

https://dspace.org/

Estimating biophysical variables of the winter wheat crop using multispectral remotely sensed data

Zikhali, Siyabonga

2025
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/50304
Zikhali, Siyabonga. (2025). Estimating biophysical variables of the winter wheat crop using multispectral remotely sensed data [Master's dissertation, University of the Witwatersrand, Johannesburg]. https://hdl.handle.net/10539/50304

Downloaded from a DSpace repository. https://dspace.org/

Estimating biophysical variables of the winter wheat crop using multispectral remotely sensed data



A Research Report Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science (MSc) in Geographic Information Systems and Remote Sensing from the University of the Witwatersrand, Johannesburg

Siyabonga Zikhali (2108480)

GEOG 7000A

Supervisor: Prof Elhadi Adam

Plagiarism Declaration

I, Siyabonga Knowledge Zikhali, registered for an MSc in Geographic Information Systems (GIS) and Remote Sensing in 2024. I hereby declare that I understand that it is improper to utilise someone else's work without their consent and without giving them credit. I certify that the research report below is entirely original to me, and I have referenced other people's ideas and opinions per the required convention.

Date: 2025/08/17



Acknowledgement

Firstly, I would like to acknowledge my supervisor, Professor Elhadi Adam, for his supervision through this research project, which might not have been a success without his guidance.

Moreover, I would like to express my thanks to the South African National Space Agency (SANSA) for their support in terms of funding my studies and providing a platform for me to present and share my research project with students and experts within the Earth Observation sector. I also like to thank the University of the Witwatersrand for the Merit Bursary Award, which they awarded me towards the funding of my studies. Thus, I would like to sincerely thank both institutions that made it possible to continue with my studies in 2024.

Lastly, I want to express my appreciation to my friends and family for their encouragement and support throughout this project. I would particularly like to thank my mother and brothers for always supporting me and ensuring that I am in a good state of mind to conduct my research to the best of my abilities.

Abstract

Winter wheat is a vital crop for ensuring global food security. For example, wheat accounts for about 21% of the world's food production. However, its cultivation faces threats from environmental stress. Furthermore, significant climate changes, coupled with the necessity for sustainable agriculture and the global imperative to address food security issues, increase the urgency of adopting sustainable agricultural practices. This can be achieved by monitoring and estimating crop biophysical parameters such as leaf area index (LAI) and chlorophyll a and b (Cab), which offer valuable insights into crop health and development, helping farmers make better-informed decisions. Nonetheless, traditional and recent UAV-based crop monitoring methods have limitations, including the requirement for field presence and poor temporal resolution. Therefore, this study aimed to estimate biophysical parameters of winter wheat (i.e., LAI and Cab) using multispectral Sentinel-2 imagery.

The support vector regression (SVR) and random forest (RF) algorithms were utilised in conjunction with Sentinel-2 imagery to estimate the winter wheat LAI and Cab, and both models performed sufficiently well for most parts. However, the RF model performed better than SVR. RF LAI models exhibited an average predictive R^2 of 0.87, with RMSE of 0.11 and MAE of 0.09, while the SVR LAI model exhibited an average predictive R^2 of 0.63, with RMSE of 0.18 and MAE of 0.13. Similarly, the RF Cab model had an average predictive R^2 of 0.89, with RMSE of 1.24 and MAE of 0.95; while the SVR Cab model had an average predictive R^2 of 0.57, with RMSE of 1.96 and MAE of 1.29. As a result, RF was used to create estimation maps for LAI and Cab across different growth stages. These results indicated that open-access data from Sentinel-2 and machine learning algorithms, such as RF, provide economical means in terms of both time, computer, and financial costs to produce realistic estimations of crop variables such as LAI and Cab, even with minimal resources available.

Keywords: Random Forest, Machine Learning, Support Vector Regression, Leaf Area Index, Chlorophyll Content

Contents

PLAGIARISM DECLARATION II
ACKNOWLEDGEMENT III
ABSTRACT IV
LIST OF FIGURES VII
LIST OF TABLES VIII
LIST OF ABBREVIATIONS IX

CHAPTER 1: INTRODUCTION 1

1.1. BACKGROUND 1
1.2. PROBLEM STATEMENT 4
1.3. AIMS AND OBJECTIVES 5
1.4. RESEARCH OVERVIEW OR STRUCTURE OF THE PAPER 5

CHAPTER 2: LITERATURE REVIEW 6

2.1. INTRODUCTION 6
2.2. REMOTE SENSING FOR CROP MONITORING 6
2.2.1. RECENT TRENDS IN CROP MONITORING 6
2.2.2. OPPORTUNITIES AND LIMITATIONS OF TRADITIONAL SATELLITE-BASED CROP MONITORING 8
2.3. PERFORMANCE OF REGRESSION AND MACHINE LEARNING MODELS IN CROP MONITORING 11
2.4. FUTURE TRENDS IN SATELLITE-BASED CROP MONITORING 12
2.5. CONCLUSION AND SUMMARY 12

CHAPTER 3: MATERIALS AND METHODS 14

3.1. INTRODUCTION 14
3.2. STUDY AREA 14
3.3. FIELD MEASUREMENTS DATA 16
3.4. MULTISPECTRAL SATELLITE IMAGERY COLLECTION AND PROCESSING 17
3.5. PRE-PROCESSING OF MULTISPECTRAL IMAGERY 18
3.6. MODELLING AND STATISTICAL ANALYSIS 18
3.6.1. RANDOM FOREST (RF) 18
3.6.2. SUPPORT VECTOR REGRESSION (SVR) 19
3.6.3. HYPERPARAMETER TUNING AND MODEL SELECTION 19
3.7. TEMPORAL VARIATION IN WINTER WHEAT CROP BIO-PARAMETERS 21
3.8. FEATURE IMPORTANCE ANALYSIS 22
3.9. METHODOLOGICAL FLOWCHART 23

CHAPTER 4: RESULTS 24

4.1. INTRODUCTION 24
4.2. TRAINING ACCURACY 24
4.2.1. LAI: SVR VS RF RESULTS 24
4.2.2. CAB: SVR VS RF RESULTS 26
4.3. PREDICTIVE ACCURACY (VALIDATION) 27
4.4. TEMPORAL AND FEATURE DYNAMICS 35
4.4.1. TEMPORAL VARIATION ACROSS GROWTH STAGES 35
4.4.2. FEATURE IMPORTANCE AT EACH STAGE 36
4.4.2.1. LAI FEATURE IMPORTANCE 36
4.4.2.2. CAB FEATURE IMPORTANCE 39
4.5. PREDICTION MAPS 42
4.5.1. RF LAI PREDICTION MAPS 42
4.5.2. RF CAB PREDICTION MAPS 45

CHAPTER 5: DISCUSSIONS 47

5.1. INTRODUCTION 47
5.2. RF AND SVR MODEL'S PREDICTION PERFORMANCE 47
5.2.1. TRAINING ACCURACY 47
5.2.2. PREDICTIVE ACCURACY 48
5.3. TEMPORAL VARIATION AND FEATURE IMPORTANCE OF LAI AND CAB 50
5.3.1. TEMPORAL VARIATION IN LAI AND CAB 51
5.3.2. FEATURE IMPORTANCE 51
5.4. RF LAI AND CAB PREDICTION MAPS 53
CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS 55
6.1. INTRODUCTION 55
6.2. CONCLUSION 55
6.3. LIMITATIONS 56
6.4. RECOMMENDATIONS 56
REFERENCES 57

List of Figures

FIGURE 1: THE STUDY REGION IN DIPALESENG LOCAL MUNICIPALITY, FOUND IN GERT SIBANDE DISTRICT IN MPUMALANGA PROVINCE, SOUTH AFRICA. 16
FIGURE 2: METHODOLOGICAL FLOWCHART SUMMARISING THE STEPS FOLLOWED TO CONDUCT THE STUDY 23
FIGURE 3: THE COMPARISON OF ACTUAL AND PREDICTED BIO-VARIABLE OR PARAMETER, NAMELY LAI OF WINTER WHEAT, PRODUCED BY THE RANDOM FOREST (RF) AND SUPPORT VECTOR REGRESSION (SVR) MODELS IN CONJUNCTION WITH SENTINEL-2 IMAGERY IN THE FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY. 30
FIGURE 4: THE TEMPORAL VARIATION IN THE PREDICTION ACCURACY METRICS (I.E., RMSE, MAE AND R^2) OF THE RF LAI AND SVR LAI MODELS, IN THE FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY 31
FIGURE 5: THE COMPARISON OF ACTUAL AND PREDICTED BIO-VARIABLE OR PARAMETER, NAMELY CAB OF WINTER WHEAT, PRODUCED BY THE RANDOM FOREST (RF) AND SUPPORT VECTOR REGRESSION (SVR) MODELS, IN CONJUNCTION WITH SENTINEL-2 IMAGERY IN THE FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY. 34
FIGURE 6: THE TEMPORAL VARIATION IN THE PREDICTION ACCURACY METRICS OF THE RF CAB AND SVR CAB MODELS, IN THE FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY. 35
FIGURE 7: THE TEMPORAL VARIATION IN WINTER WHEAT BIO-VARIABLES OR PARAMETERS, NAMELY LAI AND CAB, IN THE FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY. 36
FIGURE 8: THE RELATIVE FEATURE IMPORTANCE AND CONTRIBUTION OF EACH FEATURE IN THE RANDOM FOREST (RF) AND SUPPORT VECTOR REGRESSION (SVR) ESTIMATION OF THE WINTER WHEAT CROP BIO-PARAMETER, NAMELY LAI AT DEPALESENG LOCAL MUNICIPALITY 39
FIGURE 9: THE RELATIVE FEATURE IMPORTANCE AND CONTRIBUTION OF EACH FEATURE IN RANDOM FOREST (RF) AND SUPPORT VECTOR REGRESSION (SVR) ESTIMATION OF THE WINTER WHEAT CROP BIO-PARAMETER, NAMELY CAB AT DEPALESENG LOCAL MUNICIPALITY 42
FIGURE 10: MAPS OF WINTER WHEAT CROP BIO-PARAMETER, NAMELY LAI, PRODUCED BY THE BEST MODEL, THE RANDOM FOREST (RF), IN CONJUNCTION WITH SENTINEL-2 IMAGERY IN A FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY. 44
FIGURE 11: MAPS OF WINTER WHEAT CROP BIO-PARAMETER, NAMELY CAB, PRODUCED BY THE BEST MODEL, THE RANDOM FOREST (RF), IN CONJUNCTION WITH SENTINEL-2 IMAGERY IN A FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY. 46

List of Tables

TABLE 1: THE COMPARISON OF TRAINING ACCURACY OF THE RF AND SVR MODELS IN COMBINATION WITH SENTINEL-2 IMAGERY IN ESTIMATING WINTER WHEAT CROP BIO-PARAMETER, NAMELY LAI, IN A FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY. 25
TABLE 2: THE COMPARISON OF TRAINING ACCURACY OF THE RF AND SVR MODELS IN CONJUNCTION WITH SENTINEL-2 IMAGERY IN ESTIMATING WINTER WHEAT CROP BIO-PARAMETER, NAMELY CAB, IN A FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY. 27
TABLE 3: THE WILCOXON SIGNED-RANK TEST STATISTICS SHOWING WHETHER THE DIFFERENCE IN THE PREDICTION ACCURACY OF THE RF LAI AND SVR LAI MODELS WAS STATISTICALLY SIGNIFICANT ACROSS THE DIFFERENT WINTER WHEAT GROWTH STAGES IN THE FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY. 31
TABLE 4: THE WILCOXON SIGNED-RANK TEST STATISTICS AT 95% CONFIDENCE LEVEL SHOWING WHETHER THE DIFFERENCE IN THE PREDICTION ACCURACY OF THE RF LAI AND SVR LAI MODELS WAS STATISTICALLY SIGNIFICANT ACROSS THE DIFFERENT WINTER WHEAT GROWTH STAGES IN THE FARM PLOT AT DEPALESENG LOCAL MUNICIPALITY.....35

List of Abbreviations

Acronyms	
RF	Random Forest
SVR	Support Vector Regression
SVM	Support Vector Machine
LAI	Leaf Area Index
Cab	Chlorophyll a and b
LCC	Leaf Chlorophyll Content
UAV	Unmanned Aerial Vehicles
R^2	Co-efficient of Determination
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
NDVI	Normalized Difference Vegetation Index
GCHL	Green Chlorophyll Index

Chapter 1: Introduction

1.1. Background

Winter wheat is one of the most essential cereal crops for ensuring global food security, contributing about 21% to the worldwide food production (Shiferaw et al., 2013; Cheng et al., 2024). However, its cultivation and production are threatened by environmental stress, including climate variability and change, environmental degradation, and poor agricultural practices, all of which reduce yields and compromise food security (H. Zhang et al., 2023). These challenges, coupled with population growth and the need for sustainable agricultural practices, highlight the importance and urgency of winter wheat research to find efficient ways to produce more with minimal environmental impact (Kanning et al., 2018; C. Zhang et al., 2024).

