < 0.05$
$FI_{412,708}$	0.32	0.67	50.23	$P < 0.05$
$FI_{412,713}$	**0.40**	**0.62**	**49.93**	$P < 0.05$
				$P < 0.05$
$FI_{417,688}$	0.04	1.70	48.76	$P < 0.05$
$FI_{417,693}$	**0.63**	**0.24**	**32.80**	$P < 0.05$
$FI_{417,708}$	0.34	0.62	54.44	$P < 0.05$
$FI_{417,713}$	**0.44**	**0.43**	**40.97**	$P < 0.05$
				$P < 0.05$
$FI_{422,688}$	0.00	1.56	54.01	$P < 0.05$
$FI_{422,693}$	0.05	1.52	58.10	$P < 0.05$
$FI_{422,708}$	0.39	0.40	40.01	$P < 0.05$
$FI_{422,713}$	0.35	0.35	36.11	$P < 0.05$
				$P < 0.05$
$FI_{427,688}$	0.02	1.67	51.82	$P < 0.05$
$FI_{427,693}$	0.01	1.50	65.00	$P < 0.05$
$FI_{427,708}$	0.31	0.29	34.10	$P < 0.05$
$FI_{427,713}$	0.35	0.24	29,47	$P < 0.05$
				$P < 0.05$
$FRI_{412,462}$	0.01	1.12	23.28	$P < 0.05$

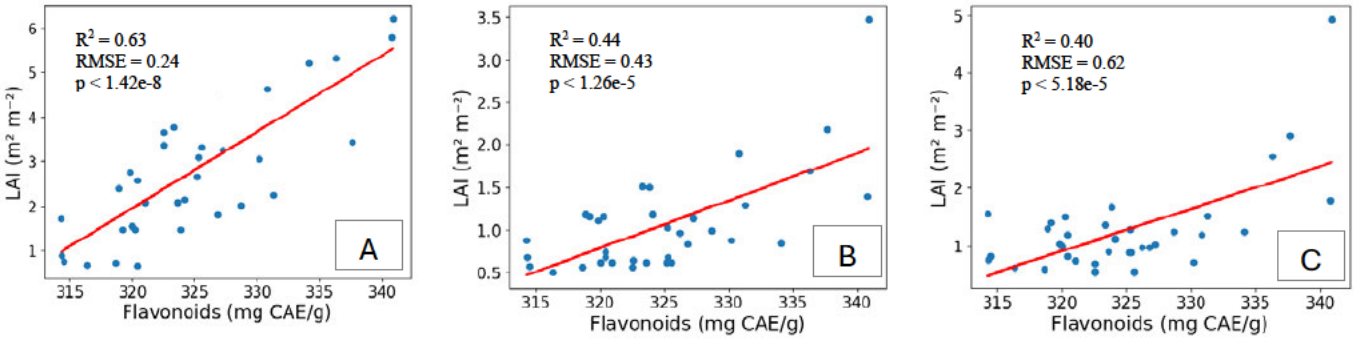


Figure 6.6. Accuracy of top three best performing Flavonoids Indices from PROSAIL-D RTM: (A) $FI_{417,693}$, (B) $FI_{417,713}$, and (C) $FI_{412,713}$.

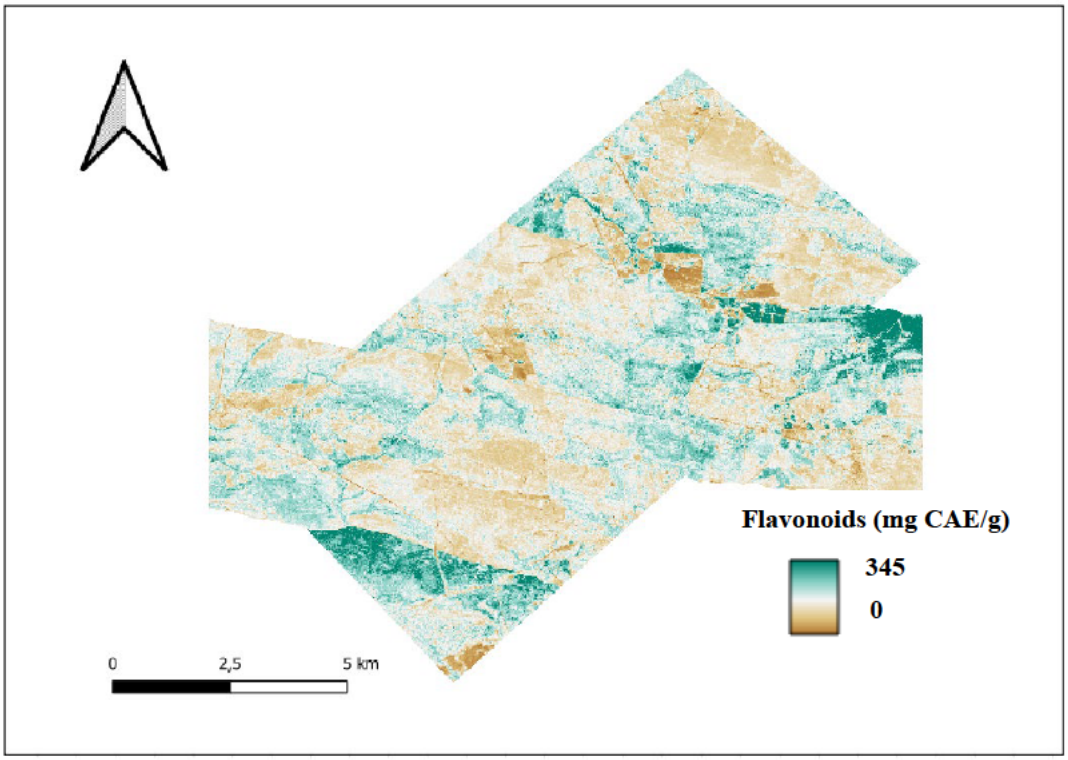


Figure 6.7. Flavonoids maps derived from PROSAIL-D- $FI_{417,693}$ index showing distribution pattern in the GCFR.

6.4. Discussion

In this study, we evaluated the performance of flavonoid estimation using spectral indices derived from AVIRIS-NG imagery. To our knowledge, this is the first attempt to map flavonoid indices at the canopy level utilizing the PROSAIL-D model and AVIRIS-NG image.

6.4.1. Impact of AVIRIS-NG Image bandwidth on flavonoid estimation

Since AVIRIS imagery does not possess the 420 nm band proposed by Zhou et al., (2022) (2023), the present study utilized all spectral regions within the 410–430 nm range, which includes spectral bands at 412, 417, 422, and 427 nm. These wavelengths are known for their sensitivity to flavonoids, and this approach aimed to maximize the detection of flavonoid signals within the available spectral resolution (Gitelson et al., 2017).

However, within this spectral region, chlorophyll and carotenoid pigments significantly influence the reflectance signal (Gitelson et al., 2017). To improve the accuracy of flavonoid estimation, it is essential to mitigate the confounding effects caused by chlorophylls and other pigments. This can be achieved by subtracting their contributions, thereby isolating the flavonoids-specific spectral signature. For wavelengths sensitive to chlorophyll, Gitelson et al. (2017) and Zhou et al. (2022) proposed utilizing spectral bands around 690 and 710 nm,

respectively. However, AVIRIS does not cover these specific wavelengths, so we focused on bands in the proximity of these regions, including 688, 693, 708, and 713 nm.

Therefore, in this study, for the bands sensitive to flavonoids, we used 412, 417, 422, and 427 nm and subtracted the contribution of chlorophyll using 688, 693, 708, and 713 nm in conjunction with the PROSAIL-D model.

6.4.2. The utility of the PROSAIL-D model

The PROSAIL-D model was selected for this study because it includes parameters related to secondary metabolites, such as carotenoids and anthocyanins (Feret et al., 2017). Notably, anthocyanins are a subgroup of flavonoids and are responsible for the pink, red, or purple hues observed in leaf tissues and play a vital role in plant physiology and adaptation (Lu et al., 2021).

Considering the environment of the GCFR, which is exceptionally biodiversity-rich with nearly 10,000 plant species (Cardoso et al., 2025), the application of PROSAIL-D becomes particularly advantageous. The region hosts a wide array of species exhibiting diverse colours and flowering phenology, and is abundant in phenolic compounds, including flavonoids (Kumar et al., 2020). In such a complex environment, a model that accounts for secondary metabolites provides a more accurate representation of optical properties and biochemical variation.

Furthermore, PROSAIL-D can simulate leaf optical properties throughout the entire plant lifecycle, covering juvenile and senescent stages as well as various environmental stresses (Feret et al., 2017). In contrast, PROSAIL-4 and PROSAIL-5 are primarily designed to simulate the optical characteristics of photosynthetic vegetation but have a limited set of parameters such as anthocyanins (Lu et al., 2021). Consequently, in an environment like the GCFR, where secondary metabolites are abundant, these models may fall short. PROSAIL-D, developed and tested in deciduous plants such as maple trees (Feret et al., 2017), has also been applied in diverse landscapes, including grasslands and mangroves (Lu et al., 2021; Zhao et al., 2023). However, it has yet to be tested in a phenolic-rich environment like the GCFR.

6.4.3. Assessment of flavonoid indices

This study tested flavonoid indices in the GCFR using the AVIRIS image and PROSAIL-D model. The results indicated that the selected spectral indices were highly sensitive for flavonoid estimation, as all tested indices achieved p-values below 0.05, demonstrating statistically significant correlations. Among these, the spectral index $FI_{417,693}$ achieved the highest accuracy, with an R^2 value of 0.63 and RMSE of 0.24, indicating a relatively strong

correlation between the spectral index and flavonoid content. The second most effective index, $FI_{417,713}$, showed an R^2 of 0.44 and an RMSE of 0.43, while $FI_{412,713}$ performed comparably with an R^2 of 0.40 and an RMSE of 0.62. These findings are consistent with previous studies that identified this spectral region as highly sensitive to flavonoids. For example, Zhou et al. (2022) reported that the 420 nm wavelength is particularly responsive to flavonoid content across various plant species. Overall, these results reinforce the importance of the 410–430 nm spectral range for effective flavonoid detection (Gitelson et al., 2017).

Although this study was conducted using a limited sample size of 35 observations, the application of 5-fold cross-validation provided robust validation of the models developed. This suggests that the spectral bands selected are reliably sensitive to flavonoid variation, even within a small dataset, by reducing potential overfitting and demonstrating that the models have good generalizability across different subsets of data. This is particularly important when considering the transferability of the PROSAIL-D model to other sites (Feret et al., 2017).

6.4.4. Factors affecting the distribution of flavonoids in the GCFR

The distribution of flavonoids is influenced by several factors, including species type, topography, canopy structure, fire occurrence, age, light availability, temperature, and nutrient levels (Zhou et al., 2023; Agati et al., 2013; Wu et al., 2020; Gitelson et al., 2017). Flavonoids are typically produced in response to both biotic and abiotic stresses, thereby enhancing plant resilience against various stressors (Chomel et al., 2016). Fynbos is characterized by heavily leached and nutrient-deficient acidic sandy soils, particularly low in nitrogen and phosphorus (Magadlela et al., 2015; Bengtsson et al., 2011). To cope with these stress factors, plants produce phenolic compounds, such as flavonoids, which help them survive in challenging environments; hence, very rich in secondary metabolites (Chomel et al., 2016).

The map showing the distribution of flavonoids indicates that higher flavonoid concentrations are found in lower elevations, next to rivers or valleys, highlighting the significant influence that topography can have on plant secondary metabolite distribution (Figure 6.7). The elevated flavonoid levels in valley regions may be attributed to variations in species composition. For example, *Berzelia lanuginosa*, commonly referred to as the common button bush or vleiknopbos, which is one of the dominet species in the study site, is typically found in habitats characterized by consistent moisture. This species thrives on damp sandstone slopes, seepages, and along stream banks, particularly within the fynbos biome of the Western Cape in South Africa (Brown et al., 2004; Jackson, W.P.U., 1990). Such environmental conditions are

conducive to the growth of specific plant species that produce higher levels of flavonoids, further illustrating the interaction between topography and plant chemistry.

Furthermore, the GCFR is characterized as a fire-driven ecosystem. Thus, fire plays a crucial role in rejuvenating vegetation. Sibiya et al. (2025, preprint) found that areas with recent fire events exhibit higher flavonoid concentrations. This increase is linked to the beneficial effects of fire on specific plant species adapted to thrive in post-fire environments. Fire is an essential ecological process within the GCFR, promoting the growth of new vegetation and enhancing biodiversity (Kraaij and van Wilgen, 2014; Smit et al., 2024). The rise in phenolic compounds may improve plants' resilience to herbivory and environmental stress, significantly contributing to the ecosystem's overall survival. This dynamic between fire and vegetation underscores the importance of fire as a regenerative force in maintaining the ecological balance and health of fynbos ecosystems (Bengtsson et al., 2011). There is an indirect relationship between species diversity and soil nutrients, such as nitrogen and NPK (nitrogen, phosphorus, and potassium), which can intensify competition among species in the fynbos, particularly during early growth stages following fire events. This heightened competition often leads to an increase in secondary metabolites, as plants adapt to thrive in conditions where they must compete more fiercely for limited resources.

Additionally, flavonoid concentration is influenced by canopy structure and light availability. Research indicates a positive relationship between flavonoid levels and LAI. This observation aligns with the findings of Zhou et al. (2023), who utilized the SAIL model along with hyperspectral data to assess flavonoid content in Ginkgo plantations. LAI reflects the total leaf surface area, which enhances light capture for photosynthesis. Increased light exposure is known to stimulate flavonoid production (Ni et al., 2020). The synthesis of flavonoids shows a strong positive correlation with light intensity, suggesting that leaves receiving more sunlight are likely to accumulate higher flavonoid levels. The fynbos ecosystem, which experiences dry conditions and intense sunlight, benefit from increased flavonoid production to resist photodamage caused by high light and UV radiation (Gitelson et al., 2017).

In summary, the interplay of these factors' topography, fire dynamics, light, and canopy structure play a critical role in shaping flavonoid distribution and abundance in plant communities.