To this end, precision agriculture remains one of the viable solutions and practical responses to these issues (Kganyago et al., 2024). For example, in regions such as Southern Africa, where approximately 70% of the population relies on agriculture for their livelihoods (Sibanda et al., 2023) effective crop management is essential. This can be achieved by monitoring various crop parameters, such as leaf biophysical properties, which are crucial for sustainable agriculture (Bhadra et al., 2024). Additionally, accurate and timely monitoring of these leaf biophysical properties helps provide reliable yield estimates and supports efforts to ensure food security (Zhu et al., 2020). Furthermore, non-destructive approaches for estimating biophysical variables such as the leaf chlorophyll content (LCC) of winter wheat can guide or direct field tillage and agricultural practices, helping produce wheat with higher yields and better quality, ultimately improving the overall standard of agricultural production (Wang et al., 2022).

The leaf area index (LAI), which can be defined as the entire one-sided leaf area divided by the ground surface area (LI et al., 2017; Chen et al., 2022), and the leaf chlorophyll content (LCC), which is a pigment that is vital to the process of photosynthesis in plants (Gao et al., 2022; Singhal et al., 2024), are among the some of the most widely used biophysical variables in crop monitoring. Furthermore, timely and accurate assessment of vegetation characteristics such as LCC, LAI, and canopy water content (CWC) can improve crop health monitoring and maximise yield potential (Verma et al., 2022). For example, LCC is a vital measure or indicator of photosynthetic capacity and plant physiological status, providing valuable insights into

biotic and nutritional stresses, particularly nitrogen deficiency (Yan et al., 2024; Singhal et al., 2024). Additionally, it is also used to determine crop nitrogen fertiliser requirements while minimising nitrogen losses to soil (Zhu et al., 2020). Meanwhile, LAI is closely linked to the vegetation structure, crop development dynamics, and ecological processes at scales from individual stands to global systems (Bajocco et al., 2024). Therefore, together, LAI and LCC are essential for biophysical modelling used in growth and yield prediction, as they reveal valuable insights about crop health status at different growth stages (Bajocco et al., 2024; Jin et al., 2025). However, the approaches for measuring these crop biophysical parameters have evolved over the years, with the advancements in technology. Particularly, the availability of satellite imagery has revolutionised crop monitoring and the remote sensing field.

In the past, field-based instruments were used to estimate crop biophysical variables (i.e., LAI and LCC); moreover, these in-situ field surveys provided accurate and reliable results (Singhal et al., 2024). However, these methods or approaches are generally limited or applicable to small areas, are time-consuming and labour-intensive, and can be destructive at times (Han et al., 2021; Yue et al., 2024). Moreover, the measurements often need to be taken in the afternoon when the sun is almost or directly overhead (Choudhury et al., 2019). Finally, the high costs in terms of time and resources, alongside resolution limitations inherent in these field-based methods, make large-scale crop health monitoring challenging, especially given the rapid changes in crop conditions.

To overcome these issues, crop monitoring has evolved in recent years with increasing use of manned and unmanned aerial vehicles (UAVs), which provide greater efficiency, higher spatial and spectral resolution, and more frequent data collection (Shen et al., 2024). Further, remote sensing technology, which includes airborne platforms, satellites, and handheld spectral equipment usually used on-site, has been proven successful in tracking crop growth and health over the years (Yao et al., 2017; Han et al., 2021). However, despite this, traditional satellite-based monitoring faces various constraints, such as cloud cover and limited spatiotemporal resolution (Han et al., 2021), while UAVs and field-based approaches require physical access to sites and are less suitable for very large-scale and frequent crop monitoring (Xiao et al., 2024).

As a result, optical satellites such as Sentinel-2 and Landsat remain some of the most widely used sensors for crop monitoring across various regions of the world; moreover, these satellites

are especially useful in commercial agriculture, where one is dealing with large-scale monitoring. Furthermore, these satellite methods or approaches are increasingly complemented by non-destructive spectral techniques that lessen the need for destructive sampling (Křížová et al., 2022). For example, instead of collecting samples from the entire field, carefully selected reference samples are used to calibrate and validate satellite-based monitoring, combining the advantages of ground truth and remote sensing, allowing for more accurate estimation of crop biophysical parameters and thus crop health status.

1.2. Problem Statement

Monitoring the biophysical and biochemical parameters of winter wheat, specifically leaf area index (LAI) and leaf chlorophyll content (Cab) as a proxy for nitrogen, is essential for assessing crop health, enhancing yield estimates, and enabling precision agriculture (Peng et al., 2019). Conventional uniform fertiliser practices often overlook spatial variability in soil and topography, leading to economic losses and environmental impacts (Revill et al., 2019). Additionally, traditional field-based methods, such as the SPAD-502 chlorophyll meter and LI-COR Plant Canopy Analyser, provide accurate measurements (Han et al., 2021); however, these methods are labour-intensive, time-consuming, and unsuitable for large-scale, frequent monitoring (Mercier et al., 2019).

Although UAVs and satellites enhance spatial coverage and revisit frequency, UAVs still require field access, and traditional satellites often suffer from cloud cover and coarse spatiotemporal resolution (Han et al., 2021). Sentinel-2, with its 5-day revisit cycle, red-edge bands, and free access, holds significant potential for monitoring crop biophysical parameters (Ienco et al., 2019; Madonsela et al., 2023). Nevertheless, there is no standardised, validated approach for estimating LAI and Cab from Sentinel-2 in specific agricultural contexts such as African winter wheat systems. Addressing this gap could optimise fertiliser and water use through variable rate applications, improving both productivity and environmental sustainability. Therefore, this study evaluates the effectiveness of Sentinel-2 data and machine learning algorithms in predicting LAI and Cab at the plot level, aiming to develop a scalable, reliable framework for winter wheat crop monitoring.

1.3. Aims and Objectives

The study aims to estimate biophysical parameters of winter wheat, specifically the leaf area index (LAI) and chlorophyll a and b (Cab), using multispectral remote sensing imagery. The specific objectives are:

- To estimate LAI and Cab using Sentinel-2 multispectral imagery with Support Vector Regression (SVR) and Random Forest (RF) machine learning algorithms.
- To examine the changes in winter wheat crop parameters across different growth stages and identify the most important features for estimation of crop parameters (i.e., LAI and Cab) of winter wheat at various growth stages using Sentinel-2 satellite imagery.
- To apply the best-performing regression model to the multispectral remote sensing images to create estimation maps for LAI and Cab.

1.4. Research Overview or Structure of the Paper

This research report has six chapters; the first chapter above provides an overview of the study's background, goals, and motivations. The second chapter examines the literature review, which summarises the findings of earlier research and explains how the methodological approach of those studies might be used in the current investigation. The third chapter describes the research region, its location, and the methodology used to obtain the results. In addition, the fourth chapter examines the study's results, which are shown in tables and figures along with a brief synopsis or explanation. The results are discussed in the fifth chapter, which also examines their applicability by contrasting them with findings from earlier studies. The final part, chapter six, offers the conclusions that can be drawn from this research, along with its limits and suggestions for enhancing its findings.

Chapter 2: Literature Review

2.1. Introduction

This literature review was conducted using a narrative search approach, synthesising relevant studies to summarise key findings and methodological trends in the field. The review is organised into thematic segments. The first segment gives a brief overview of the evolution of optical satellites, and then the second section delves into the recent trends in crop monitoring. The third section looks at the use of optical satellites such as Sentinel-2 in crop monitoring. Furthermore, the fourth section provides details about the use of various machine learning regression models, such as Random Forest (RF) and Support Vector Regression (SVR), in crop monitoring, while the second-to-last section looks at the expected future trends in remote sensing. Finally, the last section provides a summary of the literature reviewed.

2.2. Remote Sensing for Crop Monitoring

Crop monitoring has evolved over the years with the improvement in technology and capabilities of satellite and UAV platforms. Thus, crop monitoring transitioned from traditional approaches to current satellite and UAV-based methods. The transition across time from traditional methods has been because there are various constraints or limitations associated with the field-based methods of crop monitoring that have been outlined in the introduction; these include labour costs and covering small spatial areas.

2.2.1. Recent Trends in Crop Monitoring

Unmanned Aerial Vehicles (UAVs) offer various capabilities, including higher spatial resolution and a good, flexible revisit time (Kganyago et al., 2024; Shen et al., 2024), as this is determined by the user or analyst based on their specific needs. These capabilities overcome the shortcomings associated with most satellites that have been or are being used for crop monitoring. For example, it is often difficult to match the date of reference sample collection and the satellite imagery taken over the study area because of challenges, including cloud cover contamination. This further makes it difficult to monitor crops at the field level or scale using these satellite platforms; thus, as a result, the nearest cloud-free imagery is usually used (Tomíček et al., 2022) instead of the imagery matching the actual field sampling date. In

addition, UAVs overcome the revisit time and spatial resolution limitations of most traditional satellite imagery.

Thus, there has been an increase in studies focusing on the use of UAVs in crop monitoring in recent years, especially in Asian and European nations. Moreover, these studies have focused on the use of both multispectral and hyperspectral UAV-acquired imagery or data, because they offer various opportunities when it comes to crop monitoring. Furthermore, these studies in recent years have demonstrated the effectiveness of UAV-based data in estimating the crop's various parameters. For example, the study by (Han et al., 2021), which aimed to predict various crop variables including LAI, chlorophyll, and yield at various growth stages of winter wheat, also found that multivariate linear regression in conjunction with UAV data accurately predicts LAI, chlorophyll content, and yield under different water treatments.

Meanwhile, the study by Sahoo et al. (2024), which aimed to investigate the benefits of employing hybrid models to estimate crop characteristics using hyperspectral, high-resolution UAV data, further offered cost-effective methods for determining farmland conditions to promote improved farm management techniques, with all the models used in the study having fairly good validation accuracies. For example, Sahoo et al. (2024) obtained calibration coefficient of determination (R^2) ranging from 0.968 to 0.990 and validation R^2 ranging between 0.948 and 0.990 for leaf area index (LAI), leaf chlorophyll content (LCC), and canopy chlorophyll content (CCC) estimation, additionally, the Extreme Learning Model (ELM) outperformed other models used in the study including RF, ANN, SVR and MARS in the prediction of all three variables (i.e., LAI, LCC and CCC).

Moreover, the research by C. Zhang et al. (2024), which aimed to create a machine-learning regression model and a modified variable selection strategy for estimating winter wheat's bio parameters at different growth stages using multispectral UAV imagery, attained fairly good accuracies also with a coefficient of determination (R^2), root mean square error (RMSE), and relative predictive deviation (RPD) of 0.967, 0.225, and 4.905 at the early filling stage, 0.912, 2.711, and 2.872 at the heading stage, and 0.968, 0.147, and 5.279 at the booting stage, respectively, with the regression model built using sequential forward selection in combination with support vector regression (SFS-SVR) for LAI, LCC, and CCC having the best predictive accuracy. This study by C. Zhang et al. (2024) further serves as a guide for

monitoring and predicting winter wheat bio-parameters using multispectral UAV imagery during specific crop phenology periods.

Thus, these studies demonstrate that the UAV-based crop monitoring approach, which is one of the approaches being used a lot in crop monitoring in recent years, has potential in the field of precision agriculture. However, the use of UAVs in crop monitoring is still premature in certain regions of the world and introduces some challenges, which include the expensive nature of the use of these platforms when compared to satellite platforms such as Sentinel-2, which offer fairly good quality open-source data for crop monitoring at the field or farm scale.

Moreover, unlike satellites, which cover large spatial areas, UAV platforms usually cover a smaller spatial extent and have limited flying times due to battery limitations (Risbøl & Gustavsen, 2018; Shahrooz et al., 2020). Additionally, there is a negative relationship between payloads and the duration of drone flights (Singh et al., 2023). Moreover, they must be within a certain distance (i.e., the line of sight) from the person piloting them for safety and regulatory purposes (Risbøl & Gustavsen, 2018). Furthermore, UAVs have a major limitation, which is the fact that they have poor temporal resolution in terms of going back in time and continuously monitoring crops. This is probably one of the major challenges of using these platforms, especially if one is mainly focusing on the evolution of the farmland or landscape over time. For these reasons, satellite-based crop monitoring remains a viable solution in most regions, especially Africa, and research on the use of these satellite platforms is still important.

2.2.2. Opportunities and Limitations of Traditional Satellite-Based Crop Monitoring

Research on the use of satellite-borne multispectral sensors to monitor crops increased drastically after the launch of optical satellites such as Landsat and, more recently, Sentinel-2. Moreover, various studies have demonstrated that satellites can still be utilised successfully for crop monitoring. For example, Xiao et al. (2024) used Sentinel 2 data together with machine learning and deep learning algorithms to develop models for estimating winter wheat yield at the field scale. The models attained satisfactory results, especially the Attention-based One-dimensional Convolutional Neural Network (ACNN) model. In addition, the study by Zhou et al. (2020), which aimed to evaluate the suitability of Landsat-8 OLI data for LCC modelling using various retrieval techniques, concluded that Landsat-8 Operational Land Imager (OLI)

data is appropriate for estimating crop LCC and may serve as a foundation for LCC calculation when combined with comparable multispectral datasets.

Meanwhile, H. Zhang et al. (2023) used the combined or integrated Landsat 8 and Sentinel-2 vegetation indices and ML algorithms to predict winter wheat yield data. The findings showed that the wide dynamic range vegetation index (WDRVI) reflected the growth of wheat with more realism and the best-performing model BO-CatBoost model producing an R^2 of 0.7, 0.63, and 0.68, and the RMSE values of 0.2, 0.73, and 0.62 with other models used in this study including RF, SVM and least absolute shrinkage and selection operator (LASSO). Moreover, the study by Xie et al. (2019) assessed the potential or viability of Sentinel-2 imagery, in particular the red-edge bands, for estimating the winter wheat bio-parameters (i.e., LAI and LCC) in Shunyi District, Beijing, China. They noted that Sentinel-2 imagery, in particular, the red-edge bands, were suitable and improved the estimation of crop (i.e., winter wheat) bio-parameters, LAI and LCC. Furthermore, the R^2 accuracies from the artificial neural network (ANN) model for LAI, LCC, and CCC were 0.5, 0.5, and 0.72, respectively.

Moreover, various studies have also been conducted in various regions of Africa to predict the bio-variables of crops. For example, the study by Adeluyi et al. (2021), which aimed to assess how well two hybrid inversion techniques work with Sentinel-2 to determine rice LAI in Nasarawa State, North-central Nigeria. The results of this study demonstrated that when it came to obtaining LAI values from Sentinel-2 imagery, the hybrid Gaussian processes regression (GPR) model's overall prediction accuracy with R^2 of 0.82 and RMSE of 1.65 outperformed the hybrid ANN model's with R^2 of 0.66 and RMSE of 3.89, their findings further demonstrated how hybrid GPR models can be used to estimate the phenological dynamics of LAI in irrigated rice using Sentinel-2 data. Compared to hybrid ANN models, they offered a more precise estimation of LAI patterns from various nitrogen and water applications.

Another study in Nigeria by Ekwe et al. (2024) utilised Sentinel-2 to estimate the LAI of rainfed groundnuts at the seedling stage by methodically evaluating the performance of machine learning regression algorithms (MLRAs) and VIs models in the Gwagwalada Area Council of the Federal Capital Territory, North Central, Nigeria. The study tested ten machine learning regression algorithms (MLRAs), however, three models namely Gaussian processes regression (GPR) with R^2 of 0.73 and RMSE of 0.81, together with Kernel ridge regression

(KRR) with R^2 of 0.72 and RMSE of 0.82, and Support vector regression (SVR) with R^2 of 0.70 and RMSE of 0.85 were found to be the most suitable for estimating LAI for rainfed groundnut. Moreover, Chemura et al. (2017), assessed how well Sentinel-2 MSI data performed when used to predict and map leaf Chl in coffee plants at the landscape scale at Jersey Tea and Coffee Estates in Chipinge district, Zimbabwe, and concluded or found that Sentinel-2 MSI is a useful dataset for predicting coffee Chl, and recommended that coffee Chl be modelled using finer spatial resolutions of 10 m applied to mature coffee stands.

In the context of South Africa (SA) studies, such as the one by Kganyago et al. (2021), which used Sentinel-2 imagery with machine learning algorithms, to assess the accuracy of various machine learning algorithms in estimating crop biophysical parameters in their study site at Bothaville in the Free State province of SA, managed to get good estimations of LAI, LCC and CCC, especially with the random forest (RF) algorithm with an RMSE of 0.5 and R^2 of 0.76. Meanwhile, the study by Madonsela et al. (2023) tested the assumptions that the red-edge band above 730 nm may therefore help measure LAI during peak development stages of maize when the chlorophyll content is high, as the reflectance beyond 730 nm would then be greatly influenced by the chlorophyll content by using Sentinel-2 spectral data to assess these crop bio-parameters at different growth stages on maize farms located in Vanderbijlpark, Bronkhorstspuit and Magaliesburg in Gauteng South Africa. The study found that Sentinel-2 data can be used to monitor maize biophysical factors at the satellite level, especially with the use of red-edge bands, which yielded an R^2 of 0.65 and 0.63 during the early stages and an R^2 of 0.79 and 0.81 during the tassel stage.

These studies demonstrate the power and applicability of optical multispectral satellites such as Sentinel-2 in crop monitoring; thus, they can be used for monitoring crops in most African countries with limited access to high spatial resolution data. Additionally, with improvements in open-access, freely accessible satellites and refinement of machine learning and deep learning algorithms, crop monitoring at the field scale will be easier and more precise. Moreover, these satellites and machine learning algorithms seemed to be versatile and flexible, and thus have been utilised with success and good accuracy across different regions of the world, thus providing opportunities for crop monitoring even here in South Africa.

2.3. Performance of Regression and Machine Learning Models in Crop Monitoring

Various machine learning algorithms have been utilised successfully in estimating crop parameters. For instance, the study by C. Zhang et al. (2024) to estimate winter wheat biophysical parameters at different growth stages, used models such as Random Forest (RF) regression, Support Vector Regression (SVR) and Least Absolute Shrinkage and Selection (LASSO) regression and their findings demonstrated that RF regression had better accuracy but had overfitting models, while SFS-SVR was robust and better estimated the wheat crop bio-parameters at different growth stages. Moreover, the study by Zhu et al. (2020) to predict or estimate the leaf chlorophyll content (LCC) of crops, used multivariable linear regression (LM), backpropagation neural network (BP), support vector machine (SVM), and random forest (RF); all these algorithms had fairly good accuracies when it came to estimating LCC; however, RF and SVM had slightly better accuracies.

Meanwhile, Yuan et al. (2017) found that RF is more appropriate for the prediction of LAI at large scales, for example, when the sample plots are large and have a lot of variation, which usually happens when focusing on the whole growth period from the emergence stage to the harvesting stage. However, when the sample plots and variations are minimal, the Artificial Neural Network (ANN) is more suitable for predicting LAI (Yuan et al., 2017). Furthermore, Kganyago et al. (2021) contrasted the performance of sparse Partial Least Squares (SPLS), Gradient Boosting Machines (GBM), and RF for predicting crop bio parameters such as LAI, LCC, and CCC. The study found that RF had better overall accuracy when compared to other algorithms.

Overall, these findings show that many machine learning regression algorithms have been effectively used to predict various crop bio-parameters. However, RF and SVM models appear to be some of the most commonly employed for predicting crop parameters. Furthermore, these two algorithms typically perform better than other statistical or regression models when estimating different crop bio-parameters. Additionally, the widespread use of these algorithms across different regions worldwide demonstrates their versatility and flexibility, providing a useful reference for comparing results and ensuring proper model calibration.

2.4. Future Trends in Satellite-Based Crop Monitoring

UAV imagery has emerged as one of the most widely adopted crop monitoring tools in recent years due to its high spatial resolution and flexible deployment. However, a key limitation is its temporal resolution, making it less suited for continuous, long-term monitoring. For extended seasonal or multi-year observation, satellite platforms remain the more viable option. Satellite technology and sensors have steadily advanced since the launch of early environmental monitoring missions such as Landsat in the 1970s (Kganyago et al., 2024), and further improvements in both spatial and temporal resolution are anticipated. These advances are expected to enhance the precision of crop monitoring and facilitate the seamless integration of satellite imagery with UAV data, enabling multi-scale analyses with faster processing and greater accuracy. Coupled with improvements in hardware, software, and analytical algorithms, such integration will likely deliver more timely, refined, and actionable insights for crop management.

Furthermore, cloud platforms such as Google Earth Engine (GEE) eliminate the need for extensive storage and high computational power because satellite data is available within the platform, and the processing burdens are handled by GEE. Additionally, the availability of algorithms such as RF and SVM, along with readily accessible satellite imagery, can enable real-time or near-real-time monitoring of crop bio-parameters. Lastly, improvements in statistical algorithms can also facilitate transferable learning across different sites, thereby reducing the need to go directly to the field to collect reference samples.

2.5. Conclusion and Summary

Studies in the past used field-based methods to monitor crops; however, these methods had certain limitations, including the area that could be covered or monitored. These limitations resulted in the rapid development of remote sensing based on multispectral satellites and, more recently, the UAV-based crop monitoring methods. The use of satellite-based methods to monitor crops allows for coverage of larger spatial areas; however, these satellites have certain limitations, which include coarse resolution in terms of revisit time and spatial resolution.

These limitations of satellites resulted in the development and use of UAV-based methods of crop monitoring, which offer finer spatial resolutions and revisit time, thus allowing for better and more accurate estimations of various crop bio-parameters. Additionally, the UAV-based methods have improved the accuracy of crop bio-parameters (i.e., LAI and Cab) estimation, thus allowing for timely information that farmers can use to better manage their farms. However, the UAV-based approaches came with various limitations, including poor temporal resolution in terms of going back, costs in terms of both labour and economic costs, and finally, the UAV-based methods, just like the traditional methods, require one to go to the field. Thus, multispectral satellites such as Sentinel-2 remain a viable solution for crop monitoring in most regions of the world.

Chapter 3: Materials and Methods

3.1. Introduction

This segment looks at the methodological approach and materials used in this study. The first part provides the description and map of the study area, together with information about the topography and the overall picture of the importance of agriculture in the study region, which is the Dipaleseng Local Municipality. Then, the second section to the second last section describes how the data used in the study was acquired, both the in-situ measurements and the satellite data. Moreover, it also looks at and describes the models used in the study and how the measured samples were utilised to train the models. It then looks at how the accuracy assessment was done to check the validity of the models and estimations. Furthermore, it then looks at the temporal variation in the crop parameters and the feature importance. The final part provides the methodological flow chart showing how the data was collected and pre-processed, including all the steps followed to get the results.