6.4.5. Limitations and potentials

This study demonstrated that spectral indices can significantly enhance the detection of flavonoid content at the canopy level using AVIRIS imagery in conjunction with the PROSAIL-D model. The findings are promising and highlight the potential for remote sensing approaches in vegetation biochemical analysis; however, several limitations should be acknowledged to inform future research directions. Firstly, the sample size was limited to 35 field samples, which may constrain the statistical power and generalizability of the models. Expanding the number of samples would improve model robustness and enable more refined calibration and validation procedures. Moreover, it is worth noting that the high R^2 value of 0.63 achieved in this analysis might be somewhat optimistic given the small sample size. Such a relatively modest sample can introduce potential biases, as the conclusions drawn from it may not accurately represent broader trends and the risk of spatial autocorrelation.

Additionally, the current study only captures a single season, thus limiting insights into temporal dynamics. Flavonoid content can vary significantly across different seasons and phenological stages, and future studies should explore seasonal variations to better understand the temporal stability and variability of spectral-flavonoid relationships.

Another notable limitation pertains to the spectral capabilities of AVIRIS, which lacks wavelength bands such as 420 nm. The absence of this critical wavelength may hinder the precision of flavonoid estimation, as some flavonoid absorption features at shorter wavelengths could be missed. To address this, future work could utilize hyperspectral data from sensors such as EnMAP, which offer 420 spectral bands, thereby improving the spectral resolution and potentially enhancing the accuracy of flavonoid detection (Locherer et al., 2015).

This study only explored flavonoid-related indices. Future studies should also explore FI normalized by structure indices and FI normalized by chlorophyll indices proposed by Zhou et al. (2023). Furthermore, this research employed PROSAIL-D coupled with a LUT as the cost function for retrieval. Future work should therefore explore the synergy between machine learning algorithms and radiative transfer models like PROSIL to develop more precise and adaptable tools for remote sensing-based flavonoid detection. Such advancements would significantly contribute to the development of rapid, non-destructive monitoring methods for plant health and phytochemical content.

6.5. Conclusion

In this study, we utilized spectral indices to enhance the sensitivity of Flavonoids using AVIRIS-NG image and PROSAIL-D model. The key findings are as follows:

- Among these, $FI_{417,693}$ demonstrated the highest accuracy, followed by $FI_{417,713}$ and the third-best performer was $FI_{412,713}$.
- Conversely, $FI_{422,688}$ displayed the lowest predictive accuracy among the indices.

To our knowledge, this study is the first to employ AVIRIS imagery in conjunction with the PROSAIL RTM to estimate Flavonoids content through vegetation indices. The results underscore the significant potential of combining AVIRIS data with PROSAIL for non-destructive, remote sensing-based assessment of Flavonoids content at the canopy level. This approach could pave the way for more accurate and efficient monitoring of plant phytochemical traits, with broad applications in agriculture, ecology, and plant health management.

6.6. Chapter summary

This chapter focused on estimating flavonoid content using PROSAIL-D and AVIRIS hyperspectral data. Because chlorophyll influences the spectral sensitivity around 410 to 430 nm, we targeted specific wavelengths to develop spectral indices optimized for flavonoid detection. The findings showed that the spectral index FI417,693 provided the highest accuracy in predicting flavonoid content, closely followed by FI417,713. These results highlight the importance of narrow spectral bands for reliable flavonoid retrieval and demonstrate the potential of targeted spectral indices to improve remote sensing estimates of secondary metabolites in complex vegetation canopies. The final chapter summarizes all the research findings, discusses their implications, and proposes recommendations for future studies, emphasizing the significance of the conclusions drawn in this thesis.

Chapter Seven:
Synthesis

7.1. Introduction

The primary objective of this study was to retrieve plant secondary metabolites across diverse biomes using remotely sensed data in conjunction with RTMs. The results demonstrate that the application of RTMs such as PROSPECT and PROSAIL, when integrated with hyperspectral data, significantly enhances the accuracy of secondary metabolite estimations, particularly at large spatial scales. A literature review indicates PROSPECT and PROSAIL are the most widely used models over the past two decades. The main variables retrieved are LAI and chlorophyll content, often using LUT and ANNs. Applications focus on grassland and forest ecosystems, commonly using MODIS and Landsat imagery. RTMs like PROSPECT and PROSAIL have demonstrated success with multispectral data. However, their use with hyperspectral data is still limited and needs more refinement. While these models succeed with multispectral data, their use with hyperspectral data is limited and needs improvement, especially with new sensors like EnMap and PRISMA. Furthermore, there is a pressing need to evaluate and adapt these models for environments rich in phenolic compounds.

To address this, we used the PROSPECT-5 model to estimate carotenoid content in three datasets (savanna, tropical forest, and combined). We found that the green wavelength region (510–580 nm) gave the best performance across all datasets. This was followed by spectral bands selected via a random forest algorithm (i.e., green, yellow, and red-edge regions). Notably, the visible spectrum (400–700 nm) also exhibited strong predictive capability. These findings highlight the critical importance of spectral band selection in RTM inversion, suggesting that restricting analysis to specific, physiologically relevant spectral regions can improve both the accuracy and efficiency of carotenoid estimation, rather than employing the entire spectral range.

In a different biome (Fynbos biome), We estimated leaf-level phenolic compounds in relation to fire occurrence. We found that burned areas (2015, 2016, and 2021) had higher concentrations of phenolics than unburned areas. The spectral indices $FI_{415,710}$ and $FI_{420,710}$ had the strongest correlations with flavonoid content. These results highlight the crucial role of fire and phenolic compounds in the survival of fynbos biomes. Also, the importance of targeted spectral indices for identifying phenolic compounds associated with fire history and stress adaptation in phenolic-rich ecosystems.

Flavonoids (Chapters 4 and 6) and carotenoids (Chapter 3) exhibit distinct light absorption properties that are essential for interpreting plant physiological status and for understanding pigment interactions within the visible spectral region. In general, flavonoids show relatively strong absorption in the blue domain, with sensitivity in the shortwave blue region (approximately 410–430 nm), and typically absorb only weakly in the red (Gitelson et al., 2017). Carotenoids also absorb strongly in the blue region, commonly spanning approximately 400–500 nm (Huang et al., 2018). Chlorophyll, in contrast, has broader absorption features in both the blue and red, while reflecting more in the green wavelengths (Zhou et al., 2022). Because chlorophyll is often the dominant pigment, its spectral signature strongly shapes canopy reflectance and can mask or confound the signals attributable to less abundant pigments.

This overlap is less straightforward to disentangle in reflectance spectra because chlorophyll's relatively strong reflectance in the green wavelengths can obscure or dilute pigment-related signals, reducing the separability of individual pigments when only raw reflectance is considered. In our study, we found that the 415–420 nm region and nearby wavelengths were particularly sensitive to flavonoids, whereas the 510–580 nm interval showed a high correlation with carotenoid presence. Notably, the strong reflectance in the green region, dominated by chlorophyll reflectance, co-occurred with carotenoid retrieval. This indicates that even when chlorophyll controls a large portion of the visible signal, carotenoids can still be detected because their effects alter the overall spectral shape through overlapping absorption features and subsequent changes in canopy reflectance.

These findings highlight an important trade-off in pigment retrieval: increasing chlorophyll abundance can enhance green reflectance, but it can simultaneously complicate the detection of other pigments by masking their spectral signatures or by introducing parameter coupling in the inversion process. For this reason, numerous studies have developed carotenoid indices intended to minimize chlorophyll interference (Chappelle et al., 1992; Datt, 1998; Blackburn, 1998; Gitelson et al., 2002, 2006; Hernández-Clemente et al., 2012). Simmilary, flavonoid-focused studies have proposed indices designed to reduce chlorophyll influence (Zhou et al., 2022; 2023). Consistent with this rationale, our methodology incorporated specific flavonoid indices to better isolate flavonoid contributions from the chlorophyll-dominated background.

Moving from the leaf to canopy level, we evaluated whether the inclusion of secondary metabolites within the PROSAIL model enhances the estimation of PAI in phenolic-rich

landscapes, such as the GCFR. The findings indicate that the PROSAIL-D model, which explicitly accounts for secondary metabolites such as carotenoids and anthocyanins, outperformed the standard PROSAIL-4 model that does not include these compounds. This result underscores the importance of incorporating secondary metabolites when modelling in phenolic-rich environments, as it improves the accuracy of PAI estimates and highlights the need for models tailored to ecosystems with high phenolic content.

Expanding from Chapter 5, we utilized the PROSAIL-D model, which explicitly accounts for secondary metabolites, we estimated flavonoid content through the use of vegetation indices. The results indicate that the spectral index FI417,693 achieved the highest accuracy in predicting flavonoid levels, closely followed by FI417,713. These findings highlight the critical importance of narrow spectral bands for reliable flavonoid retrieval. The sensitivity of these bands is influenced by chlorophyll absorption in the visible region, which underscores the necessity of selecting specific spectral wavelengths to improve the precision of secondary metabolite estimation. In summary, methodologically, the RTM inversion proved more robust in the Tropical Forest dataset (Chapter 3) due to canopy homogeneity, whereas the Fynbos application (Chapter 5 & 6) required the inclusion of specific pigment parameters to overcome the complex spectral mixing of sclerophyllous vegetation.

7.2. Challenges and limitations

- Although carotenoid estimation has shown promising results at the leaf level, in the current study, carotenoid retrieval was not conducted at the canopy level, primarily due to the limitations of available canopy hyperspectral data.
- The sample size for flavonoid and PAI estimation was limited to 35 field samples, which constrains the statistical power of the models and reduces their generalizability. A larger, more diverse sampling effort would improve the robustness of the predictive models, facilitate more precise calibration, and enable more rigorous validation.
- Although we acknowledge the limitations associated with the small sample size, the implementation of k-fold cross-validation in Chapter 5 and 6 served as a vital safeguard against overfitting. This technique, which involved partitioning the available data into subsets to ensure that each sample was utilized for both training and validation, helped mitigate the risks of model overfitting, thereby enhancing the reliability of the model estimates. Nonetheless, it is essential to note that while k-fold cross-validation provides an additional layer of validation, it cannot entirely compensate for the limitations

presented by the small sample size. Expanding the number of samples in future studies would not only bolster the models' predictive accuracy but also improve their applicability to a wider range of environmental conditions and plant species.

- The current study provides a snapshot of carotenoid, flavonoids, and PAI spectral properties within a single season, thereby limiting insights into their temporal variability. Seasonal dynamics, driven by phenological stages, environmental conditions, and stress factors, play a crucial role in pigment composition and, therefore, should be examined through multi-temporal datasets to capture year-round variability and improve the predictive capabilities of remote sensing models.
- The AVIRIS sensor used in this study lacked some key wavelengths critical for flavonoid estimation, notably around 420 nm, a spectral region known for its sensitivity to flavonoid absorption features. The absence of this wavelength limits the effectiveness of current spectral indices and models in accurately capturing flavonoid variations, emphasizing a need for future sensor designs or data sources that include this important spectral band.
- The study focused on three distinct biomes including savanna, forest, and fynbos, each characterized by unique vegetation structures and pigment compositions. However, other ecosystem types, such as grasslands and temperate meadows, warrant attention to broaden understanding of pigment variability across different ecological contexts. Including such regions would enhance model transferability and contribute to a more comprehensive understanding of pigment dynamics across diverse landscapes.

7.3. Recommendation and prospects

This research represents a significant step forward in leveraging remote sensing and radiative transfer models for the characterization of plant secondary metabolites across diverse biomes. The findings highlight the value of RTMs for estimating key biochemical and biophysical properties and provide critical insights for ecological monitoring, conservation, and land management, especially in the face of environmental change. To further advance this field, several key recommendations and prospects emerge:

- Future studies should prioritize upscaling carotenoid estimates to the canopy level. This necessitates addressing complexities in canopy structure by utilizing PROSAIL model. Lidar integration can offer superior high-resolution structural parameters (e.g., canopy height, biomass), to create better models.

- While hyperspectral data has proven invaluable in providing detailed spectral information for the identification and quantification of secondary metabolites, their high cost and processing requirements may limit their application in some contexts. Future research should explore the utility of multispectral sensors such as Sentinel-2. The development of new, specifically tailored spectral indices designed for multispectral data will be crucial for maximizing the utility of these readily accessible datasets.

- Furthermore, new as well as the planned hyperspectral missions, including EnMAP, PRISMA, FLEX, and HyspIRI, will provide even more detailed environmental data (Verrelst et al., 2016). Moreover, advancements in airborne platforms, particularly unmanned aerial vehicles (UAVs), have expanded the range of remotely sensed data available for analysing vegetation (Kükenbrink et al., 2021; Yin et al., 2022). This wide variety of data necessitate the development of improved processing techniques that are accurate, reliable, and efficient.

- Future studies should focus on multi-temporal datasets to capture year-round variability and improve the predictive capabilities of remote sensing models. Detailed analysis of phenological stages (e.g., timing of leaf flush, flowering, senescence) using high temporal resolution data is needed to account for pigment changes at different points during the growing season. Long-term ecological monitoring sites can facilitate the collection of phenological data, which can then be integrated with remotely sensed data to improve model accuracy.

- Future work should integrate machine learning (e.g. active learning) with radiative transfer models like PROSAIL to improve secondary metabolite mapping. Active learning, in particular, can help address outliers and non-informative points in RTM-simulated data.

- Future work should use EnMap when estimating flavonoids since it has 420nm wavelength which is important for flavonoids.

References

Adeline, K., Roth, K., Huesca, M., Gastellu-Etchegorry, J.P., Baldocchi, D. and Ustin, S., 2016, August. Spectral sensitivity of radiative transfer inversion for seasonal canopy pigments estimation from aviris data in a woodland savanna ecosystem. In *2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)* (pp. 1-5). IEEE.

Agati, G., Brunetti, C., Di Ferdinando, M., Ferrini, F., Pollastri, S. and Tattini, M., 2013. Functional roles of flavonoids in photoprotection: new evidence, lessons from the past. *Plant Physiology and Biochemistry*, *72*, pp.35-45.

Agati, G., Tuccio, L., Kusznierewicz, B., Chmiel, T., Bartoszek, A., Kowalski, A., Grzegorzewska, M., Kosson, R. and Kaniszewski, S., 2016. Nondestructive optical sensing of flavonoidsonols and chlorophyll in white head cabbage (Brassica oleracea L. var. capitata subvar. alba) grown under different nitrogen regimens. *Journal of Agricultural and Food Chemistry*, *64*(1), pp.85-94.

Ahmad, S., Pandey, A.C., Kumar, A. and Lele, N.V., 2021. Potential of hyperspectral AVIRIS-NG data for vegetation characterization, species spectral separability, and mapping. *Applied Geomatics*, *13*(3), pp.361-372.

Ali, A.M., Darvishzadeh, R., Skidmore, A.K., van Duren, I., Heiden, U. and Heurich, M., 2016. Estimating leaf functional traits by inversion of PROSPECT: Assessing leaf dry matter content and specific leaf area in mixed mountainous forest. *International Journal of Applied Earth Observation and Geoinformation*, *45*, pp.66-76.

Ali, A.M., Skidmore, A.K., Darvishzadeh, R., van Duren, I., Holzwarth, S. and Mueller, J., 2016. Retrieval of forest leaf functional traits from HySpex imagery using radiative transfer models and continuous wavelet analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, *122*, pp.68-80.

Ali, A.M., Darvishzadeh, R., Skidmore, A., Heurich, M., Paganini, M., Heiden, U. and Mücher, S., 2020. Evaluating prediction models for mapping canopy chlorophyll content across biomes. *Remote Sensing*, *12*(11), p.1788.

Ali, A.M., Darvishzadeh, R., Skidmore, A., Gara, T.W. and Heurich, M., 2021. Machine learning methods' performance in radiative transfer model inversion to retrieve plant traits from Sentinel-2 data of a mixed mountain forest. *International Journal of Digital Earth*, *14*(1), pp.106-120.

ASD Analytical Spectral Devices. 2016. FieldSpec4 User Manual. A PANalytical Company. Boulder, CO, USA: Analytical Spectral Devices.

Ashraf, M.A., Iqbal, M., Rasheed, R., Hussain, I., Riaz, M. and Arif, M.S., 2018. Environmental stress and secondary metabolites in plants: an overview. *Plant Metabolites And Regulation Under Environmental Stress*, pp.153-167.

Asner, G.P., Palace, M., Keller, M., Pereira Jr, R., Silva, J.N. and Zweede, J.C., 2002. Estimating canopy structure in an Amazon forest from laser range finder and IKONOS satellite observations 1. *Biotropica*, *34*(4), pp.483-492.

Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. *Remote sensing of Environment*, *64*(3), pp.234-253.

Arellano, P., Tansey, K. and Balzter, H., 2016. Detection of Amazon Forest Degradation Caused by Land Use Changes. *Tropical Forests-The Challenges of Maintaining Ecosystem Services while Managing the Landscape*, pp.51-77.

Arellano, P., Tansey, K., Balzter, H. and Boyd, D.S., 2017. Field spectroscopy and radiative transfer modelling to assess impacts of petroleum pollution on biophysical and biochemical parameters of the Amazon rainforest. *Environmental Earth Sciences*, *76*, pp.1-14a.

Arellano, P., Tansey, K., Balzter, H. and Tellkamp, M., 2017. Plant family-specific impacts of petroleum pollution on biodiversity and leaf chlorophyll content in the Amazon rainforest of Ecuador. *PLOS ONE*, *12*(1), p.e0169867b.

Atzberger, C., Jarmer, T., Schlerf, M., Kötz, B. and Werner, W., 2003, May. Retrieval of wheat bio-physical attributes from hyperspectral data and SAILH+ PROSPECT radiative transfer model. In *Proceedings of the 3rd EARSeL Workshop on Imaging Spectroscopy* (Vol. 3, pp. 473-482). Citeseer.

Atzberger, C., Darvishzadeh, R., Schlerf, M. and Le Maire, G., 2013. Suitability and adaptation of PROSAIL radiative transfer model for hyperspectral grassland studies. *Remote Sensing Letters*, *4*(1), pp.55-64.

Atzberger, C., Darvishzadeh, R., Immitzer, M., Schlerf, M., Skidmore, A. and Le Maire, G., 2015. Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, *43*, pp.19-31.

Bailey, B.N., Overby, M., Willemsen, P., Pardyjak, E.R., Mahaffee, W.F. and Stoll, R., 2014. A scalable plant-resolving radiative transfer model based on optimized GPU ray tracing. *Agricultural And Forest Meteorology*, *198*, pp.192-208.

Banskota, A., Serbin, S.P., Wynne, R.H., Thomas, V.A., Falkowski, M.J., Kayastha, N., Gastellu-Etchegorry, J.P. and Townsend, P.A., 2015. An LUT-based inversion of DART model to estimate forest LAI from hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(6), pp.3147-3160.

Barbehenn, R.V. and Constabel, C.P., 2011. Tannins in plant–herbivore interactions. *Phytochemistry*, *72*(13), pp.1551-1565.

Baret, F. and Buis, S., 2008. Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems. *Advances In Land Remote Sensing: System, Modeling, Inversion and Application*, pp.173-201.

Bartwal, A., Mall, R., Lohani, P., Guru, S.K. and Arora, S., 2013. Role of secondary metabolites and brassinosteroids in plant defense against environmental stresses. *Journal Of Plant Growth Regulation*, *32*(1), pp.216-232.

Bayat, B., van der Tol, C. and Verhoef, W., 2018. Integrating satellite optical and thermal infrared observations for improving daily ecosystem functioning estimations during a drought episode. *Remote Sensing of Environment*, *209*, pp.375-394.

Bayat, B., van der Tol, C. and Verhoef, W., 2020. Retrieval of land surface properties from an annual time series of Landsat TOA radiances during a drought episode using coupled radiative transfer models. *Remote Sensing of Environment*, *238*, p.110917.

Belsky, A.J., 1992. Effects of grazing, competition, disturbance and fire on species composition and diversity in grassland communities. *Journal Of Vegetation Science*, *3*(2), pp.187-200.

Bengtsson, J., Janion, C., Chown, S.L. and Leinaas, H.P., 2011. Variation in decomposition rates in the fynbos biome, South Africa: the role of plant species and plant stoichiometry. *Oecologia*, *165*(1), pp.225-235.

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F. and Hank, T., 2018. Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: A review study. *Remote Sensing*, *10*(1), p.85.

Berni, J.A.J., Kljun, N., Van Gorsel, E., Haverd, V., Leuning, R., Cabello-Leblic, A., Held, A., Hopkinson, C., Chasmer, L. and Youngentob, K., 2011. 3D spatial distribution of biophysical parameters derived from hyperspectral and lidar remote sensing. Improving the constraints in land surface modelling. In *34th International Symposium on Remote Sensing of Environment-The GEOSS Era: Towards Operational Environmental Monitoring*.

Binh, N.A., Hauser, L.T., Viet Hoa, P., Thi Phuong Thao, G., An, N.N., Nhut, H.S., Phuong, T.A. and Verrelst, J., 2022. Quantifying mangrove leaf area index from Sentinel-2 imagery using hybrid models and active learning. *International Journal of Remote Sensing*, *43*(15-16), pp.5636-5657.

Biriukova, K., Celesti, M., Evdokimov, A., Pacheco-Labrador, J., Julitta, T., Migliavacca, M., Giardino, C., Miglietta, F., Colombo, R., Panigada, C. and Rossini, M., 2020. Effects of varying solar-view geometry and canopy structure on solar-induced chlorophyll fluorescence and PRI. *International Journal of Applied Earth Observation and Geoinformation*, *89*, p.102069.

Bhattacharya, B.K., Green, R.O., Rao, S., Saxena, M., Sharma, S., Kumar, K.A., Srinivasulu, P., Sharma, S., Dhar, D., Bandyopadhyay, S. and Bhatwadekar, S., 2019. An overview of AVIRIS-NG airborne hyperspectral science campaign over India. *Current Science*, *116*(7), pp.1082-1088.

Blackburn, G.A., 1998. Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. *International Journal of Remote Sensing*, *19*(4), pp.657-675.

Blackburn, G.A., 2007. Hyperspectral remote sensing of plant pigments. *Journal Of Experimental Botany*, *58*(4), pp.855-867.

Bond, W.J., Woodward, F.I. and Midgley, G.F., 2005. The global distribution of ecosystems in a world without fire. *New Phytologist*, *165*(2), pp.525-538.

Breiman, L., 2001. Random forests. *Machine Learning*, *45*, pp.5-32.

Britton, G., Liaaen-Jensen, S. and Pfander, H. eds., 2004. *Carotenoids: Handbook*. Springer Science & Business Media.

Broge, N.H. and Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, *76*(2), pp.156-172.

Brown, N.A.C., Botha, P.A. and Van Staden, J., 2004. Smoke seed germination studies and a guide to seed propagation of plants from the major families of the Cape Floristic Region, South Africa. *South African Journal of Botany*, *70*(4), pp.559-581.

Bye, I.J., North, P.R., Los, S.O., Kljun, N., Rosette, J.A.B., Hopkinson, C., Chasmer, L. and Mahoney, C., 2017. Estimating forest canopy parameters from satellite waveform LiDAR by inversion of the FLIGHT three-dimensional radiative transfer model. *Remote Sensing of Environment*, *188*, pp.177-189.

Campbell, P.E., Rock, B.N., Martin, M.E., Neefus, C.D., Irons, J.R., Middleton, E.M. and Albrechtova, J., 2004. Detection of initial damage in Norway spruce canopies using hyperspectral airborne data. *International Journal of Remote Sensing*, *25*(24), pp.5557-5584.

Campos-Taberner, M., García-Haro, F.J., Confalonieri, R., Martínez, B., Moreno, Á., Sánchez-Ruiz, S., Gilabert, M.A., Camacho, F., Boschetti, M. and Busetto, L., 2016. Multitemporal monitoring of plant area index in the Valencia rice district with PocketLAI. *Remote Sensing*, *8*(3), p.202.

Cardoso, A.W., Hestir, E.L., Slingsby, J.A., Forbes, C.J., Moncrieff, G.R., Turner, W., Skowno, A.L., Nesslage, J., Brodrick, P.G., Gaddis, K.D. and Wilson, A.M., 2025. The biodiversity survey of the Cape (BioSCape), integrating remote sensing with biodiversity science. *npj Biodiversity*, *4*(1), p.2.

Casas, A., Riaño, D., Ustin, S.L., Dennison, P. and Salas, J., 2014. Estimation of water-related biochemical and biophysical vegetation properties using multitemporal airborne hyperspectral data and its comparison to MODIS spectral response. *Remote Sensing of Environment*, *148*, pp.28-41.

Cerasoli, S., Campagnolo, M., Faria, J., Nogueira, C. and Caldeira, M.D.C., 2018. On estimating the gross primary productivity of Mediterranean grasslands under different fertilization regimes using vegetation indices and hyperspectral reflectance. *Biogeosciences*, *15*(17), pp.5455-5471.

Chapin III, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E. and Mack, M.C., 2000. Consequences of changing biodiversity. *Nature*, *405*(6783), pp.234-242.

Chappelle, E.W., Kim, M.S. and McMurtrey III, J.E., 1992. Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. *Remote sensing of environment*, *39*(3), pp.239-247.

Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).

Chen, L., Zhang, Y., Nunes, M.H., Stoddart, J., Khoury, S., Chan, A.H. and Coomes, D.A., 2022. Predicting leaf traits of temperate broadleaf deciduous trees from hyperspectral reflectance: can a general model be applied across a growing season?. *Remote Sensing of Environment*, *269*, p.112767.

Cheng, Y.B., Zarco-Tejada, P.J., Riaño, D., Rueda, C.A. and Ustin, S.L., 2006. Estimating vegetation water content with hyperspectral data for different canopy scenarios: Relationships between AVIRIS and MODIS indexes. *Remote Sensing of Environment*, *105*(4), pp.354-366.

Cho, M.A., Mathieu, R., Asner, G.P., Naidoo, L., Van Aardt, J.A.N., Ramoelo, A., Debba, P., Wessels, K., Main, R., Smit, I.P. and Erasmus, B., 2012. Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. *Remote Sensing of Environment*, *125*, pp.214-226.

Cho, M.A., Ramoelo, A. and Skidmore, A., 2014, July. Exploring various spectral regions for estimating chlorophyll from ASD leaf reflectance using prospect radiative transfer model. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 4754-4757). IEEE.a

Cho, M.A., Ramoelo, A. and Math, R., 2014, July. Estimation of leaf area index (LAI) of South Africa from MODIS imager by inversion of PROSAIL radiative transfer model. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 2590-2593). IEEE.b

Cho, M.A., Malahlela, O. and Ramoelo, A., 2015. Assessing the utility WorldView-2 imagery for tree species mapping in South African subtropical humid forest and the conservation implications: Dukuduku forest patch as case study. *International Journal of Applied Earth Observation and Geoinformation*, *38*, pp.349-357.

Chomel, M., Guittonny-Larchevêque, M., Fernandez, C., Gallet, C., DesRochers, A., Paré, D., Jackson, B.G. and Baldy, V., 2016. Plant secondary metabolites: a key driver of litter decomposition and soil nutrient cycling. *Journal of Ecology*, *104*(6), pp.1527-1541.

City of Cape Town. City of Cape Town Biodiversity Report 2018; City of Cape Town: Cape Town, South Africa, 2018.

Clark, R.N. and Roush, T.L., 1984. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth*, *89*(B7), pp.6329-6340.

Clevers, J.G. and Kooistra, L., 2011. Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(2), pp.574-583.

Close, D.C. and McArthur, C., 2002. Rethinking the role of many plant phenolics–protection from photodamage not herbivores? *Oikos*, *99*(1), pp.166-172.

Collins, S.L., 1990. Effects of fire on community structure in tallgrass and mixed-grass prairie. *Fire In North American Tallgrass Prairies*, pp.81-98.

Coops, N.C., Tompalski, P., Goodbody, T.R., Queinnec, M., Luther, J.E., Bolton, D.K., White, J.C., Wulder, M.A., van Lier, O.R. and Hermosilla, T., 2021. Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends. *Remote Sensing of Environment*, *260*, p.112477.