3.2. Study Area

The study area is a farm plot located in the Dipaleseng Local Municipality within Gert Sibande District Municipality, Southwest of Mpumalanga Province, South Africa (Figure 1). The Dipaleseng Local Municipality is located at latitudes: -26.500890 to -27.066960 and longitudes: 28.243465 to 28.987960 at an average altitude or elevation of 1574 m and a minimum altitude of 1478 m (Dipaleseng Local Municipality Topographic Map, Elevation, Terrain, n.d.). Additionally, the Dipaleseng municipality covers an area of about 2619.4 km^2 and plays a huge role in agricultural services, as agriculture is one of the main land uses within this municipality (Zama, 2013). For example, the Gert Sibande District utilises most of its land for agricultural purposes, with approximately 22% of the total land area dedicated to crop cultivation (Dipaleseng Local Economic Development Strategy, 2011). In addition, Dipaleseng contributed 70% to the Gert Sibande agriculture sector in 2009 (Dipaleseng Local Economic Development Strategy, 2011); however, agricultural income has been decreasing (Zama, 2013).

Moreover, the main crops being cultivated in Dipaleseng include maize, sunflower, soya beans, and various animal feed (Phase 2: Spatial Challenges and Opportunities Report, 2019). Dipaleseng is located in the Highveld, which is made up of wooded hills and lush grasslands. The region's game farms are home to a wide variety of game species as well as birds. Furthermore, in terms of climate, Dipaleseng usually experiences variable temperatures with the highest average temperatures of 20.5°C to 25 °C being recorded in summer around January, while the lowest temperatures of 8.4 °C are experienced in June (Phase 2: Spatial Challenges and Opportunities Report, 2019). In addition, the area often experiences severe cold conditions and frost, which results in wildlife and crops being damaged. Thus, with warm summers and cold winters, this area usually experiences summer rainfall, with most regions in the municipality getting an average yearly rainfall between 595-794 mm (Phase 2: Spatial Challenges and Opportunities Report, 2019). Finally, the fairly substantial rainfall and availability of water lead to various economic opportunities for agricultural services dependent on both rainfall and irrigation, especially in areas located near the Vaal River banks (Phase 2: Spatial Challenges and Opportunities Report, 2019). This is one of the reasons why agriculture is one of the main services within this municipality.

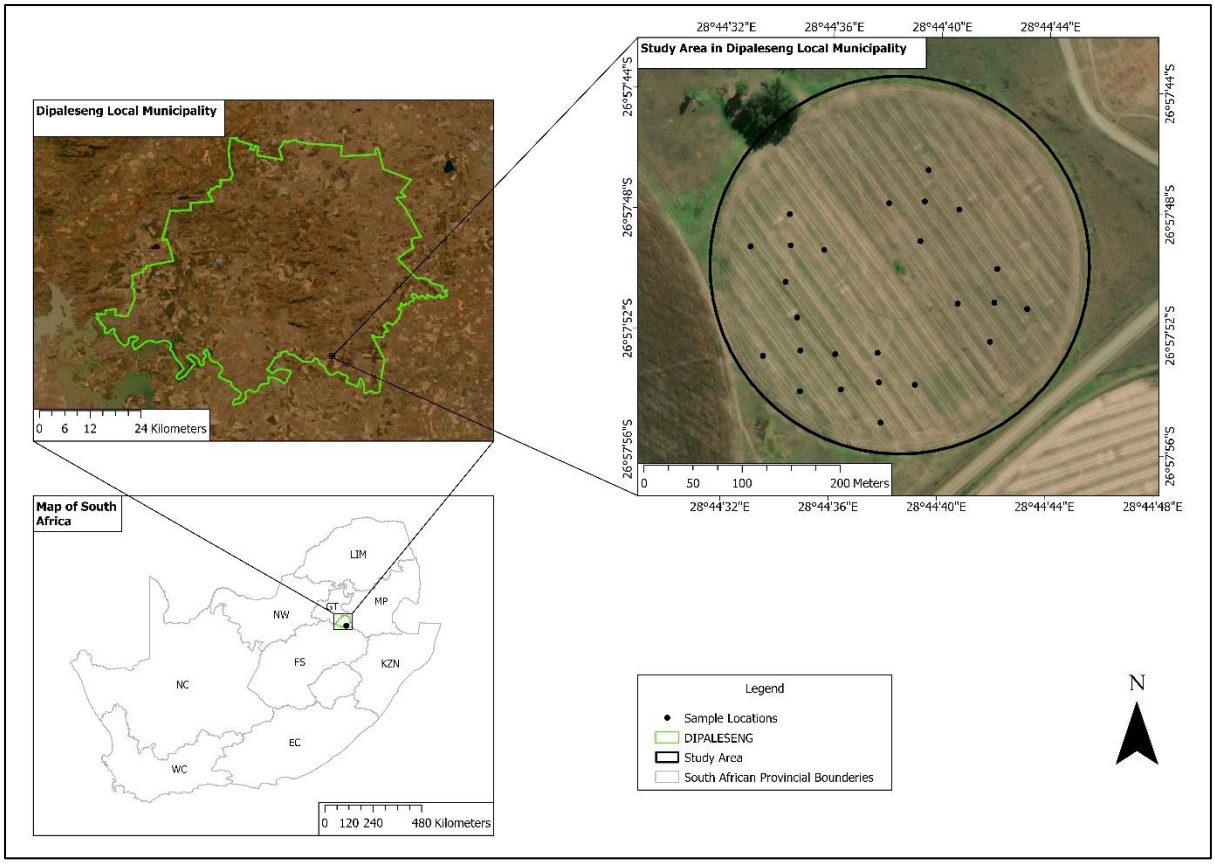


Figure 1: The Study region in Dipaleseng Local Municipality, found in Gert Sibande District in Mpumalanga Province, South Africa.

3.3. Field Measurements Data

The LAI and Cab measurements were collected in situ in the field. First, the Cab was measured using a portable leaf chlorophyll meter, while the LAI was measured with the LI-22000 C Plant Canopy Analyser. Five 50-by-50-meter plots were established in each field. From each plot, five 10-by-10-meter subplots were selected to gather representative data. Five separate measurements were taken and averaged to accurately reflect the conditions within each subplot. This method ensured a comprehensive and reliable dataset that correctly represented each plot's environmental conditions. Consequently, approximately 25 measurements or samples were collected from the portable leaf chlorophyll meter and the LI-22000 C Plant Canopy Analyser. The samples were taken on different days and at various growth stages. For instance, the winter wheat crop has four growth stages: jointing, booting, heading, and filling (Zhang et al., 2024).

The first field visit was on the 20th of August 2022, and the samples were then collected weekly after the first sample, with the second sample being collected on the 27th of August 2022, while the third field visit was on the 03rd of September 2022. The fourth field visit was on the 11th of September 2022, then the fifth field visit was on the 18th of September 2022, and the sixth visit was on the 25th of September 2022. Then, the seventh sample was collected on the 03rd of October 2022, with the eighth sample being collected on the 18th of October 2022, and the ninth sample was collected on the 25th of October 2022. The tenth sample was collected on the 03rd of November 2022, and the last sample was collected on the 09th of November 2022. Thus, the samples collected cover or account for the four main growth stages as outlined by C. Zhang et al. (2024) namely the first stage right after emerging (jointing stages), the second stage, which is the vegetative or booting stage, the third stage, which is the heading stage, and finally, right before harvest, the grain filling stage.

3.4. Multispectral Satellite Imagery Collection and Processing

The Sentinel-2 multispectral surface reflectance imagery, which has 13 spectral bands, was used for the prediction of crop bio-parameters. The Sentinel-2 satellite comprises two twin satellites, Sentinel-2A and Sentinel-2 B (Ienco et al., 2019). The revisit time of Sentinel-2 is 5 days because these two twin satellites make it possible to monitor crops more regularly, especially when there are clear skies and no interferences to the image collection. Additionally, Sentinel-2 imagery also has bands that are separated as follows: the 10-meter resolution bands are blue (B2), green (B3), red (B4), and near-infrared (B8); the 20-meter resolution bands are the red edge (B5), short-wave infrared (SWIR) bands B11 and B12, and near-infrared (NIR) bands B6, B7, and B8A (Shoko et al., 2020). Lastly, the coastal aerosol (B1) and cirrus band (B10) have a 60-meter spatial resolution. The aerosol band 1 and the cirrus band 10 were not included in the study because they do not contribute much to vegetation or crop monitoring.

Furthermore, the Sentinel-2 imagery has a wavelength or bandwidth ranging from 443 nm to 2190 nm. These Sentinel-2 images, which are offered by Copernicus, were collected from the Google Earth Engine data catalogue. Moreover, the multispectral Sentinel-2 imagery was collected on the same dates or at the closest dates possible to when the Cab and LAI in situ field measurements were collected using precise hand-held field equipment. Moreover, Sentinel-2 imagery has the red-edge bands, which are usually used for vegetation monitoring,

namely bands 5, 6, and 7, which gives it an edge over most freely accessible optical multispectral satellites when it comes to crop monitoring.

3.5. Pre-processing of Multispectral Imagery

The imagery was pre-processed to ensure that it was suitable for use in the analysis in the current study. The main pre-processing that was conducted included image mosaicking or harmonisation, and the extraction of the region of interest B. Zhang et al. (2024). These are the most crucial pre-processing steps that need to be taken to make sure that the data is ready for analysis. Then, the region of interest and the sample of reflectance values were extracted from the Sentinel-2 imagery; however, there were times when the imagery was not available, and the images were collected from the closest cloud-free date with imagery available.

3.6. Modelling and Statistical Analysis

The Random Forest and Support Vector Regression models were utilised in the R-Studio software, using various libraries including the randomForest library and Support Vector Regression library (i.e., e1071). The Machine Learning Regression Algorithms (MLRA) used in this study were selected based on their extensive use in previous studies and the satisfactory results these algorithms achieved in previous studies. Moreover, the extensive use of these models in previous studies provides a good reference for comparison and calibration of the models to ensure that the results are more reliable. The Support Vector Machine (i.e., support vector regression specifically) and Random Forest algorithms are both nonparametric approaches that can be used for both classification and regression problems. These two algorithms were contrasted, and the best-performing algorithm was discussed in detail in the results and discussion section. Additionally, the best-performing algorithm was applied to the acquired multispectral imagery to create LAI and Cab estimation maps, which showed the variation in LAI and Cab across the field and time.

3.6.1. Random Forest (RF)

The RF regression is a prediction or estimation algorithm that fits multiple decision trees to different dataset subsamples, and it then averages the results to increase the predicted accuracy

(Chen et al., 2022; Sahoo et al., 2024). Moreover, it predicts a variable value by combining the outputs of several decision trees. Additionally, the RF can work with highly dimensional data, and it has an advantage in that it limits or reduces over-fitting (Yuan et al., 2017). Thus, this model is usually used for predicting the surface characteristics of crops and soils.

3.6.2. Support Vector Regression (SVR)

The SVR is a prediction or estimation algorithm that converts a nonlinear problem into a linear one (Sahoo et al., 2024). By using kernel functions like radial, polynomial, and linear functions. Moreover, SVR provides flexibility in determining the best hyperplane in a higher dimension to fit the data. The primary goal of this model is to reduce error by customising the hyperplane to maximise the margin while maintaining a reasonable error tolerance.

3.6.3. Hyperparameter Tuning and Model Selection

The hyperparameters for both the RF and SVR models were tuned; for example, the number of decision trees was tuned using grid search for RF, thus these varied across the different growth stages based on the optimal combination chosen based on the RMSE, along with MAE and R^2. While the kernel type (i.e., radial, polynomial, and linear) was tuned for SVR, moreover, the cost and gamma were specified for the SVR model based on the grid search method until the optimum results were obtained. The best model was selected automatically based on the RMSE, along with the MAE and R^2 values.

Firstly, both the RF and SVR models had multiple predictors, and the RMSE, MAE, and R^2 were used to choose the best predictor model for each model at each collection period. The predictor model with the smallest RMSE was selected as the best model since RMSE provides context about the error between the estimated and actual values. However, the process of selecting the best model was automated for both RF and SVR. Then the best RF and SVR models, which were carefully chosen based on the root mean square error (RMSE), R^2, and MAE, were compared. These two models were further used to compare the actual versus predicted values across time or at different sampling periods to determine which model was the best in terms of prediction. The best model was ultimately used to create LAI and Cab estimation maps across the plot and growth stages.