Croft, H., Chen, J.M., Zhang, Y. and Simic, A., 2013. Modelling leaf chlorophyll content in broadleaf and needle leaf canopies from ground, CASI, Landsat TM 5 and MERIS reflectance data. *Remote Sensing of Environment*, *133*, pp.128-140.

Dai, E., Huang, Y., Wu, Z. and Zhao, D., 2016. Analysis of spatio-temporal features of a carbon source/sink and its relationship to climatic factors in the Inner Mongolia grassland ecosystem. *Journal of Geographical Sciences*, *26*, pp.297-312.

Danner, M., Berger, K., Wocher, M., Mauser, W. and Hank, T., 2017. Retrieval of biophysical crop variables from multi-angular canopy spectroscopy. *Remote Sensing*, *9*(7), p.726.

Darvishzadeh, R., Skidmore, A., Schlerf, M. and Atzberger, C., 2008. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment*, *112*(5), pp.2592-2604.a

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F. and Cho, M., 2008. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*(4), pp.409-426.b

Darvishzadeh, R., Atzberger, C., Skidmore, A. and Schlerf, M., 2010. *Retrieval Of Vegetation Biochemicals Using a Radiative Transfer Model and Hyperspectral Data*. na.

Darvishzadeh, R., Atzberger, C., Skidmore, A. and Schlerf, M., 2011. Mapping grassland leaf area index with airborne hyperspectral imagery: A comparison study of statistical approaches and inversion of radiative transfer models. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(6), pp.894-906.

Darvishzadeh, R., Wang, T., Skidmore, A., Vrieling, A., O'Connor, B., Gara, T.W., Ens, B.J. and Paganini, M., 2019. Analysis of Sentinel-2 and RapidEye for retrieval of leaf area index in a saltmarsh using a radiative transfer model. *Remote Sensing*, *11*(6), p.671.

Darvishzadeh, R., Skidmore, A., Abdullah, H., Cherenet, E., Ali, A., Wang, T., Nieuwenhuis, W., Heurich, M., Vrieling, A., O'Connor, B. and Paganini, M., 2019. Mapping leaf chlorophyll content from Sentinel-2 and RapidEye data in spruce stands using the invertible forest reflectance model. *International Journal of Applied Earth Observation and Geoinformation*, *79*, pp.58-70.

Dasgupta, S., Qu, J.J. and Bhoi, S., 2009. Constrained radiative transfer inversions for vegetation moisture retrievals in grasslands. *Journal of Applied Remote Sensing*, *3*(1), p.031503.

Datt, B., 1998. Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+ b, and total carotenoid content in eucalyptus leaves. *Remote Sensing of Environment*, *66*(2), pp.111-121.

Dehghan-Shoar, M.H., Orsi, A.A., Pullanagari, R.R. and Yule, I.J., 2023. A hybrid model to predict nitrogen concentration in heterogeneous grassland using field spectroscopy. *Remote Sensing of Environment*, *285*, p.113385a.

Dehghan-Shoar, M.H., Pullanagari, R.R., Orsi, A.A. and Yule, I.J., 2023. Simulating spaceborne imaging to retrieve grassland nitrogen concentration. *Remote Sensing Applications: Society and Environment*, *29*, p.100912b.

De Klerk, H., 2008. A pragmatic assessment of the usefulness of the MODIS (Terra and Aqua) 1-km active fire (MOD14A2 and MYD14A2) products for mapping fires in the fynbos biome. *International Journal of Wildland Fire*, *17*(2), pp.166-178.

Dong, Z.B., Li, S.P., Hong, M. and Zhu, Q., 2005. Hypothesis of potential active components in Angelica sinensis by using biomembrane extraction and high-performance liquid chromatography. *Journal of Pharmaceutical and Biomedical Analysis*, *38*(4), pp.664-669.

Döpper, V., Rocha, A.D., Berger, K., Gränzig, T., Verrelst, J., Kleinschmit, B. and Förster, M., 2022. Estimating soil moisture content under grassland with hyperspectral data using radiative transfer modelling and machine learning. *International Journal of Applied Earth Observation and Geoinformation*, *110*, p.102817.

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J., Brazile, J., Singh, R. and Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal Of Applied Earth Observation And Geoinformation*, *9*(2), pp.165-193.

Dorigo, W., Richter, R., Baret, F., Bamler, R. and Wagner, W., 2009. Enhanced automated canopy characterization from hyperspectral data by a novel two step radiative transfer model inversion approach. *Remote Sensing*, *1*(4), pp.1139-1170.

Du, K., Huang, H., Feng, Z., Hakala, T., Chen, Y. and Hyyppä, J., 2021. Using Microwave Profile Radar to Estimate Forest Canopy Leaf Area Index: Linking 3D Radiative Transfer Model and Forest Gap Model. *Remote Sensing*, *13*(2), p.297.

Duan, S.B., Li, Z.L., Wu, H., Tang, B.H., Ma, L., Zhao, E. and Li, C., 2014. Inversion of the PROSAIL model to estimate leaf area index of maize, potato, and sunflower fields from unmanned aerial vehicle hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, *26*, pp.12-20.

Duan, P., Wang, Y. and Yin, P., 2020. Remote sensing applications in monitoring of protected areas: A bibliometric analysis. *Remote sensing*, *12*(5), p.772.

Dusseux, P., Hubert-Moy, L., Lecerf, R., Gong, X. and Corpetti, T., 2011, July. Identification of grazed and mown grasslands using a time series of high-spatial-resolution remote sensing images. In *2011 6th International Workshop on The Analysis of Multi-Temporal Remote Sensing Images (Multi-Temp)* (pp. 145-148). IEEE.

Farmonov, N., Walden, S., Martinée, E., Lampei, C., Schreiber, M., Opgenoorth, L., Rakotomalala, A.A.N.A., Müller, T., Farwig, N., Pinkert, S. and Saueressig, L., 2025. Optimizing hybrid models for forest leaf and canopy trait mapping from EnMAP hyperspectral data with limited field samples. *Science of Remote Sensing*, p.100253.

Fang, H. and Liang, S., 2005. A hybrid inversion method for mapping leaf area index from MODIS data: Experiments and application to broadleaf and needleleaf canopies. *Remote Sensing of Environment*, *94*(3), pp.405-424.

Feret, J.B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P., Ustin, S.L., Le Maire, G. and Jacquemoud, S., 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment*, *112*(6), pp.3030-3043.

Féret, J.B., François, C., Gitelson, A., Asner, G.P., Barry, K.M., Panigada, C., Richardson, A.D. and Jacquemoud, S., 2011. Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling. *Remote Sensing of Environment*, *115*(10), pp.2742-2750.

Féret, J.B., Gitelson, A.A., Noble, S.D. and Jacquemoud, S.J.R.S.O.E., 2017. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sensing of Environment*, *193*, pp.204-215.

Féret, J.B., Le Maire, G., Jay, S., Berveiller, D., Bendoula, R., Hmimina, G., Cheraiet, A., Oliveira, J.C., Ponzoni, F.J., Solanki, T. and De Boissieu, F., 2019. Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning. *Remote Sensing of Environment*, *231*, p.110959.\

Féret, J.B., Berger, K., De Boissieu, F. and Malenovský, Z., 2021. PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. *Remote Sensing of Environment*, *252*, p.112173.

Fernández-Guisuraga, J.M., Suárez-Seoane, S., Quintano, C., Fernández-Manso, A. and Calvo, L., 2022. Comparison of physical-based models to measure forest resilience to fire as a function of burn severity. *Remote Sensing*, *14*(20), p.5138.

Gao, S., Huete, A., Kobayashi, H., Doody, T.M., Liu, W., Wang, Y., Zhang, Y. and Lu, X., 2022. Simulation of solar-induced chlorophyll fluorescence in a heterogeneous forest using 3-D radiative transfer modelling and airborne LiDAR. *ISPRS Journal of Photogrammetry and Remote Sensing*, *191*, pp.1-17.

Gamon, J., Serrano, L. and Surfus, J.S., 1997. The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia*, *112*(4), pp.492-501.

Gascon, F., Gastellu-Etchegorry, J.P., Lefevre-Fonollosa, M.J. and Dufrene, E., 2004. Retrieval of forest biophysical variables by inverting a 3-D radiative transfer model and using high and very high-resolution imagery. *International Journal of Remote Sensing*, *25*(24), pp.5601-5616.

Gastellu-Etchegorry, J.P., Demarez, V., Pinel, V. and Zagolski, F., 1996. Modeling radiative transfer in heterogeneous 3-D vegetation canopies. *Remote sensing of environment*, *58*(2), pp.131-156.

Gastellu-Etchegorry, J.P., 2008. 3D modeling of satellite spectral images, radiation budget and energy budget of urban landscapes. *Meteorology And Atmospheric Physics*, *102*(3), pp.187-207.

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

Gitelson, A.A., Zur, Y., Chivkunova, O.B. and Merzlyak, M.N., 2002. Assessing carotenoid content in plant leaves with reflectance spectroscopy¶. *Photochemistry And Photobiology*, *75*(3), pp.272-281.

Gitelson, A.A. and Merzlyak, M.N., 2004. Non-destructive assessment of chlorophyll carotenoid and anthocyanin content in higher plant leaves: principles and algorithms.

Gitelson, A.A., Keydan, G.P. and Merzlyak, M.N., 2006. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, *33*(11).

Gitelson, A., Chivkunova, O., Zhigalova, T. and Solovchenko, A., 2017. In situ optical properties of foliar flavonoidsonoids: Implication for non-destructive estimation of flavonoidsonoid content. *Journal Of Plant Physiology*, *218*, pp.258-264.

Gould, K.S., 2004. Nature′ s swiss army knife: the diverse protective roles of anthocyanins in leaves. *Biomed Research International*, *2004*(5), pp.314-320.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C. and Straif, C., 2015. The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing*, *7*(7), pp.8830-8857.

Gui-bin, W.A.N.G., Xu-qin, G.U.O., Li, C.H.A.N.G. and Fu-liang, C.A.O., 2013. Effects of air temperature and soil moisture on flavonoidsonoids accumulation in Ginkgo biloba leaves. *Yingyong Shengtai Xuebao*, *24*(11).

Han, Y., Dong, Y., Zhu, Y. and Huang, W., 2023. Remote sensing inversion of vegetation parameters with IPROSAIL-Net. *IEEE Transactions On Geoscience And Remote Sensing*.

Hansen, P.M. and Schjoerring, J.K., 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, *86*(4), pp.542-553.

Hati, J.P., Goswami, S., Samanta, S., Pramanick, N., Majumdar, S.D., Chaube, N.R., Misra, A. and Hazra, S., 2021. Estimation of vegetation stress in the mangrove forest using AVIRIS-NG airborne hyperspectral data. *Modeling Earth Systems and Environment*, *7*, pp.1877-1889.

He, B., Quan, X., Xu, D., Yin, C., Liao, Z., Qiu, S., Ge, J. and Zhang, Z., 2017. Retrieving grassland canopy water content by considering the information from neighboring pixels. *Photogrammetric Engineering & Remote Sensing*, *83*(8), pp.553-565.

He, L., Li, A., Yin, G., Nan, X. and Bian, J., 2019. Retrieval of grassland aboveground biomass through inversion of the PROSAIL model with MODIS imagery. *Remote Sensing*, *11*(13), p.1597.

He, C., Sun, J., Chen, Y., Wang, L., Shi, S., Qiu, F., Wang, S., Yang, J. and Tagesson, T., 2023. PROSPECT-GPR: Exploring spectral associations among vegetation traits in wavelength selection for leaf mass per area and water contents. *Science of Remote Sensing*, *8*, p.100100.

He, S., Qi, J., Wang, D., Yan, K. and Huang, H., 2024. Estimation of canopy photon recollision probability from airborne laser scanning. *Remote Sensing of Environment*, *311*, p.114264.

Hengl, T., Leenaars, J.G., Shepherd, K.D., Walsh, M.G., Heuvelink, G.B., Mamo, T., Tilahun, H., Berkhout, E., Cooper, M., Fegraus, E. and Wheeler, I., 2017. Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutrient Cycling in Agroecosystems*, *109*, pp.77-102.

Hernández-Clemente, R., Navarro-Cerrillo, R.M. and Zarco-Tejada, P.J., 2012. Carotenoid content estimation in a heterogeneous conifer forest using narrow-band indices and PROSPECT+ DART simulations. *Remote Sensing of Environment*, *127*, pp.298-315.

Hernández-Clemente, R., North, P.R., Hornero, A. and Zarco-Tejada, P.J., 2017. Assessing the effects of forest health on sun-induced chlorophyll fluorescence using the FluorFLIGHT 3-D radiative transfer model to account for forest structure. *Remote Sensing of Environment*, *193*, pp.165-179.

Hiruma, K., 2019. Roles of plant-derived secondary metabolites during interactions with pathogenic and beneficial microbes under conditions of environmental stress. *Microorganisms*, *7*(9), p.362.

Holmberg, K. and Hellsten, I., 2016. Integrating and differentiating meanings in tweeting about the fifth Intergovernmental Panel on Climate Change (IPCC) report. *First Monday*.

Holst, T., Hauser, S., Kirchgässner, A., Matzarakis, A., Mayer, H. and Schindler, D., 2004. Measuring and modelling plant area index in beech stands. *International Journal of Biometeorology*, *48*(4), pp.192-201.

Hornero, A., North, P.R., Zarco-Tejada, P.J., Rascher, U., Martín, M.P., Migliavacca, M. and Hernández-Clemente, R., 2021. Assessing the contribution of understory sun-induced chlorophyll fluorescence through 3-D radiative transfer modelling and field data. *Remote Sensing of Environment*, *253*, p.112195.

Huang, W., Zhou, X., Kong, W. and Ye, H., 2018. Monitoring Crop Carotenoids Concentration by Remote Sensing. In *Progress in Carotenoid Research*. London, UK: IntechOpen.

Huemmrich, K.F., 2001. The GeoSail model: a simple addition to the SAIL model to describe discontinuous canopy reflectance. *Remote Sensing of Environment*, *75*(3), pp.423-431.

Ishaq, R.A.F., Zhou, G., Tian, C., Tan, Y., Jing, G. and Jiang, H., 2023. A systematic review of radiative transfer models for crop yield prediction and crop traits retrieval. *Remote Sensing*, *16*(1), p.121.?

Jackson, W.P.U., 1990. *Origins and meanings of names of South African plant genera*. University of Cape Town.

Jacquemoud, S. and Baret, F., 1990. PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, *34*(2), pp.75-91.? – Consistency??

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C. and Ustin, S.L., 2009. PROSPECT+ SAIL models: A review of use for vegetation characterization. *Remote Sensing Of Environment*, *113*, pp.S56-S66.

Jiang, J., Weiss, M., Liu, S., Rochdi, N. and Baret, F., 2020. Speeding up 3D radiative transfer simulations: A physically based metamodel of canopy reflectance dependency on wavelength, leaf biochemical composition and soil reflectance. *Remote Sensing of Environment*, *237*, p.111614.

Jiang, Y., Kong, P., Xu, H., Zhang, Q., Quan, X. and He, B., 2021, July. Temporal Mapping of Grassland Aboveground Biomass in Qinghai Province from Landsat 8 and Sentinel-2. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS* (pp. 6901-6904). IEEE.

Jin, D., Qi, J., Huang, H. and Li, L., 2021. Combining 3D radiative transfer model and convolutional neural network to accurately estimate Forest canopy cover from very high-resolution satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, pp.10953-10963.

Juszak, I., Iturrate-Garcia, M., Gastellu-Etchegorry, J.P., Schaepman, M.E., Maximov, T.C. and Schaepman-Strub, G., 2017. Drivers of shortwave radiation fluxes in Arctic tundra across scales. *Remote Sensing of Environment*, *193*, pp.86-102.

Kallel, A., 2020. FluLCVRT: Reflectance and fluorescence of leaf and canopy modeling based on Monte Carlo vector radiative transfer simulation. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *253*, p.107183.

Keeley, S.C., Keeley, J.E., Hutchinson, S.M. and Johnson, A.W., 1981. Postfire succession of the herbaceous flora in southern California chaparral. *Ecology*, *62*(6), pp.1608-1621.

Kennedy, B.E., King, D.J. and Duffe, J., 2020. Comparison of empirical and physical modelling for estimation of biochemical and biophysical vegetation properties: field scale analysis across an Arctic bioclimatic gradient. *Remote Sensing*, *12*(18), p.3073.

Khoo, H.E., Azlan, A., Tang, S.T. and Lim, S.M., 2017. Anthocyanidins and anthocyanins: Colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food & Nutrition Research*, *61*(1), p.1361779.

Kimes, D.S., Nelson, R.F., Manry, M.T. and Fung, A.K., 1998. Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. International journal of remote sensing, 19(14), pp.2639-2663.

Koetz, B., Sun, G., Morsdorf, F., Ranson, K.J., Kneubühler, M., Itten, K. and Allgöwer, B., 2007. Fusion of imaging spectrometer and LIDAR data over combined radiative transfer models for forest canopy characterization. *Remote Sensing of Environment*, *106*(4), pp.449-459.

Kokaly, R.F. and Clark, R.N., 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment*, *67*(3), pp.267-287.

Kozlowski, T.T. ed., 2012. *Fire and Ecosystems*. Elsevier.

Kraaij, T. and van Wilgen, B.W., 2014. Drivers, ecology, and management of fire in fynbos. *Fynbos: Ecology, Evolution, And Conservation of A Megadiverse Region*, *15*(1), pp.47-72.

Kükenbrink, D., Schneider, F.D., Schmid, B., Gastellu-Etchegorry, J.P., Schaepman, M.E. and Morsdorf, F., 2021. Modelling of three-dimensional, diurnal light extinction in two contrasting forests. *Agricultural and Forest Meteorology*, *296*, p.108230.

Kumar, S., Abedin, M.M., Singh, A.K. and Das, S., 2020. Role of phenolic compounds in plant-defensive mechanisms. *Plant Phenolics in Sustainable Agriculture: volume 1*, pp.517-532.

Labate, D., Ceccherini, M., Cisbani, A., De Cosmo, V., Galeazzi, C., Giunti, L., Melozzi, M., Pieraccini, S. and Stagi, M., 2009. The PRISMA payload optomechanical design, a high-performance instrument for a new hyperspectral mission. *Acta Astronautica*, *65*(9-10), pp.1429-1436.

Lally, H.T., O'Connor, I., Jensen, O.P. and Graham, C.T., 2019. Can drones be used to conduct water sampling in aquatic environments? A review. *Science Of the Total Environment*, *670*, pp.569-575.

Lawes, M.J., Macfarlane, D.M. and Eeley, H.A., 2004. Forest landscape pattern in the KwaZulu–Natal midlands, South Africa: 50 years of change or stasis? *Austral Ecology*, *29*(6), pp.613-623.

Lehnert, L.W., Meyer, H., Meyer, N., Reudenbach, C. and Bendix, J., 2014. A hyperspectral indicator system for rangeland degradation on the Tibetan Plateau: A case study towards spaceborne monitoring. *Ecological Indicators*, *39*, pp.54-64.

Lehmann, J.R.K., Große-Stoltenberg, A., Römer, M. and Oldeland, J., 2015. Field spectroscopy in the VNIR-SWIR region to discriminate between Mediterranean native plants and exotic-invasive shrubs based on leaf tannin content. *Remote Sensing*, *7*(2), pp.1225-1241.

Leonenko, G., Los, S.O. and North, P.R.J., 2013. Retrieval of leaf area index from MODIS surface reflectance by model inversion using different minimization criteria. *Remote Sensing Of Environment*, *139*, pp.257-270.

Li, D., Cheng, T., Jia, M., Zhou, K., Lu, N., Yao, X., Tian, Y., Zhu, Y. and Cao, W., 2018. PROCWT: Coupling PROSPECT with continuous wavelet transform to improve the retrieval of foliar chemistry from leaf bidirectional reflectance spectra. *Remote Sensing of Environment*, *206*, pp.1-14.

Liang, S., 2005. *Quantitative Remote Sensing of Land Surfaces*. John Wiley & Sons.

Liang, S., 2007. Recent developments in estimating land surface biogeophysical variables from optical remote sensing. *Progress in Physical Geography*, *31*(5), pp.501-516.

Li-Cor, I., 1992. LAI-2000 plants canopy analyzer instruction manual. *LI-COR Inc.: Lincoln, NE, USA*.

Lichtenthaler, H.K. and Wellburn, A.R., 1983. Determinations of total carotenoids and chlorophylls a and b of leaf extracts in different solvents.

Ließ, M., Schmidt, J. and Glaser, B., 2016. Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches. *PLoS One*, *11*(4), p.e0153673.

Liu, W., Atherton, J., Mõttus, M., Gastellu-Etchegorry, J.P., Malenovský, Z., Raumonen, P., Åkerblom, M., Mäkipää, R. and Porcar-Castell, A., 2019. Simulating solar-induced chlorophyll fluorescence in a boreal forest stand reconstructed from terrestrial laser scanning measurements. *Remote Sensing of Environment*, *232*, p.111274.

Liu, L., Li, S., Yang, W., Wang, X., Luo, X., Ran, P. and Zhang, H., 2023. Forest Canopy Water Content Monitoring Using Radiative Transfer Models and Machine Learning. *Forests*, *14*(7), p.1418.

Locherer, M., Hank, T., Danner, M. and Mauser, W., 2015. Retrieval of seasonal leaf area index from simulated EnMAP data through optimized LUT-based inversion of the PROSAIL model. *Remote Sensing*, *7*(8), pp.10321-10346.

López-Serrano, F.R., Landete-Castillejos, T., Martínez-Millán, J. and del Cerro-Barja, A., 2000. LAI estimation of natural pine forest using a non-standard sampling technique. *Agricultural and Forest Meteorology*, *101*(2-3), pp.95-111.

Lu, B., Proctor, C. and He, Y., 2021. Investigating different versions of PROSPECT and PROSAIL for estimating spectral and biophysical properties of photosynthetic and non-photosynthetic vegetation in mixed grasslands. *GIScience & Remote Sensing*, *58*(3), pp.354-371.

Lukeš, P., Neuwirthová, E., Lhotáková, Z., Janoutová, R. and Albrechtová, J., 2020. Upscaling seasonal phenological course of leaf dorsiventral reflectance in radiative transfer model. *Remote Sensing of Environment*, *246*, p.111862.

Lunagaria, M.M. and Patel, H.R., 2019. Evaluation of PROSAIL inversion for retrieval of chlorophyll, leaf dry matter, leaf angle, and leaf area index of wheat using spectrodirectional measurements. *International Journal of Remote Sensing*, *40*(21), pp.8125-8145.

Luo, K., He, B., Quan, X., Liu, X. and Wen, C., 2019, July. Preliminarily analysis of the relation between satellite derived fuel moisture content and wildfire activity in southwestern China. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 9906-9909). IEEE.a

Luo, K., Quan, X., He, B. and Yebra, M., 2019. Effects of live fuel moisture content on wildfire occurrence in fire-prone regions over southwest China. *Forests*, *10*(10), p.887.b

Madonsela, S., Cho, M.A., Ramoelo, A. and Mutanga, O., 2017. Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS Journal of Photogrammetry and Remote Sensing*, *133*, pp.116-127.

Madonsela, S., 2018. *Exploring The Relationship Between Spectral Reflectance and Tree Species Diversity in The Savannah Woodlands* (Doctoral dissertation).

Magadlela, A., Vardien, W., Kleinert, A., Dreyer, L.L. and Valentine, A.J., 2015. The role of phosphorus deficiency in nodule microbial composition, and carbon and nitrogen nutrition of a native legume tree in the Cape fynbos ecosystem. *Australian Journal of Botany*, *63*(5), pp.379-386.

Makkar, H., 1999, Quantification of tannins in tree foliage. A laboratory manual forthe FAO/IAEA coordinated research project on 'Use of nuclear and relatedtechniques to develop simple tannin assay for predicting and improving thesafety and efficiency of feeding ruminants on the tanniniferous tree foliage', pp.1–29, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture,Vienna.

Malenovský, Z., Martin, E., Homolová, L., Gastellu-Etchegorry, J.P., Zurita-Milla, R., Schaepman, M.E., Pokorný, R., Clevers, J.G. and Cudlín, P., 2008. Influence of woody elements of a Norway spruce canopy on nadir reflectance simulated by the DART model at very high spatial resolution. *Remote Sensing of Environment*, *112*(1), pp.1-18.

Manjunath, K.R., Ray, S.S. and Vyas, D., 2016. Identification of indices for accurate estimation of anthocyanin and carotenoids in different species of flowers using hyperspectral data. *Remote Sensing Letters*, *7*(10), pp.1004-1013.

Martínez-Ferrer, L., Moreno-Martínez, Á., Campos-Taberner, M., García-Haro, F.J., Muñoz-Marí, J., Running, S.W., Kimball, J., Clinton, N. and Camps-Valls, G., 2022. Quantifying uncertainty in high resolution biophysical variable retrieval with machine learning. *Remote Sensing of Environment*, *280*, p.113199.

Masemola, C., Cho, M.A. and Ramoelo, A., 2016. Comparison of Landsat 8 OLI and Landsat 7 ETM+ for estimating grassland LAI using model inversion and spectral indices: case study of Mpumalanga, South Africa. *International Journal of Remote Sensing*, *37*(18), pp.4401-4419.

Masenyama, A., Mutanga, O., Dube, T., Bangira, T., Sibanda, M. and Mabhaudhi, T., 2022. A systematic review on the use of remote sensing technologies in quantifying grasslands ecosystem services. *GIScience & Remote Sensing*, *59*(1), pp.1000-1025.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I. and Huang, M., 2021. Climate change 2021: the physical science basis. *Contribution Of Working Group I To the Sixth Assessment Report of The Intergovernmental Panel on Climate Change*, *2*(1), p.2391.

Mediavilla, S., González-Zurdo, P., García-Ciudad, A. and Escudero, A., 2011. Morphological and chemical leaf composition of Mediterranean evergreen tree species according to leaf age. *Trees*, *25*, pp.669-677.

Melendo-Vega, J.R., Martín, M.P., Pacheco-Labrador, J., González-Cascón, R., Moreno, G., Pérez, F., Migliavacca, M., García, M., North, P. and Riaño, D., 2018. Improving the performance of 3-D radiative transfer model FLIGHT to simulate optical properties of a tree-grass ecosystem. *Remote Sensing*, *10*(12), p.2061.

Meng, B., Ge, J., Liang, T., Yang, S., Gao, J., Feng, Q., Cui, X., Huang, X. and Xie, H., 2017. Evaluation of remote sensing inversion error for the above-ground biomass of alpine meadow grassland based on multi-source satellite data. *Remote Sensing*, *9*(4), p.372.

Merzlyak, M.N., Solovchenko, A.E., Smagin, A.I. and Gitelson, A.A., 2005. Apple flavonoidsonols during fruit adaptation to solar radiation: spectral features and technique for non-destructive assessment. *Journal Of Plant Physiology*, *162*(2), pp.151-160.

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.S., Cheng, K., Das, B.S. and Field, D.J., 2017. Soil carbon 4 per mille. *Geoderma*, *292*, pp.59-86.

Miraglio, T., Adeline, K., Huesca, M., Ustin, S. and Briottet, X., 2019. Monitoring LAI, chlorophylls, and carotenoids content of a woodland savanna using hyperspectral imagery and 3D radiative transfer modeling. *Remote Sensing*, *12*(1), p.28.

Mitchell, D.T., Coley, P.G.F., Webb, S. and Allsopp, N., 1986. Litterfall and decomposition processes in the coastal fynbos vegetation, south-western Cape, South Africa. *The Journal of Ecology*, pp.977-993.

Mngadi, M., Odindi, J. and Mutanga, O., 2021. The utility of sentinel-2 spectral data in quantifying above-ground carbon stock in an urban reforested landscape. *Remote Sensing*, *13*(21), p.4281.

Mngadi, M., Odindi, J., Mutanga, O. and Sibanda, M., 2022. Quantitative remote sensing of forest ecosystem services in sub-Saharan Africa's urban landscapes: A review. *Environmental Monitoring and Assessment*, *194*(4), p.242.