3.6.4. Accuracy Assessment Metrics and Tests

To assess the model accuracy in terms of accurately estimating the values of the winter wheat bio-parameters (i.e., LAI and Cab), the mean absolute error (MAE), coefficient of determination (R^2), and root mean square error (RMSE) were used (C. Zhang et al., 2024). A vast number of studies have utilised these measures to evaluate the accuracy of the regression or machine learning algorithms, such as RF and SVR, thus making it easier to compare the accuracies of the studies. Moreover, to accomplish this, the data samples collected from the field were used for training the models. However, instead of using the 70% to 30% split of the samples for training the model, and 30% of the samples to test and evaluate the accuracy of the models, all samples were utilised for training. The collection of 25 samples per collection date resulted in not enough samples to split; thus, the k-fold cross-validation method was employed instead to assess the accuracy of the models.

When using the k-fold cross-validation, the dataset is split into an equal number of sub-datasets at random (Kganyago et al., 2022). Then, k-1 sub-datasets make up a training dataset, and a one k sub-dataset makes up a validation dataset. The iterative validation steps, that is, use one of the k sub-datasets each time, combine to provide the final estimation value (Adeluyi et al., 2021; Kganyago et al., 2022). However, in this study, the k method that was used was the 10-fold cross-validation; thus, the dataset was split into ten equal data subsets at random for 10-fold cross-validation, with nine of these subsets acting as the training set and one as the validation set. Thus, overall, the regression is trained 10 times (Kganyago et al., 2024).

$$R^2 = \frac{\sum(y_i^n - \bar{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|$$

Where n is the number of samples, $\bar{y}_i$ is the estimated mean, and x_i and y_i stand for the estimated and observed sample values, respectively. In terms of mathematics, a model is often seen as having more accuracy if it produces a higher R^2 and a lower RMSE for the same set of sample data (Kganyago et al., 2022).

Furthermore, the Wilcoxon signed-rank test (w), a non-parametric method for comparing paired datasets, was utilised to evaluate whether the predictive accuracies of RF and SVR differed significantly for both LAI and Cab. This test was chosen because it does not assume normality of the error distributions, which is advantageous when the underlying distribution of the sample data is unknown (Couch et al., 2018), making it suitable for remote sensing-derived bio-parameter estimates. The Wilcoxon Signed Rank test (w) can be computed as follows:

$$w = \sum_i s_i \, r_i$$

Where the difference between each pair of values is $d_i = u_i - v_i$, and pairs with $d_i = 0$ are not included in the test. The rank awarded to the absolute value $|d_i|$ is r_i, and tied magnitudes get the average rank. Then, $s_i = \text{sign}(d_i)$ is either ±1 depending on whether d_i is positive or negative. The summation $\sum_i s_i \, r_i$ gives the test statistic w, which reflects the degree and direction of the differences between the paired observations (Couch et al., 2018).

3.7. Temporal Variation in Winter Wheat Crop Bio-Parameters

The LAI and Cab measurements collected in the field on the same collection day were averaged to get one value, representing these crop bio-variables on that specific date. The averaged field measurements were then used to assess the variation in LAI and Cab at different periods and growth stages. To achieve this, the averaged LAI and Cab values were then plotted against the collection dates on the x-axis, creating a time series graph or plot which showed how these variables were changing over time (i.e., showing whether these variables increased or decreased over time), thus showing the overall trend or change of these variables with change in time.

3.8. Feature Importance Analysis

Accurate crop yield estimation requires elucidating the importance of input features or parameters (Xiao et al., 2024). Additionally, feature importance is crucial in separating the most useful and least useful features, and hence lowering the complexity or dimensionality of the data (Srinet et al., 2019). Thus, the RF model's built-in feature importance function was used to assess which features contributed most or least to the prediction model. Meanwhile, caret was used to assess feature importance for SVR since it does not have a built-in function for feature importance. In this study, all the Sentinel-2 surface reflectance bands except bands 1 and 10 were used in the prediction model. Additionally, the normalised difference vegetation index (NDVI) and the Green Chlorophyll Index were some of the features that were in the LAI and Cab prediction models. Moreover, features that did not contribute significantly to the model were not removed since these features varied across time and growth stages. Additionally, assessing the most important features is usually very important in crop monitoring. For example, assessing feature importance gives context to the farmers and future researchers as to which features contribute the most in assessing crop health at different growth stages, and this also gives context as to which features contributed the most and the least in the prediction of LAI and Cab.

3.9. Methodological Flowchart

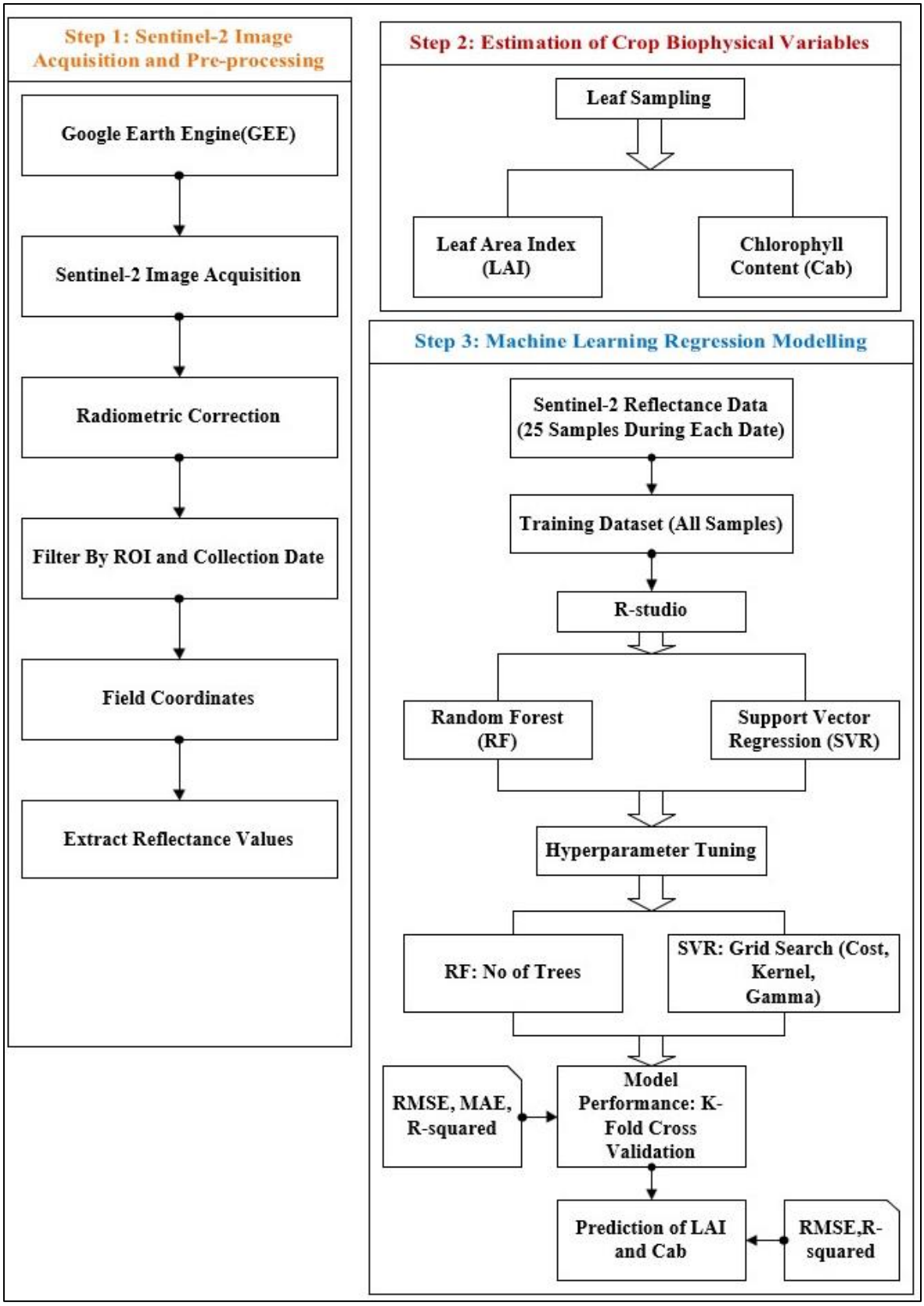


Figure 2: Methodological Flowchart summarising the steps followed to conduct the study

Chapter 4: Results

4.1. Introduction

This section presents the findings through figures and tables, along with brief descriptions of each. The first two sections show the performance results of the RF and SVR LAI and Cab trained models across different growth stages. The third and fourth sections compare the actual and estimated values of LAI and Cab to evaluate how well the RF and SVR models performed. Next, the section displays the results of the RF and SVR LAI and Cab models, which were ultimately selected as the best models. Following this, the section illustrates the temporal variations in LAI and Cab. Subsequently, there is a section highlighting which features contributed most and least to predicting LAI and Cab at various times and growth stages. The final section presents the LAI and Cab maps generated using the best-performing model, the RF.

4.2. Training Accuracy

4.2.1. LAI: SVR vs RF Results

The results below show the RF and SVR, LAI models that were ultimately selected as the best from multiple predictors created for each collection date (Table 1). On 20 August 2022, the SVR LAI model appeared to outperform the RF model in terms of training accuracy, as measured by RMSE and MAE, with RMSE values of 0.164 and 0.184 for SVR and RF, respectively, and MAE values of 0.128 and 0.155 for SVR and RF, respectively (Table 1). However, the SVR LAI model's coefficient of determination (R^2) was slightly lower than that of the RF LAI model, with values of 0.617 and 0.710 for SVR and RF, respectively (Table 1).

Furthermore, the same trend where the SVR LAI model is better than the RF model in terms of RMSE and MAE is observed for almost all the collection dates except for the 25th of September 2022, where RF performed better than SVR in terms of RMSE values, however, the SVR MAE was slightly lower than that of RF (Table 1). Moreover, in terms of R^2, the performance of both the RF and SVR models seems to vary across time; however, when looking at the metrics (RMSE, MAE and R^2), SVR seems to perform better than RF in terms of training accuracy (Table 1). For example, the LAI RF model seems to be less stable and has

the highest variation between the lowest and highest R^2, with a difference of 0.339, while the LAI SVR has the lowest difference between the smallest and highest $R^{2,}$ with a difference of 0.255. Additionally, the SVR LAI model had a slight edge over the RF LAI model for most periods, with the SVR LAI model having an average R^2, RMSE and MAE of 0.792, 0.202, and 0.176, respectively, while the RF LAI model had an average R^2, RMSE and MAE of 0.755, 0.221, and 0.196, respectively. Furthermore, the RF LAI model R^2 ranged between 0.619 and 0.958; meanwhile, the SVR LAI model R^2 values ranged between 0.617 and 0.872. Finally, the Wilcoxon signed-rank test results showed that there was a significant difference at a 95% confidence level between the RF and SVR accuracy metrics during training, with RMSE P-value of 0.009 and MAE P-value of 0.004 (Table 1). However, there was no statistically significant difference in their ability to explain the variation within the data, with an R^2 P-value of 0.278 between the two models' accuracy metrics (Table 1).

Table 1: The comparison of training accuracy of the RF and SVR models in combination with Sentinel-2 imagery in estimating winter wheat crop bio-parameter, namely LAI, in a farm plot at Depaleseng Local Municipality.