Mngadi, M., Odindi, J. and Mutanga, O., 2022. Quantifying carbon stock variability of species within a reforested urban landscape using texture measures derived from remotely sensed imagery. *Advances in Remote Sensing for Forest Monitoring*, pp.150-166.

Montesano, P.M., Rosette, J., Sun, G., North, P., Nelson, R.F., Dubayah, R.O., Ranson, K.J. and Kharuk, V., 2015. The uncertainty of biomass estimates from modeled ICESat-2 returns across a boreal forest gradient. *Remote Sensing of Environment*, *158*, pp.95-109.

Morgan, J.W., 1999. Defining grassland fire events and the response of perennial plants to annual fire in temperate grasslands of south-eastern Australia. *Plant Ecology*, *144*, pp.127-144.

Moses, W.J., Bowles, J.H., Lucke, R.L. and Corson, M.R., 2012. Impact of signal-to-noise ratio in a hyperspectral sensor on the accuracy of biophysical parameter estimation in case II waters. *Optics Express*, *20*(4), pp.4309-4330.

Mthembu, N., Lottering, R. and Kotze, H., 2024. Estimating leaf area index of Eucalyptus dunnii over the dry and wet seasons using vegetation indices and image texture derived from WorldView-3 imagery. *International Journal of Remote Sensing*, *45*(12), pp.3871-3896.

Musungu, K., Shoko, M. and Smit, J., 2025. Determining the Spectral Characteristics of Fynbos Wetland Vegetation Species Using Unmanned Aerial Vehichle Data. *Geomatics*, *5*(2), p.17.

Mutanga, O. and Skidmore, A.K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal Of Remote Sensing*, *25*(19), pp.3999-4014.

Mutanga, O., Skidmore, A.K., Kumar, L. and Ferwerda, J., 2005. Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain. *International Journal of Remote Sensing*, *26*(6), pp.1093-1108.

Mutanga, O., Van Aardt, J. and Kumar, L., 2009. Imaging spectroscopy (hyperspectral remote sensing) in southern Africa: an overview. *South African Journal of Science*, *105*(5), pp.193-198.

Naczk, M. and Shahidi, F., 2004. Extraction and analysis of phenolics in food. *Journal of Chromatography A*, *1054*(1-2), pp.95-111.

Naidoo, L., M. A. Cho, R. Mathieu, and G. Asner. 2012. "Classification of Savanna Tree Species, in theGreater Kruger National Park Region, by Integrating Hyperspectral and LiDAR Data in a RandomForest Data Mining Environment." *ISPRS Journal of Photogrammetry & Remote Sensing* 69:167–179.https://doi.org/10.1016/j.isprsjprs.2012.03.005.

Nandave, M., Ojha, S.K., Joshi, S., Kumari, S. and Arya, D.S., 2009. Moringa oleifera leaf extract prevents isoproterenol-induced myocardial damage in rats: evidence for an antioxidant, antiperoxidative, and cardioprotective intervention. *Journal Of Medicinal Food*, *12*(1), pp.47-55.

Ndebele, N.E., Grab, S. and Hove, H., 2022. Wet season rainfall characteristics and temporal changes for Cape Town, South Africa, 1841–2018. *Climate of the Past*, *18*(11), pp.2463-2482.

Ndlovu, N., Luck-Vogel, M., Schloms, B. and Cho, M., 2011. The quantification of human impact on the Dukuduku indigenous forest from 1960 to 2008 using GIS techniques as a basis for sustainable management. In *Fifth Natural Forest and Woodland Symp*.

Ndlovu, H.S., Odindi, J., Sibanda, M. and Mutanga, O., 2024. A systematic review on the application of UAV-based thermal remote sensing for assessing and monitoring crop water status in crop farming systems. *International Journal of Remote Sensing*, *45*(15), pp.4923-4960.

Neary, G.D. and Leonard, M.J., 2020. Effects of fire on grassland soils and water: A review. *Grasses And Grassland Aspects*, pp.1-22.

Nelson, C.R., Hallett, J.G., Romero Montoya, A.E., Andrade, A., Besacier, C., Boerger, V., Bouazza, K., Chazdon, R., Cohen-Shacham, E., Danano, D. and Diederichsen, A., 2024. *Standards Of Practice to Guide Ecosystem Restoration: A Contribution to The United Nations Decade on Ecosystem Restoration 2021–2030*. Food & Agriculture Org.

Ni, Y.W., Lin, K.H., Chen, K.H., Wu, C.W. and Chang, Y.S., 2020. Flavonoid compounds and photosynthesis in Passiflora plant leaves under varying light intensities. *Plants*, *9*(5), p.633.

North, P.R., 1996. Three-dimensional forest light interaction model using a Monte Carlo method. *IEEE Transactions on Geoscience and Remote Sensing*, *34*(4), pp.946-956.

North, P.R.J., Rosette, J.A.B., Suárez, J.C. and Los, S.O., 2010. A Monte Carlo radiative transfer model of satellite waveform LiDAR. *International Journal of Remote Sensing*, *31*(5), pp.1343-1358.

Nouvellon, Y., Moran, M.S., Chehbouniz, A., Seen, D.L., Bryant, R., Nichols, M., Prévot, L., Rambal, S., Ni, W., Bégué, A. and Heilman, P., 2000, July. Assimilating LANDSAT data in an ecosystem model for multi-year simulation of grassland carbon, water and energy budget. In *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No. 00CH37120)* (Vol. 5, pp. 1966-1968). IEEE.

Odebiri, O.O., 2022. *The Application of Deep Learning for Remote Sensing of Soil Organic Carbon Stocks Distribution In South Africa* (Doctoral dissertation).

Odebiri, O., Mutanga, O. and Odindi, J., 2022. Deep learning-based national scale soil organic carbon mapping with Sentinel-3 data. *Geoderma*, *411*, p.115695.

Odebiri, O., Mutanga, O., Odindi, J. and Naicker, R., 2023. Mapping soil organic carbon distribution across South Africa's major biomes using remote sensing-topo-climatic covariates and Concrete Autoencoder-Deep neural networks. *Science of the Total Environment*, *865*, p.161150.

Owens, J.D., Luebke, D., Govindaraju, N., Harris, M., Krüger, J., Lefohn, A.E. and Purcell, T.J., 2007, March. A survey of general-purpose computation on graphics hardware. In *Computer Graphics Forum* (Vol. 26, No. 1, pp. 80-113). Oxford, UK: Blackwell Publishing Ltd.

Pacheco-Labrador, J., El-Madany, T.S., Van der Tol, C., Martin, M.P., Gonzalez-Cascon, R., Perez-Priego, O., Guan, J., Moreno, G., Carrara, A., Reichstein, M. and Migliavacca, M., 2021. senSCOPE: Modeling mixed canopies combining green and brown senesced leaves. Evaluation in a Mediterranean Grassland. *Remote Sensing of Environment*, *257*, p.112352.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E. and Chou, R., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, *372*.

Pant, P., Pandey, S. and Dall'Acqua, S., 2021. The influence of environmental conditions on secondary metabolites in medicinal plants: A literature review. *Chemistry & Biodiversity*, *18*(11), p.e2100345.

Paramanik, S., Behera, M.D., Deep, N.R., Barnwal, S., Bhattacharya, B.K., Behera, S.K., Swain, D.K. and Dash, J., 2025. Leaf Chlorophyll Concentration Estimation Using Absorption Spectroscopy of AVIRIS-NG for a Mangrove Forest in India. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, *93*(3), pp.285-298.

Pascual-Venteo, A.B., Garcia, J.L., Berger, K., Estévez, J., Vicent, J., Pérez-Suay, A., Van Wittenberghe, S. and Verrelst, J., 2024. Gaussian Process Regression Hybrid Models for the Top-of-Atmosphere Retrieval of Vegetation Traits Applied to PRISMA and EnMAP Imagery. *Remote Sensing*, *16*(7), p.1211.

Pasolli, E., Melgani, F., Alajlan, N. and Bazi, Y., 2012. Active learning methods for biophysical parameter estimation. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(10), pp.4071-4084.

Pillay, S.J., Bangira, T., Sibanda, M., Kebede Gurmessa, S., Clulow, A. and Mabhaudhi, T., 2024. Assessing Drone-Based Remote Sensing for Monitoring Water Temperature, Suspended Solids and CDOM in Inland Waters: A Global Systematic Review of Challenges and Opportunities. *Drones*, *8*(12), p.733.

Pinheiro, A.C., Privette, J.L. and Guillevic, P., 2006. Modeling the observed angular anisotropy of land surface temperature in a savanna. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(4), pp.1036-1047.

Porter, L.J., 1992. Structure and chemical properties of the condensed tannins. In *Plant Polyphenols: Synthesis, Properties, Significance* (pp. 245-258). Boston, MA: Springer US.

Prabhakar, M., Gopinath, K.A., Ravi Kumar, N., Thirupathi, M., Sai Sravan, U., Srasvan Kumar, G., Samba Siva, G., Chandana, P. and Singh, V.K., 2024. Mapping Leaf Area Index at Various Rice Growth Stages in Southern India Using Airborne Hyperspectral Remote Sensing. *Remote Sensing*, *16*(6), p.954.

Proctor, C., Lu, B. and He, Y., 2017. Determining the absorption coefficients of decay pigments in decomposing monocots. *Remote Sensing of Environment*, *199*, pp.137-153.

Pu, J., Yan, K., Zhou, G., Lei, Y., Zhu, Y., Guo, D., Li, H., Xu, L., Knyazikhin, Y. and Myneni, R.B., 2020. Evaluation of the MODIS LAI/FPAR algorithm based on 3D-RTM simulations: A case study of grassland. *Remote Sensing*, *12*(20), p.3391.

Qabaqaba, M., Naidoo, L., Tsele, P., Ramoelo, A. and Cho, M.A., 2023. Integrating random forest and synthetic aperture radar improves the estimation and monitoring of woody cover in indigenous forests of South Africa. *Applied Geomatics*, *15*(1), pp.209-225.

Qi, J., Jiang, J., Zhou, K., Xie, D. and Huang, H., 2023. Fast and accurate simulation of canopy reflectance under wavelength-dependent optical properties using a semi-empirical 3D radiative transfer model. *Journal of Remote Sensing*, *3*, p.0017.

Qin, G., Wu, J., Li, C. and Meng, Z., 2024. Comparison of the hybrid of radiative transfer model and machine learning methods in leaf area index of grassland mapping. *Theoretical and Applied Climatology*, *155*(4), pp.2757-2773.

Quan, X., He, B., Li, X. and Liao, Z., 2015. Retrieval of grassland live fuel moisture content by parameterizing radiative transfer model with interval estimated LAI. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(2), pp.910-920a.

Quan, X., He, B., Li, X. and Tang, Z., 2015. Estimation of grassland live fuel moisture content from ratio of canopy water content and foliage dry biomass. *IEEE Geoscience and Remote Sensing Letters*, *12*(9), pp.1903-1907b.

Quan, X., He, B., Yebra, M., Yin, C., Liao, Z., Zhang, X. and Li, X., 2017. A radiative transfer model-based method for the estimation of grassland aboveground biomass. *International Journal of Applied Earth Observation and Geoinformation*, *54*, pp.159-168.

Ramoelo, A., Skidmore, A.K., Schlerf, M., Heitkönig, I.M., Mathieu, R. and Cho, M.A., 2013. Savanna grass nitrogen to phosphorous ratio estimation using field spectroscopy and the potential for estimation with imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, *23*, pp.334-343.

Raymond Hunt Jr, E. and Daughtry, C.S., 2014. Chlorophyll meter calibrations for chlorophyll content using measured and simulated leaf transmittances. *Agronomy Journal*, *106*(3), pp.931-939.

Regaieg, O., Yin, T., Malenovský, Z., Cook, B.D., Morton, D.C. and Gastellu-Etchegorry, J.P., 2021. Assessing impacts of canopy 3D structure on chlorophyll fluorescence radiance and radiative budget of deciduous forest stands using DART. *Remote Sensing of Environment*, *265*, p.112673.

Regaieg, O., 2023. *Modeling Radiative Transfer of Sun-Induced Fluorescence, Thermal Emission and Radiative Budget Of 3D Vegetation Canopies: Towards A Comprehensive 3D SIF Model* (Doctoral dissertation, Université Paul Sabatier-Toulouse III).

Reginato, M.A., Castagna, A., Furlán, A., Castro, S., Ranieri, A. and Luna, V., 2014. Physiological responses of a halophytic shrub to salt stress by Na2SO4 and NaCl: oxidative damage and the role of polyphenols in antioxidant protection. *AoB Plants*, *6*, p.plu042.

Rocha, A.D., Groen, T.A., Skidmore, A.K., Darvishzadeh, R. and Willemen, L., 2018. Machine learning using hyperspectral data inaccurately predicts plant traits under spatial dependency. *Remote sensing*, *10*(8), p.1263.

Rosema, A., Verhoef, W., Schroote, J. and Snel, J.F.H., 1991. Simulating fluorescence light-canopy interaction in support of laser-induced fluorescence measurements. *Remote Sensing of Environment*, *37*(2), pp.117-130.

Sakai, Y., Kobayashi, H. and Kato, T., 2020. FLiES-SIF version 1.0: three-dimensional radiative transfer model for estimating solar induced fluorescence. *Geoscientific Model Development*, *13*(9), pp.4041-4066.

Sauer, J., Mariani, V.C., dos Santos Coelho, L., Ribeiro, M.H.D.M. and Rampazzo, M., 2022. Extreme gradient boosting model based on improved Jaya optimizer applied to forecasting energy consumption in residential buildings. *Evolving Systems*, pp.1-12.

Savitzky, A. and Golay, M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*(8), pp.1627-1639.

Schimel, D., Pavlick, R., Fisher, J.B., Asner, G.P., Saatchi, S., Townsend, P., Miller, C., Frankenberg, C., Hibbard, K. and Cox, P., 2015. Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology*, *21*(5), pp.1762-1776.

Schlerf, M. and Atzberger, C., 2006. Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*, *100*(3), pp.281-294.

Schlerf, M. and Atzberger, C., 2012. Vegetation structure retrieval in beech and spruce forests using spectrodirectional satellite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(1), pp.8-17.