Date	LAI RF			LAI SVR		
	RMSE	MAE	R^2	RMSE	MAE	R^2
2022-08-20	0.184	0.155	0.710	0.164	0.128	0.617
2022-08-27	0.092	0.075	0.703	0.085	0.071	0.821
2022-09-03	0.071	0.064	0.809	0.066	0.060	0.736
2022-09-11	0.150	0.132	0.845	0.129	0.110	0.770
2022-09-18	0.143	0.130	0.731	0.123	0.114	0.812
2022-09-25	0.249	0.217	0.889	0.259	0.213	0.872
2022-10-03	0.265	0.226	0.619	0.249	0.218	0.849
2022-10-18	0.283	0.239	0.663	0.244	0.205	0.759
2022-10-25	0.498	0.478	0.958	0.467	0.437	0.797
2022-11-03	0.243	0.205	0.678	0.205	0.169	0.866
2022-11-09	0.257	0.233	0.705	0.229	0.208	0.811
RMSE: p-value = 0.009, MAE: p-value = 0.004, R^2: p-value = 0.278						

4.2.2. Cab: SVR vs RF Results

The results below show the RF and SVR, Cab models that were ultimately selected as the best models from multiple predictors that were created for each collection date. On 20 August 2022, the SVR Cab model performed better than the RF Cab model in terms of RMSE, with RMSE values of 2.059 and 2.295 for SVR and RF, respectively, while the MAE values were 1.876 and 1.868 for SVR and RF, respectively (Table 2). Additionally, the SVR Cab model coefficient of determination (R^2) is also slightly higher than that of the RF Cab model, with values of 0.765 and 0.722 for SVR and RF, respectively (Table 2).

Furthermore, the same trend where the SVR Cab model is better than the RF model in terms of RMSE and MAE is observed for almost all the collection dates except the 03rd of September, and 03rd of November 2022, where RF performed better than SVR in terms of RMSE and MAE values (Table 2). Additionally, in terms of $R^{2,}$ the performance of both the RF and SVR models varies across time; however, the SVR Cab model performed better than the RF overall in terms of training accuracy (Table 2). For example, the SVR Cab model had a smaller difference between the smallest and highest R^2, with a difference of 0.201, while the Cab RF model was a bit unstable, as it had a larger variation between the lowest and highest R^2 with a difference of 0.243 (Table 2). Furthermore, in terms of R^2, RMSE and MAE, the SVR Cab model further outperformed the RF model for most periods, with the SVR Cab model having an average R^2, RMSE and MAE of 0.765, 2.271, and 2.019, respectively, while the RF Cab model had an average R^2, RMSE and MAE of 0.744, 2.242, and 2.145 (Table 2). Finally, the RF Cab model R^2 ranged between 0.620 and 0.863, while the SVR Cab model R^2 values ranged between 0.664 and 0.8765.

Furthermore, the Wilcoxon signed-rank test results showed a significant difference at a 95% confidence level between the RF and SVR accuracy metrics during training, namely RMSE (i.e., p-value = 0.014) and MAE (i.e., p-value = 0.024) (Table 1). However, there was no statistically significant difference in their ability to explain the variation within the data, with an R^2 P-value of 0.638 (Table 1).

Table 2: The comparison of training accuracy of the RF and SVR models in conjunction with Sentinel-2 imagery in estimating winter wheat crop bio-parameter, namely Cab, in a farm plot at Depaleseng Local Municipality.

Date	Cab RF			Cab SVR		
	RMSE	MAE	R^2	RMSE	MAE	R^2
2022-08-20	2.295	1.868	0.722	2.059	1.876	0.765
2022-08-27	2.664	2.432	0.800	2.594	2.376	0.744
2022-09-03	2.223	2.018	0.741	2.249	2.108	0.707
2022-09-11	1.901	1.781	0.790	1.873	1.669	0.839
2022-09-18	1.954	1.828	0.658	1.708	1.579	0.664
2022-09-25	1.713	1.627	0.647	1.428	1.327	0.843
2022-10-03	2.443	2.196	0.863	2.173	1.934	0.789
2022-10-18	2.095	1.780	0.770	1.947	1.639	0.865
2022-10-25	2.301	2.027	0.770	2.026	1.716	0.791
2022-11-03	3.852	3.490	0.620	4.038	3.598	0.673
2022-11-09	3.086	2.549	0.807	2.881	2.384	0.734
RMSE: p-value = 0.014, MAE: p-value = 0.024, R^2: p-value = 0.638						

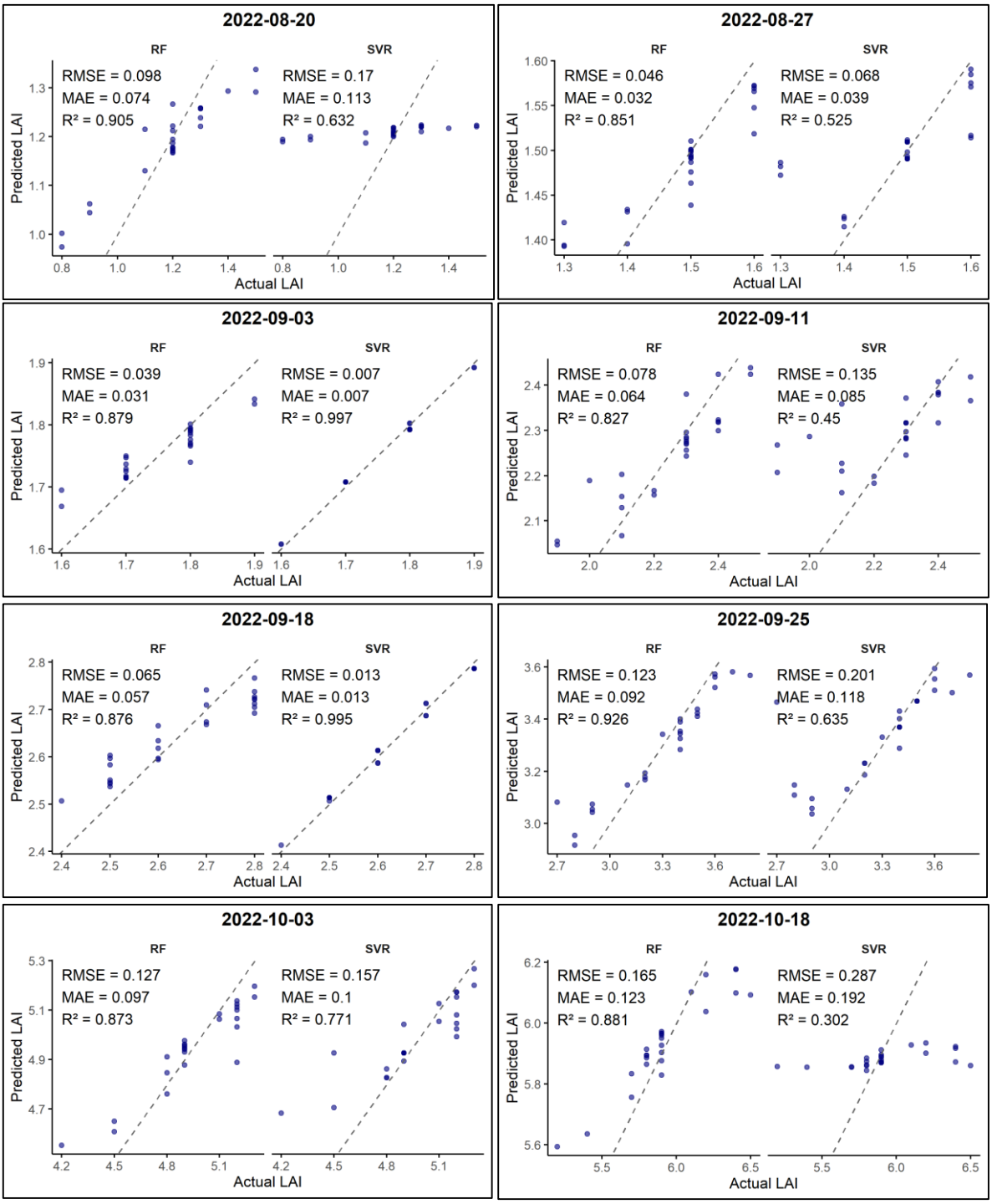
4.3. Predictive Accuracy (Validation)

4.3.1. LAI: RF and SVR (Actual vs Predicted) and Wilcoxon Test

The figure below shows the relationship between actual and predicted LAI values using the RF and SVR models across various growth stages of winter wheat. On 20 and 27 August 2022, the RF model produced R^2 values of 0.905 and 0.851, with RMSE values of 0.098 and 0.046, and MAE values of 0.075 and 0.032, respectively (Figure 3). Meanwhile, the SVR model exhibited R^2 values of 0.632 and 0.525, with RMSE values of 0.170 and 0.068, and MAE values of 0.113 and 0.039, respectively (Figure 3). On the 3rd, 11th, 18th, and 25th of September 2022, RF R^2 values were 0.879, 0.827, 0.876, and 0.926, with corresponding RMSE values of 0.039, 0.08, 0.065, and 0.123, and MAE values of 0.031, 0.064, 0.057, and 0.092, respectively (Figure 3). Meanwhile, the SVR model's R^2 values were 0.997, 0.450, 0.995, and 0.635, with RMSE values of 0.007, 0.135, 0.013, and 0.201, and MAE values of 0.007, 0.085, 0.013, and 0.118 (Figure 3). In October 2022, the RF model maintained strong performance, with R^2 values of 0.873, 0.881, and 0.883 on the 3rd, 18th, and 25th, respectively, accompanied by RMSE values of

0.127, 0.165, and 0.201, and MAE values of 0.097, 0.123, and 0.179, respectively (Figure 3). The SVR model, however, yielded R^2 values of 0.771, 0.302, and 0.297, with RMSE values of 0.157, 0.287, and 0.444, and MAE values of 0.10, 0.192, and 0.394, respectively (Figure 3). On the 3rd and 9th of November 2022, RF exhibited R^2 values of 0.768 and 0.891, with RMSE values of 0.132 and 0.132, and MAE values of 0.101 and 0.112 (Figure 3). Conversely, the SVR model had R^2 values of 0.587 and 0.718, with RMSE values of 0.211 and 0.239, and MAE values of 0.150 and 0.189 (Figure 3).

Overall, both models show a positive relationship between actual and predicted LAI values; however, the RF model demonstrates much stronger predictive power and consistency across different growth stages (Figures 3 and 4). Visually, RF samples mostly align closely with the line of best fit, indicating minimal error and robust prediction, while SVR samples tend to deviate more and respond less across the growing season (Figure 3). The difference between the lowest and highest R^2 for RF is 0.158, whereas for SVR it is 0.700, indicating higher variability and inconsistency in SVR predictions. On average, the RF and SVR models achieved R^2 values of 0.869 and 0.628, respectively, with RMSE values of 0.110 and 0.176, and MAE values of 0.087 and 0.127, respectively, highlighting the superior reliability and accuracy of the RF model in estimating LAI values.