Schucknecht, A., Reinermann, S., Kiese, R. and Bayer, A., 2023, October. Towards The Retrieval of Plant Traits in Grasslands with Hyperspectral Enmap Data. In *2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)* (pp. 1-5). IEEE.

Şenkal, B.C., 2020. The role of secondary metabolites obtained from medicinal and aromatic plants in our lives. *ISPEC Journal of Agricultural Sciences*, *4*(4), pp.1071-1079.

Shiklomanov, A.N., Dietze, M.C., Fer, I., Viskari, T. and Serbin, S.P., 2021. Cutting out the middleman: Calibrating and validating a dynamic vegetation model (ED2-PROSPECT5) using remotely sensed surface reflectance. *Geoscientific Model Development*, *14*(5), pp.2603-2633.

Si, Y., Schlerf, M., Zurita-Milla, R., Skidmore, A. and Wang, T., 2012. Mapping spatio-temporal variation of grassland quantity and quality using MERIS data and the PROSAIL model. *Remote Sensing of Environment*, *121*, pp.415-425.

Sibanda, M., Mutanga, O., Chimonyo, V.G., Clulow, A.D., Shoko, C., Mazvimavi, D., Dube, T. and Mabhaudhi, T., 2021. Application of drone technologies in surface water resources

monitoring and assessment: A systematic review of progress, challenges, and opportunities in the global south. *Drones*, *5*(3), p.84.

Sibanda, M. and Bacela, E., 2024. A Systematic Review of Literature on remote sensing tomato crops productivity. *Physics and Chemistry of the Earth, Parts A/B/C*, p.103759.

Sibiya, B., Lottering, R. and Odindi, J., 2021. Discriminating commercial forest species using image texture computed from a worldview-2 pan-sharpened image and partial least squares discriminant analysis. *Remote Sensing Applications: Society and Environment*, *23*, p.100605.

Sibiya, B., Lottering, R. and Odindi, J., 2022. Utility of texture combinations computed from fused WorldView-2 imagery in discriminating commercial forest species. *Geocarto International*, *37*(23), pp.6915-6931.

Sibiya, B.S.P., Odindi, J., Mutanga, O., Cho, M.A. and Masemola, C., 2025. The utility of radiative transfer models (RTM) on remotely sensed data in retrieving biophysical and biochemical properties of terrestrial biomes: A systematic review. *Advances in Space Research*.a

Sibiya, B., Cho, M., Mutanga, O., Odindi, J., Masemola, C. and Bonnet, W., 2025. Determining the value of different wavelength ranges of non-imaging hyperspectral reflectance to estimate carotenoid content using the PROSPECT-5 model. *International Journal of Remote Sensing*, *46*(4), pp.1696-1719.b

Sibiya, B., Odindi, J., Mutanga, O., Cho, M.A., Masemola, C., Stadenc, J.V. and Vambe, M., 2025. Leveraging Leaf Spectroscopy to Understand Variations in the Concentration of Phenolic Compounds in Post-Fire Recovery of the Cape Floristic Region. *ESS Open Archive eprints*, *548*, pp.essoar-174060575.

Sims, D.A. and Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, *81*(2-3), pp.337-354.

Singh, P., Srivastava, P.K., Malhi, R.K.M., Chaudhary, S.K., Verrelst, J., Bhattacharya, B.K. and Raghubanshi, A.S., 2020. Denoising AVIRIS-NG data for generation of new chlorophyll indices. *IEEE Sensors Journal*, *21*(5), pp.6982-6989.

Sinha, S.K., Padalia, H., Dasgupta, A., Verrelst, J. and Rivera, J.P., 2020. Estimation of leaf area index using PROSAIL based LUT inversion, MLRA-GPR and empirical models: Case study of tropical deciduous forest plantation, North India. *International Journal of Applied Earth Observation and Geoinformation*, *86*, p.102027.

Shen, N., Wang, T., Gan, Q., Liu, S., Wang, L. and Jin, B., 2022. Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity. *Food chemistry*, *383*, p.132531.

Shi, Y., Skidmore, A.K., Wang, T., Holzwarth, S., Heiden, U., Pinnel, N., Zhu, X. and Heurich, M., 2018. Tree species classification using plant functional traits from LiDAR and

hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, *73*, pp.207-219.

Singh, P., Srivastava, P.K., Verrelst, J., Mall, R.K., Rivera, J.P., Dugesar, V. and Prasad, R., 2023. High resolution retrieval of leaf chlorophyll content over Himalayan pine forest using Visible/IR sensors mounted on UAV and radiative transfer model. *Ecological Informatics*, *75*, p.102099.

Singh, P., Srivastava, P.K., Jha, P.K., Verrelst, J., Singh, P.N. and Prasad, R., 2025. Retrieval of crop traits using PROSAIL-based hybrid radiative transfer model and EnMAP hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, *143*, p.104769.

Silva, L.N., Freer-Smith, P. and Madsen, P., 2019. Production, restoration, mitigation: a new generation of plantations. *New Forests*, *50*(2), pp.153-168.

Skidmore, A.K., Ferwerda, J.G., Mutanga, O., Van Wieren, S.E., Peel, M., Grant, R.C., Prins, H.H., Balcik, F.B. and Venus, V., 2010. Forage quality of savannas—Simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. *Remote Sensing of Environment*, *114*(1), pp.64-72.

Skidmore, A.K., Coops, N.C., Neinavaz, E., Ali, A., Schaepman, M.E., Paganini, M., Kissling, W.D., Vihervaara, P., Darvishzadeh, R., Feilhauer, H. and Fernandez, M., 2021. Priority list of biodiversity metrics to observe from space. *Nature Ecology & Evolution*, *5*(7), pp.896-906.

Smit, I.P., Baard, J.A. and van Wilgen, B.W., 2024. Fire regimes and management options in mixed grassland-fynbos vegetation, South Africa. *Fire Ecology*, *20*(1), p.29.

Solovchenko, A.E., Chivkunova, O.B., Merzlyak, M.N. and Reshetnikova, I.V., 2001. A spectrophotometric analysis of pigments in apples. *Russian Journal of Plant Physiology*, *48*(5), pp.693-700.

Song, S., Gong, W., Zhu, B. and Huang, X., 2011. Wavelength selection and spectral discrimination for paddy rice, with laboratory measurements of hyperspectral leaf reflectance. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(5), pp.672-682.

Song, D., Gao, D., Sun, H., Qiao, L., Zhao, R., Tang, W. and Li, M., 2021. Chlorophyll content estimation based on cascade spectral optimizations of interval and wavelength characteristics. *Computers and Electronics in Agriculture*, *189*, p.106413.

Sonobe, R. and Hirono, Y., 2023. Carotenoid Content Estimation in Tea Leaves Using Noisy Reflectance Data. *Remote Sensing*, *15*(17), p.4303.

Sonobe, R. and Wang, Q., 2018. Nondestructive assessments of carotenoids content of broadleaved plant species using hyperspectral indices. *Computers and Electronics in Agriculture*, *145*, pp.18-26.

Soubry, I., Doan, T., Chu, T. and Guo, X., 2021. A systematic review on the integration of remote sensing and GIS to forest and grassland ecosystem health attributes, indicators, and measures. *Remote Sensing*, *13*(16), p.3262.

Spafford, L., Le Maire, G., MacDougall, A., De Boissieu, F. and Féret, J.B., 2021. Spectral subdomains and prior estimation of leaf structure improves PROSPECT inversion on reflectance or transmittance alone. *Remote Sensing of Environment*, *252*, p.112176.

Stock, W.D. and Allsopp, N., 1992. Functional perspective of ecosystems. *The Ecology of Fynbos. Nutrients, Fire and Diversity. Oxford University Press, Cape Town*, pp.241-259.

Stylinski, C., Gamon, J. and Oechel, W., 2002. Seasonal patterns of reflectance indices, carotenoid pigments and photosynthesis of evergreen chaparral species. *Oecologia*, *131*(3), pp.366-374.

Sun, J., Shi, S., Yang, J., Gong, W., Qiu, F., Wang, L., Du, L. and Chen, B., 2019. Wavelength selection of the multispectral lidar system for estimating leaf chlorophyll and water contents through the PROSPECT model. *Agricultural and forest meteorology*, *266*, pp.43-52.

Tattini, M., Galardi, C., Pinelli, P., Massai, R., Remorini, D. and Agati, G., 2004. Differential accumulation of flavonoidsonoids and hydroxycinnamates in leaves of Ligustrum vulgare under excess light and drought stress. *New Phytologist*, *163*(3), pp.547-561.

Thenkabail, P.S., Smith, R.B. and De Pauw, E., 2002. Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogrammetric Engineering and Remote Sensing*, *68*(6), pp.607-622.

Timothy, D., Onisimo, M. and Riyad, I., 2016. Quantifying aboveground biomass in African environments: A review of the trade-offs between sensor estimation accuracy and costs. *Tropical Ecology*, *57*(3), pp.393-405.

Treutter, D., 2006. Significance of flavonoidsonoids in plant resistance: a review. *Environmental Chemistry Letters*, *4*(3), pp.147-157.

Trombetti, M., Riaño, D., Rubio, M.A., Cheng, Y.B. and Ustin, S.L., 2008. Multi-temporal vegetation canopy water content retrieval and interpretation using artificial neural networks for the continental USA. *Remote Sensing of Environment*, *112*(1), pp.203-215.

Tsele, P., Ramoelo, A. and Qabaqaba, M., 2022, October. Evaluation of Sentinel-2 Spectral Reflectance and Indices to Estimate Grass LAI and CCC in Heterogeneous Grassland. In *International Conference of the African Association of Remote Sensing and the Environment* (pp. 363-372). Cham: Springer Nature Switzerland.

Tsele, P. and Ramoelo, A., 2024. Hybrid retrieval of grass biophysical variables based-on radiative transfer, active learning and regression methods using Sentinel-2 data in Marakele National Park. *Geocarto International*, *39*(1), p.2387087.

Tshabalala, T., Abdel-Rahman, E.M., Masemola, C., Ncube, B., Ndhlala, A.R. and Mutanga, O., 2021. Predicting medicinal phytochemicals of Moringa oleifera using hyperspectral reflectance of tree canopies. *International Journal of Remote Sensing*, *42*(10), pp.3955-3980.

Vambe, M., Singh, K., Coopoosamy, R.M., Naidoo, K. and Zengeni, R., 2024. Vermicompost leachates enhance morpho-physiological properties in Pelargonium sidoides DC. *Journal of Medicinal Plants for Economic Development*, *8*(1), p.256.

van der Merwe, A.M., Van der Walt, J.J.A. and Marais, E.M., 1994. Anatomical adaptations in the leaves of selected fynbos species. *South African Journal of Botany*, *60*(2), pp.99-107.

Van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A. and Su, Z., 2009. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. *Biogeosciences*, *6*(12), pp.3109-3129.

Van Eck, N. and Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), pp.523-538.

Van Staden, J., 1966. A paper chromatographic survey of free amino acids and related compounds in the Proteaceae. *JS Afr. Bot*, *32*, pp.77-86.

Varela, M.C., Arslan, I., Reginato, M.A., Cenzano, A.M. and Luna, M.V., 2016. Phenolic compounds as indicators of drought resistance in shrubs from Patagonian shrublands (Argentina). *Plant Physiology and Biochemistry*, *104*, pp.81-91.

Verger, A., Baret, F. and Camacho, F., 2011. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. *Remote Sensing of Environment*, *115*(2), pp.415-426.

Verhoef, W., 1985. Earth observation modeling based on layer scattering matrices. *Remote sensing of environment*, *17*(2), pp.165-178.

Verma, B., Prasad, R., Srivastava, P.K., Singh, P., Badola, A. and Sharma, J., 2022. Evaluation of simulated AVIRIS-NG imagery using a spectral reconstruction method for the retrieval of leaf chlorophyll content. *Remote Sensing*, *14*(15), p.3560.

Vermerris, W. and Nicholson, R., 2007. *Phenolic Compound Biochemistry*. Springer Science & Business Media.

Verrelst, J., Schaepman, M.E., Malenovský, Z. and Clevers, J.G., 2010. Effects of woody elements on simulated canopy reflectance: Implications for forest chlorophyll content retrieval. *Remote Sensing of Environment*, *114*(3), pp.647-656.

Verrelst, J., Romijn, E. and Kooistra, L., 2012. Mapping vegetation density in a heterogeneous river floodplain ecosystem using pointable CHRIS/PROBA data. *Remote Sensing*, *4*(9), pp.2866-2889.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G. and Moreno, J., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties–A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, pp.273-290.

Verrelst, J., Rivera, J.P., Gitelson, A., Delegido, J., Moreno, J. and Camps-Valls, G., 2016. Spectral band selection for vegetation properties retrieval using Gaussian processes regression. *International Journal of Applied Earth Observation and Geoinformation*, *52*, pp.554-567.

Verrelst, J., Dethier, S., Rivera, J.P., Munoz-Mari, J., Camps-Valls, G. and Moreno, J., 2016. Active learning methods for efficient hybrid biophysical variable retrieval. *IEEE Geoscience and Remote Sensing Letters*, *13*(7), pp.1012-1016.

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.P., Lewis, P., North, P. and Moreno, J., 2019. Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods. *Surveys in Geophysics*, *40*(3), pp.589-629.

Verrelst, J., Rivera-Caicedo, J.P., Reyes-Muñoz, P., Morata, M., Amin, E., Tagliabue, G., Panigada, C., Hank, T. and Berger, K., 2021. Mapping landscape canopy nitrogen content from space using PRISMA data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *178*, pp.382-395.

Vohland, M. and Jarmer, T., 2008. Estimating structural and biochemical parameters for grassland from spectroradiometer data by radiative transfer modelling (PROSPECT+ SAIL). *International Journal of Remote Sensing*, *29*(1), pp.191-209.

Vuolo, F., Neugebauer, N., Bolognesi, S.F., Atzberger, C. and D'Urso, G., 2013. Estimation of leaf area index using DEIMOS-1 data: Application and transferability of a semi-empirical relationship between two agricultural areas. *Remote Sensing*, *5*(3), pp.1274-1291.