2022-08-20
RF
RMSE = 0.098
MAE = 0.074
R² = 0.905
SVR
RMSE = 0.17
MAE = 0.113
R² = 0.632
Predicted LAI
Actual LAI
2022-08-27
RF
RMSE = 0.046
MAE = 0.032
R² = 0.851
SVR
RMSE = 0.068
MAE = 0.039
R² = 0.525
Predicted LAI
Actual LAI
2022-09-03
RF
RMSE = 0.039
MAE = 0.031
R² = 0.879
SVR
RMSE = 0.007
MAE = 0.007
R² = 0.997
Predicted LAI
Actual LAI
2022-09-11
RF
RMSE = 0.078
MAE = 0.064
R² = 0.827
SVR
RMSE = 0.135
MAE = 0.085
R² = 0.45
Predicted LAI
Actual LAI
2022-09-18
RF
RMSE = 0.065
MAE = 0.057
R² = 0.876
SVR
RMSE = 0.013
MAE = 0.013
R² = 0.995
Predicted LAI
Actual LAI
2022-09-25
RF
RMSE = 0.123
MAE = 0.092
R² = 0.926
SVR
RMSE = 0.201
MAE = 0.118
R² = 0.635
Predicted LAI
Actual LAI
2022-10-03
RF
RMSE = 0.127
MAE = 0.097
R² = 0.873
SVR
RMSE = 0.157
MAE = 0.1
R² = 0.771
Predicted LAI
Actual LAI
2022-10-18
RF
RMSE = 0.165
MAE = 0.123
R² = 0.881
SVR
RMSE = 0.287
MAE = 0.192
R² = 0.302
Predicted LAI
Actual LAI

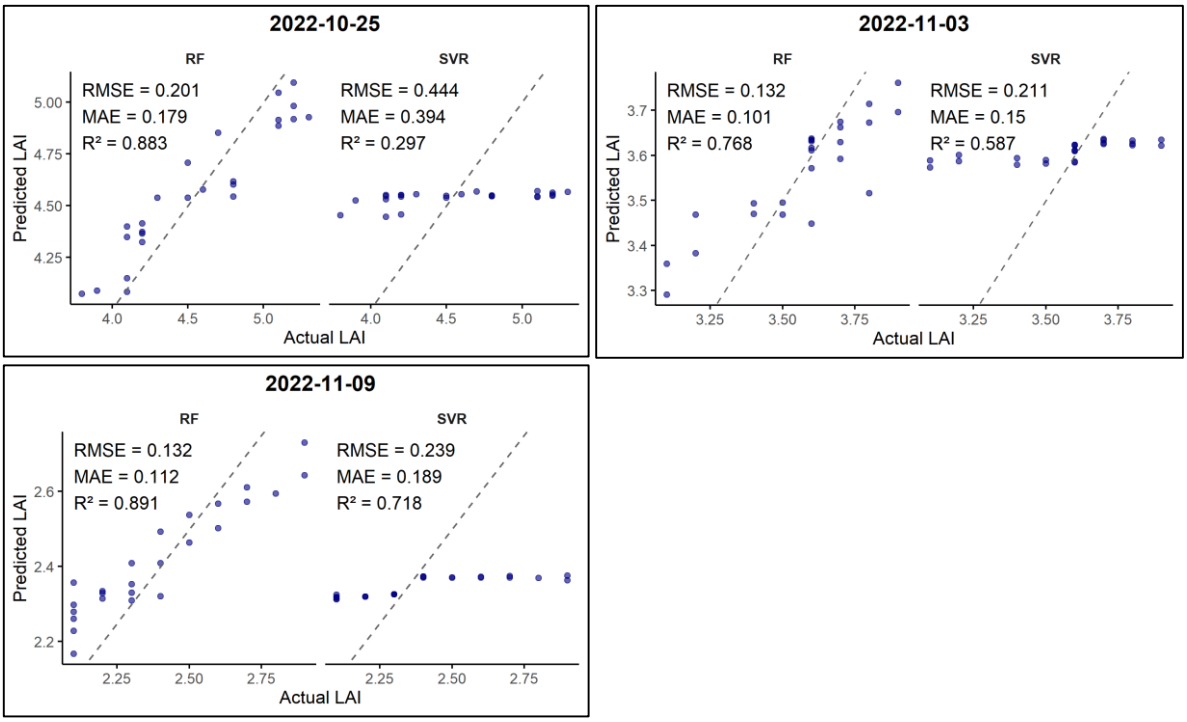


Figure 3: The comparison of actual and predicted bio-variable or parameter, namely LAI of winter wheat, produced by the Random Forest (RF) and Support Vector Regression (SVR) models in conjunction with Sentinel-2 imagery in the farm plot at Depaleseng Local Municipality.

To further assess whether the difference in prediction ability between RF LAI and SVR LAI models was significant, the Wilcoxon signed-rank test was performed. The test revealed statistically significant differences at a 95% confidence level ($\alpha = 0.05$) between RF and SVR for predicting LAI across two metrics (RMSE: $p = 0.019$, R^2: $p = 0.010$) (Table 3). However, the MAE p-value was 0.054, indicating that there was no significant difference in the MAE. Overall, these findings suggest that the model performance varies considerably, with Random Forest offering improved prediction accuracy and explanatory power, as indicated by its lower RMSE and MAE values and higher R^2 values (Figure 3). Additionally, the RF LAI and SVR LAI models' RMSE, MAE and R^2 values were plotted across the growth stages, and they indicate that the RF LAI was more stable across the different growth stages (Figure 4). While SVR, LAI models showed fluctuations across the growth stages, indicating instability or inconsistency (Figure 4).

Table 3: The Wilcoxon signed-rank test statistics showing whether the difference in the prediction accuracy of the RF LAI and SVR LAI models was statistically significant across the different winter wheat growth stages in the farm plot at Depaleseng Local Municipality.

RF vs SVR	Metric	p-value
LAI	RMSE	0.019
LAI	MAE	0.054
LAI	R^2	0.010

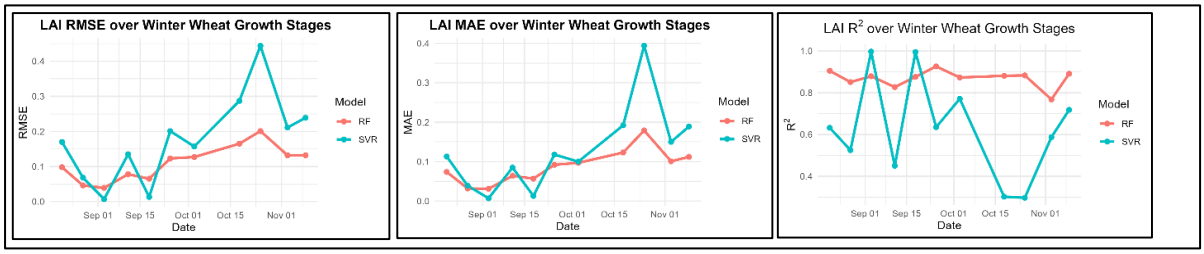


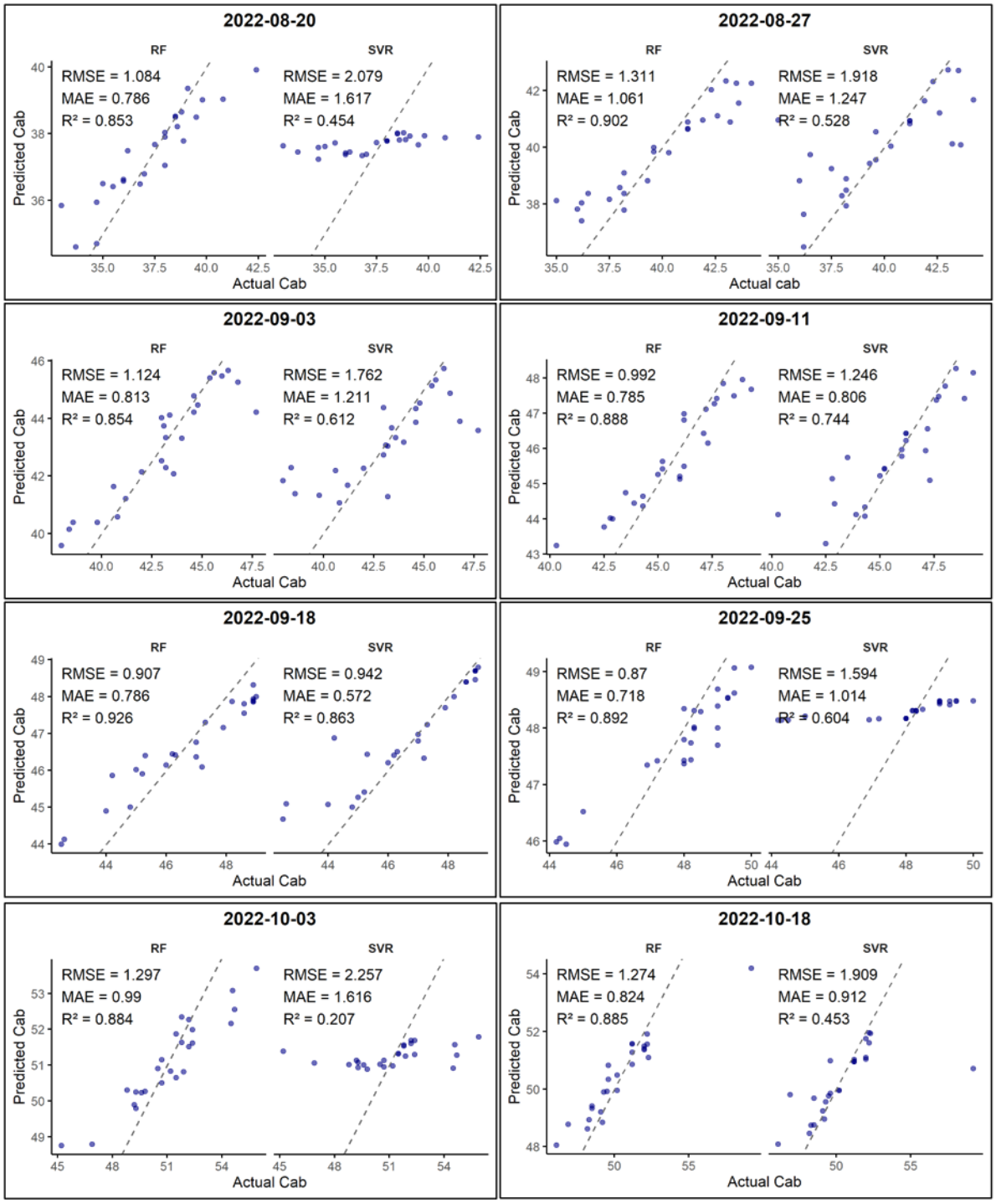
Figure 4: The temporal variation in the prediction accuracy metrics (i.e., RMSE, MAE and R^2) of the RF LAI and SVR LAI models, in the farm plot at Depaleseng Local Municipality.

4.3.2. Cab: RF and SVR (Actual vs Predicted) and Wilcoxon Test

The figure below shows the relationship between actual and predicted Cab values using the RF and SVR models across various growth stages of winter wheat. On 20 and 27 August 2022, the RF model yielded R^2 values of 0.853 and 0.902, with RMSE values of 1.084 and 1.311, and MAE values of 0.786 and 1.061, respectively (Figure 5). Meanwhile, the SVR model exhibited R^2 values of 0.454 and 0.528, with corresponding RMSE values of 2.079 and 1.918, and MAE values of 1.617 and 1.247, respectively (Figure 5). On the 3rd, 11th, 18th, and 25th of September 2022, RF R^2 values were 0.854, 0.888, 0.926, and 0.892, with corresponding RMSE values of 1.124, 0.992, 0.907, and 0.87, and MAE values of 0.813, 0.785, 0.786, and 0.718 (Figure 5). Meanwhile, the SVR model's R^2 values were 0.612, 0.744, 0.863, and 0.604, with RMSE values of 1.762, 1.246, 0.942, and 0.594, and MAE values of 1.211, 0.806, 0.572, and 1.014, respectively (Figure 5). In October 2022, the RF model maintained strong performance with R^2 values of 0.884, 0.885, and 0.904 on the 3rd, 18th, and 25th, accompanied by RMSE values of 1.297, 1.274, and 1.136, and MAE values of 0.99, 0.824, and 0.88, respectively (Figure 5). The SVR model, however, had R^2 values of 0.207, 0.453, and 0.546, with RMSE values of 2.257, 1.909 and 1.992, and MAE values of 1.616, 0.912 and 1.452. On the 3 and 9 November 2022, RF exhibited R^2 values of 0.882 and 0.937, with RMSE values of 2.159 and 1.455, and MAE values of 1.751 and 1.042. Conversely, the SVR model had R^2 values of 0.568

and 0.667, with RMSE values of 2.75 and 3.078, and MAE values of 1.626 and 2.072, respectively (Figure 5).

Overall, both models show a positive relationship between actual and predicted Cab values; however, the RF model demonstrates much stronger predictive power and consistency across different growth stages (Figure 5). Visually, RF samples mostly align closely with the line of best fit, indicating minimal error and robust prediction, while SVR samples tend to deviate more and respond less across the growing season (Figure 5). The difference between the lowest and highest R^2 for the RF Cab model is 0.084, whereas for SVR it is 0.656, indicating higher variability and inconsistency in SVR predictions. On average, the RF and SVR models achieved R^2 values of 0.892 and 0.571, with RMSE values of 1.237 and 1.957 and MAE values of 0.949 and 1.286, respectively, highlighting the superior reliability and accuracy of the RF model in estimating Cab values (Figure 5).