Wan, L., Zhou, W., He, Y., Wanger, T.C. and Cen, H., 2022. Combining transfer learning and hyperspectral reflectance analysis to assess leaf nitrogen concentration across different plant species datasets. *Remote Sensing of Environment*, *269*, p.112826.

Wand, S.J., 1995. Concentration of ultraviolet-B radiation absorbing compounds in leaves of a range of fynbos species. *Vegetatio*, *116*, pp.51-61.

Wang, G., Cao, F., Chang, L., Guo, X. and Wang, J., 2014. Temperature has more effects than soil moisture on biosynthesis of flavonoids in Ginkgo (Ginkgo biloba L.) leaves. *New Forests*, *45*(6), pp.797-812.

Wang, L., 2022. *Improve Yield Prediction with Uas-Based Leaf Area Index Estimation and a Hybrid Machine Learning-and Process-Based Model* (Doctoral dissertation, The University of Nebraska-Lincoln).

Wang, Z., Skidmore, A.K., Darvishzadeh, R., Heiden, U., Heurich, M. and Wang, T., 2015. Leaf nitrogen content indirectly estimated by leaf traits derived from the PROSPECT model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(6), pp.3172-3182.

Wang, Z., Skidmore, A.K., Wang, T., Darvishzadeh, R., Heiden, U., Heurich, M., Latifi, H. and Hearne, J., 2017. Canopy foliar nitrogen retrieved from airborne hyperspectral imagery by correcting for canopy structure effects. *International Journal of Applied Earth Observation and Geoinformation*, *54*, pp.84-94.

Wang, Z., Skidmore, A.K., Darvishzadeh, R. and Wang, T., 2018. Mapping forest canopy nitrogen content by inversion of coupled leaf-canopy radiative transfer models from airborne hyperspectral imagery. *Agricultural And Forest Meteorology*, *253*, pp.247-260.

Wang, H. and Xiao, J., 2021. Improving the capability of the SCOPE model for simulating solar-induced fluorescence and gross primary production using data from OCO-2 and flux towers. *Remote Sensing*, *13*(4), p.794.

Wang, C., Yang, C., Zhang, J., Kuai, J., Xie, J., Wu, W., Zuo, Q., Yan, M., Du, H., Ma, N. and Liu, B., 2023. A PROSAIL model with a vegetation index lookup table optimized with in-situ statistics for rapeseed leaf area index estimation using diverse unmanned aerial vehicle sensors in the Yangtze River Basin. *Computers and Electronics in Agriculture*, *215*, p.108418.

Weiss, M., Baret, F., Myneni, R., Pragnère, A. and Knyazikhin, Y., 2000. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, *20*(1), pp.3-22.

Widlowski, J.L., Mio, C., Disney, M., Adams, J., Andredakis, I., Atzberger, C., Brennan, J., Busetto, L., Chelle, M., Ceccherini, G. and Colombo, R., 2015. The fourth phase of the radiative transfer model intercomparison (RAMI) exercise: Actual canopy scenarios and conformity testing. *Remote Sensing of Environment*, *169*, pp.418-437.

Williamson, G., Day, A.J., Plumb, G.W. and Couteau, D., 2000. Human metabolic pathways of dietary flavonoidsonoids and cinnamates. *Biochemical Society Transactions*, *28*(2), pp.16-22.

Wu, D., Feng, J., Lai, M., Ouyang, J., Liao, D., Yu, W., Wang, G., Cao, F., Jacobs, D.F. and Zeng, S., 2020. Combined application of bud and leaf growth fertilizer improves leaf flavonoids yield of Ginkgo biloba. *Industrial Crops and Products*, *150*, p.112379.

Xiang, T.Z., Xia, G.S. and Zhang, L., 2019. Mini-unmanned aerial vehicle-based remote sensing: Techniques, applications, and prospects. *IEEE Geoscience and Remote Sensing Magazine*, *7*(3), pp.29-63.

Xie, J., Wang, C., Ma, D., Chen, R., Xie, Q., Xu, B., Zhao, W. and Yin, G., 2022. Generating spatiotemporally continuous grassland aboveground biomass on the tibetan plateau through PROSAIL model inversion on google earth engine. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, pp.1-10.

Xie, R., Darvishzadeh, R., Skidmore, A. and van der Meer, F., 2024. Characterizing foliar phenolic compounds and their absorption features in temperate forests using leaf spectroscopy. *ISPRS Journal of Photogrammetry and Remote Sensing*, *212*, pp.338-356.

Xie, R., Darvishzadeh, R., Skidmore, A.K., van der Meer, F., Torres-Rodriguez, A. and Heurich, M., 2025. Mapping canopy phenolics in European mixed temperate forests using air- and space-borne imaging spectroscopy. *Remote Sensing of Environment*, *331*, p.115020.

Xu, S., Wang, M. and Shi, X., 2020. Hyperspectral imaging for high-resolution mapping of soil carbon fractions in intact paddy soil profiles with multivariate techniques and variable selection. *Geoderma*, *370*, p.114358.

Xu, L., Shi, S., Gong, W., Shi, Z., Qu, F., Tang, X., Chen, B. and Sun, J., 2022. Improving leaf chlorophyll content estimation through constrained PROSAIL model from airborne hyperspectral and LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, *115*, p.103128.

Yanez-Rausell, L., Malenovský, Z., Rautiainen, M., Clevers, J.G., Lukeš, P., Hanuš, J. and Schaepman, M.E., 2015. Estimation of spruce needle-leaf chlorophyll content based on DART and PARAS canopy reflectance models. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(4), pp.1534-1544.

Yang, P., Verhoef, W. and van der Tol, C., 2017. The mSCOPE model: A simple adaptation to the SCOPE model to describe reflectance, fluorescence and photosynthesis of vertically heterogeneous canopies. *Remote Sensing of Environment*, *201*, pp.1-11.

Yang, J., Yang, S., Zhang, Y., Shi, S. and Du, L., 2021. Improving characteristic band selection in leaf biochemical property estimation considering interrelations among biochemical parameters based on the PROSPECT-D model. *Optics Express*, *29*(1), pp.400-414.

Yang, C., Song, L., Wei, K., Gao, C., Wang, D., Feng, M., Zhang, M., Wang, C., Xiao, L., Yang, W. and Song, X., 2023. Study on Hyperspectral Monitoring Model of Total Flavonoids and Total Phenols in Tartary Buckwheat Grains. *Foods*, *12*(7), p.1354.

Yebra, M., Quan, X., Riaño, D., Larraondo, P.R., Van Dijk, A.I. and Cary, G.J., 2018, July. Mapping live fuel moisture content and flammability for continental Australia using optical remote sensing. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 5903-5906). IEEE.a

Yebra, M., Quan, X., Riaño, D., Larraondo, P.R., van Dijk, A.I. and Cary, G.J., 2018. A fuel moisture content and flammability monitoring methodology for continental Australia based on optical remote sensing. *Remote Sensing of Environment*, *212*, pp.260-272.b

Yin, C., He, B., Quan, X. and Liao, Z., 2016. Chlorophyll content estimation in arid grasslands from Landsat-8 OLI data. *International Journal of Remote Sensing*, *37*(3), pp.615-632.

Yin, C., He, B., Quan, X., Zhang, X., Ge, J. and Zhang, Z., 2017, July. Estimation of grassland biophysical variables for Lake Qinghai watershed: Moving towards remote sensing products. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3491-3494). IEEE.

Yin, C., He, B., Yebra, M., Quan, X., Edwards, A.C., Liu, X. and Liao, Z., 2020. Improving burn severity retrieval by integrating tree canopy cover into radiative transfer model simulation. *Remote Sensing of Environment*, *236*, p.111454.

Yin, S., Zhou, K., Cao, L. and Shen, X., 2022. Estimating the horizontal and vertical distributions of pigments in canopies of ginkgo plantation based on UAV-borne LiDAR, hyperspectral data by coupling PROSAIL model. *Remote Sensing*, *14*(3), p.715.

Yuan, H., Ma, R., Atzberger, C., Li, F., Loiselle, S.A. and Luo, J., 2015. Estimating forest fAPAR from multispectral Landsat-8 data using the invertible forest reflectance model INFORM. *Remote Sensing*, *7*(6), pp.7425-7446.

Yuan, W., Lin, S. and Wang, X., 2022. Progress of studies on satellite-based terrestrial vegetation production models in China. *Progress in Physical Geography: Earth and Environment*, *46*(6), pp.889-908.

Zarco-Tejada, P.J., Rueda, C.A. and Ustin, S.L., 2003. Water content estimation in vegetation with MODIS reflectance data and model inversion methods. *Remote Sensing of Environment*, *85*(1), pp.109-124.

Zarco-Tejada, P., 2005. Development of a vegetation fluorescence canopy model. *ESA Scientific and Technical Publications Branch, ESTEC*.

Zhang, Q., Xiao, X., Braswell, B., Linder, E., Baret, F. and Moore III, B., 2005. Estimating light absorption by chlorophyll, leaf and canopy in a deciduous broadleaf forest using MODIS data and a radiative transfer model. *Remote Sensing of Environment*, *99*(3), pp.357-371.

Zhang, Q., Xiao, X., Braswell, B., Linder, E., Ollinger, S., Smith, M.L., Jenkins, J.P., Baret, F., Richardson, A.D., Moore III, B. and Minocha, R., 2006. Characterization of seasonal variation of forest canopy in a temperate deciduous broadleaf forest, using daily MODIS data. *Remote Sensing of Environment*, *105*(3), pp.189-203.

Zhang, J.T., Fan, L. and Li, M., 2012. Functional diversity in plant communities: theory and analysis methods. *African Journal of Biotechnology*, *11*(5), pp.1014-1022.

Zhang, Y., Shu, Q., Wang, L., Quan, X., Liu, X. and Lu, B., 2019, July. Estimation of fuel biomass for grasslands using data assimilation technique. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 9988-9991). IEEE.

Zhang, Y., Migliavacca, M., Penuelas, J. and Ju, W., 2021. Advances in hyperspectral remote sensing of vegetation traits and functions. *Remote Sensing of Environment*, *252*, p.112121.

Zhang, R., Yang, P., Liu, S., Wang, C. and Liu, J., 2022. Evaluation of the methods for estimating leaf chlorophyll content with SPAD chlorophyll meters. *Remote Sensing*, *14*(20), p.5144.

Zhang, L., Gao, H. and Zhang, X., 2023. Combining Radiative Transfer Model and Regression Algorithms for Estimating Aboveground Biomass of Grassland in West Ujimqin, China. *Remote Sensing*, *15*(11), p.2918.

Zhao, F., Li, Y., Dai, X., Verhoef, W., Guo, Y., Shang, H., Gu, X., Huang, Y., Yu, T. and Huang, J., 2015. Simulated impact of sensor field of view and distance on field measurements of

bidirectional reflectance factors for row crops. *Remote Sensing of Environment*, *156*, pp.129-142.

Zhao, F., Dai, X., Verhoef, W., Guo, Y., van der Tol, C., Li, Y. and Huang, Y., 2016. FluorWPS: A Monte Carlo ray-tracing model to compute sun-induced chlorophyll fluorescence of three-dimensional canopy. *Remote Sensing of Environment*, *187*, pp.385-399.

Zhao, Q., Yu, L., Du, Z., Peng, D., Hao, P., Zhang, Y. and Gong, P., 2022. An overview of the applications of earth observation satellite data: impacts and future trends. *Remote Sensing*, *14*(8), p.1863.

Zhao, D., Zhen, J., Zhang, Y., Miao, J., Shen, Z., Jiang, X., Wang, J., Jiang, J., Tang, Y. and Wu, G., 2023. Mapping mangrove leaf area index (LAI) by combining remote sensing images with PROSAIL-D and XGBoost methods. *Remote Sensing in Ecology and Conservation*, *9*(3), pp.370-389.

Zhao, X., Liu, C., Qi, J., Yuan, L., Yu, Z., He, S. and Huang, H., 2024. Evaluating Different Crown Reconstruction Approaches from Airborne LiDAR for Quantifying APAR Distribution Using a 3D Radiative Transfer Model. *Remote Sensing*, *17*(1), p.53.

Zhao, X., Qi, J., Jiang, J., Liu, S., Xu, H., Lin, S., Yu, Z., Li, L. and Huang, H., 2024. Fine-scale retrieval of leaf chlorophyll content using a semi-empirically accelerated 3D radiative transfer model. *International Journal of Applied Earth Observation and Geoinformation*, *135*, p.104285.

Zheng, G. and Moskal, L.M., 2009. Retrieving leaf area index (LAI) using remote sensing: theories, methods and sensors. *Sensors*, *9*(4), pp.2719-2745.

Zhou, X., Huang, W., Kong, W., Ye, H., Dong, Y. and Casa, R., 2017. Assessment of leaf carotenoids content with a new carotenoid index: Development and validation on experimental and model data. *International Journal of Applied Earth Observation and Geoinformation*, *57*, pp.24-35.

Zhou, K., Cao, L., Shen, X. and Wang, G., 2023. Novel spectral indices for enhanced estimations of 3-dimentional flavonoid contents for Ginkgo plantations using UAV-borne LiDAR and hyperspectral data. *Remote Sensing of Environment*, *299*, p.113882.

Zhou, K., Cao, L., Yin, S., Wang, G. and Cao, F., 2022. An advanced bidirectional reflectance factor (BRF) spectral approach for estimating flavonoid content in leaves of Ginkgo plantations. *ISPRS Journal of Photogrammetry and Remote Sensing*, *193*, pp.1-16.

Zhu, C., Bai, C., Sanahuja, G., Yuan, D., Farré, G., Naqvi, S., Shi, L., Capell, T. and Christou, P., 2010. The regulation of carotenoid pigmentation in flowers. *Archives of Biochemistry and Biophysics*, *504*(1), pp.132-141.

Zhu, X., Yang, Q., Chen, X. and Ding, Z., 2023. An approach for joint estimation of grassland leaf area index and leaf chlorophyll content from UAV hyperspectral data. *Remote Sensing*, *15*(10), p.2525.

Zhuang, Y., Liu, X., Nguyen, T., He, Q. and Hong, S., 2013. Global remote sensing research trends during 1991–2010: a bibliometric analysis. *Scientometrics*, *96*, pp.203-219.