2022-08-20
RF
RMSE = 1.084
MAE = 0.786
R² = 0.853
SVR
RMSE = 2.079
MAE = 1.617
R² = 0.454
Predicted Cab
Actual Cab
2022-08-27
RF
RMSE = 1.311
MAE = 1.061
R² = 0.902
SVR
RMSE = 1.918
MAE = 1.247
R² = 0.528
Predicted Cab
Actual cab
2022-09-03
RF
RMSE = 1.124
MAE = 0.813
R² = 0.854
SVR
RMSE = 1.762
MAE = 1.211
R² = 0.612
Predicted Cab
Actual Cab
2022-09-11
RF
RMSE = 0.992
MAE = 0.785
R² = 0.888
SVR
RMSE = 1.246
MAE = 0.806
R² = 0.744
Predicted Cab
Actual Cab
2022-09-18
RF
RMSE = 0.907
MAE = 0.786
R² = 0.926
SVR
RMSE = 0.942
MAE = 0.572
R² = 0.863
Predicted Cab
Actual Cab
2022-09-25
RF
RMSE = 0.87
MAE = 0.718
R² = 0.892
SVR
RMSE = 1.594
MAE = 1.014
R² = 0.604
Predicted Cab
Actual Cab
2022-10-03
RF
RMSE = 1.297
MAE = 0.99
R² = 0.884
SVR
RMSE = 2.257
MAE = 1.616
R² = 0.207
Predicted Cab
Actual Cab
2022-10-18
RF
RMSE = 1.274
MAE = 0.824
R² = 0.885
SVR
RMSE = 1.909
MAE = 0.912
R² = 0.453
Predicted Cab
Actual Cab

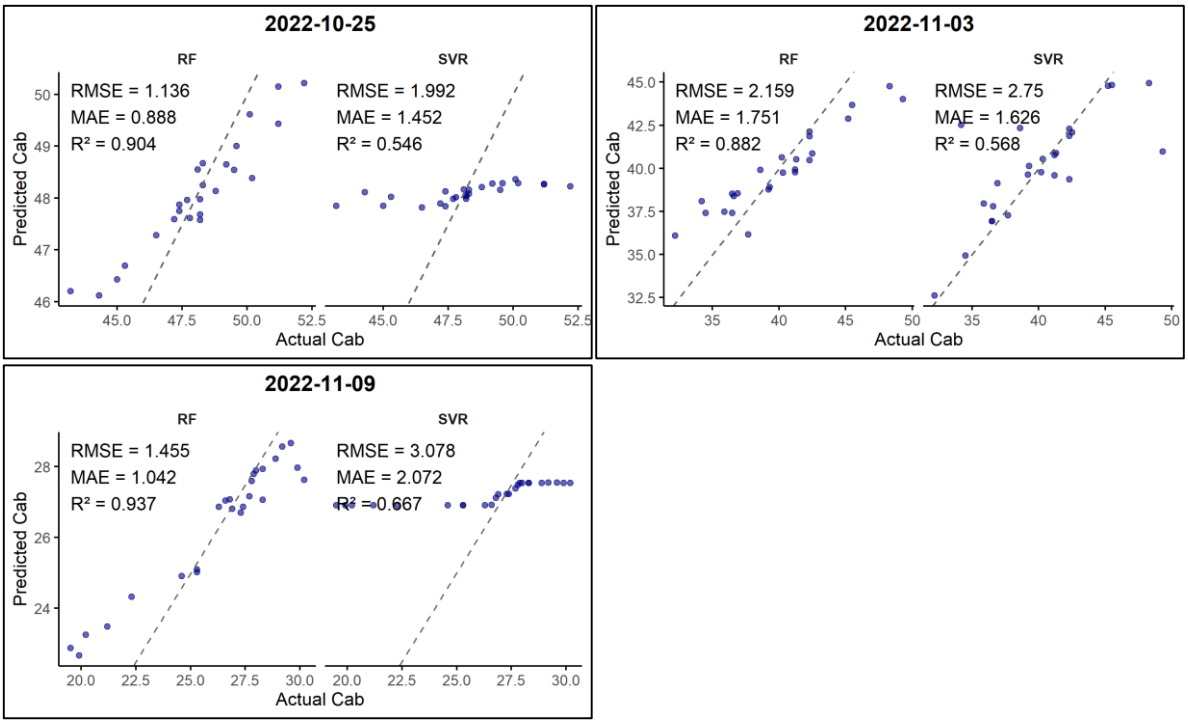


Figure 5: The comparison of actual and predicted bio-variable or parameter, namely Cab of winter wheat, produced by the Random Forest (RF) and Support Vector Regression (SVR) models, in conjunction with Sentinel-2 imagery in the farm plot at Depaleseng Local Municipality.

To further assess whether the difference in prediction ability between RF Cab and SVR Cab models was significant, the Wilcoxon signed-rank test was performed. The test showed a statistically significant difference at a 95% confidence level ($\alpha = 0.05$) between RF and SVR for predicting Cab across all metrics (RMSE: p-value = 0.001, MAE: p-value = 0.019, R^2: p-value = 0.001) (Table 4). These findings suggest that the model performance varies considerably, with RF offering improved prediction accuracy and explanatory power, as indicated by its lower RMSE and MAE values and higher R^2 values (Figure 5). Additionally, the RF Cab and SVR Cab models' RMSE, MAE and R^2 values were plotted across the growth stages, and they indicate that the RF Cab was more stable across the different growth stages (Figure 6). Meanwhile, the SVR Cab model showed a lot of variation across growth stages, indicating instability or inconsistency (Figure 6).

Table 4: The Wilcoxon signed-rank test statistics at 95% confidence level showing whether the difference in the prediction accuracy of the RF LAI and SVR LAI models was statistically significant across the different winter wheat growth stages in the farm plot at Depaleseng Local Municipality.

RF vs SVR	Metric	p-value
Cab	RMSE	0.001
Cab	MAE	0.020
Cab	R^2	0.001

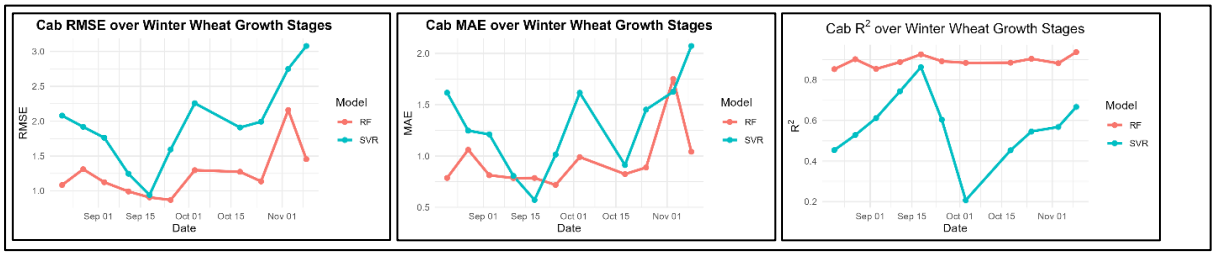


Figure 6: The temporal variation in the prediction accuracy metrics of the RF Cab and SVR Cab models, in the farm plot at Depaleseng Local Municipality.

4.4. Temporal and Feature Dynamics

4.4.1. Temporal Variation across Growth Stages

The LAI and Cab seem to vary over time and at different growth stages. For example, both the LAI and Cab show an increase from the early growth stages to the late middle growth stages and show a decreasing trend again in the late growth stages of the winter wheat crop (Figure 7). However, the Cab changes in values are a bit gradual, and there is no sharp change; meanwhile, the LAI values show a small change in the early stages and then a sharp or steep increase in the late middle stages (Figure 7). Moreover, for LAI, there is another sharp decrease in values from around 2022-10-18 to 2022-11-09 (Figure 7). Additionally, the same trend where values start decreasing can be observed for Cab; however, the decrease started earlier, around the 3rd of October 2022 to the 9th of November 2022, and was gradual when compared to the LAI decrease (Figure 7).

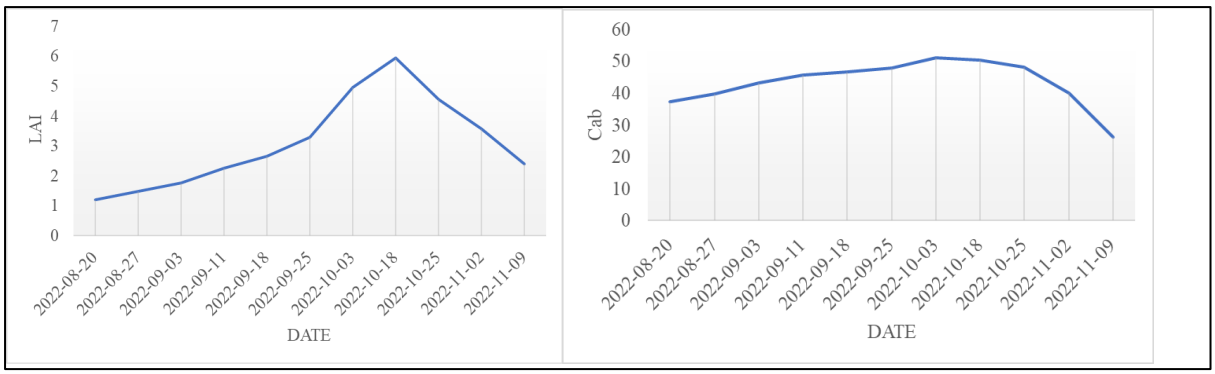


Figure 7: The temporal variation in winter wheat bio-variables or parameters, namely LAI and Cab, in the farm plot at Depaleseng Local Municipality.

4.4.2. Feature Importance at each Stage

4.4.2.1. LAI Feature Importance

The results below show which Sentinel-2 features, including the indices, contributed the most or the least to the prediction of LAI by the RF and SVR models (Figure 8). The features that contributed the most to the prediction model when it comes to predicting LAI seem to vary across time and at different growth stages (Figure 8). For example, on 20 and 27 August 2022, the most important features for the RF LAI model were band 8 and band 2, respectively; while the most important features for the SVR LAI model were band 8 and NDVI. Additionally, the least important features for RF on 20 and 27 August 2022 were band 3 and band 6, respectively, while the least important features for SVR were band 2 and band 5, respectively (Figure 8).

Moreover, on the 3rd, 11th, 18th and 25th of September 2022, the most important features for the RF LAI model were band 3, band 3, band 6 and band 5, respectively, while the most important features for the SVR LAI model were band 12, band 4, band 7 and band 5, respectively (Figure 8). The least important features on the 3rd, 11th, 18th and 25th of September 2022 for the RF LAI model were band 7, band 4, NDVI and band 7, respectively, meanwhile the least important features for the SVR LAI model were NDVI, green chlorophyll index (GCHL), band 11 and band 2, respectively (Figure 8).

On the 3rd, 18th and 25th of October 2022, the most important features for the RF LAI model were band 4, band 4, and band 6, respectively; meanwhile, the most important features for the SVR LAI model were band 4, band 11, and NDVI, respectively (Figure 8). The least important

features for the RF LAI model on the 3rd, 18th and 25th of October 2022 were band 7, band 2 and band 8A, respectively, while the least important features for the SVR LAI model were band 8, band 8A and band 8A, respectively (Figure 8).

Furthermore, on the 3rd and 9th of November 2022, the most important features for the RF LAI model were band 2 and band 8, respectively; meanwhile, the most important features for SVR LAI were band 2 and band 5, respectively (Figure 8). Additionally, on the 3rd and 9th of November 2022, the least important features for the RF LAI model were green chlorophyll index (GCHL) and NDVI, respectively, while the least important features for the SVR LAI model were band 8A and band 4, respectively (Figure 8). However, since the important bands in terms of predicting the LAI were different throughout the growing season the band selection was not performed in this study. If one were focusing on a single growing season, then the most important features were going to ultimately be used for estimation.

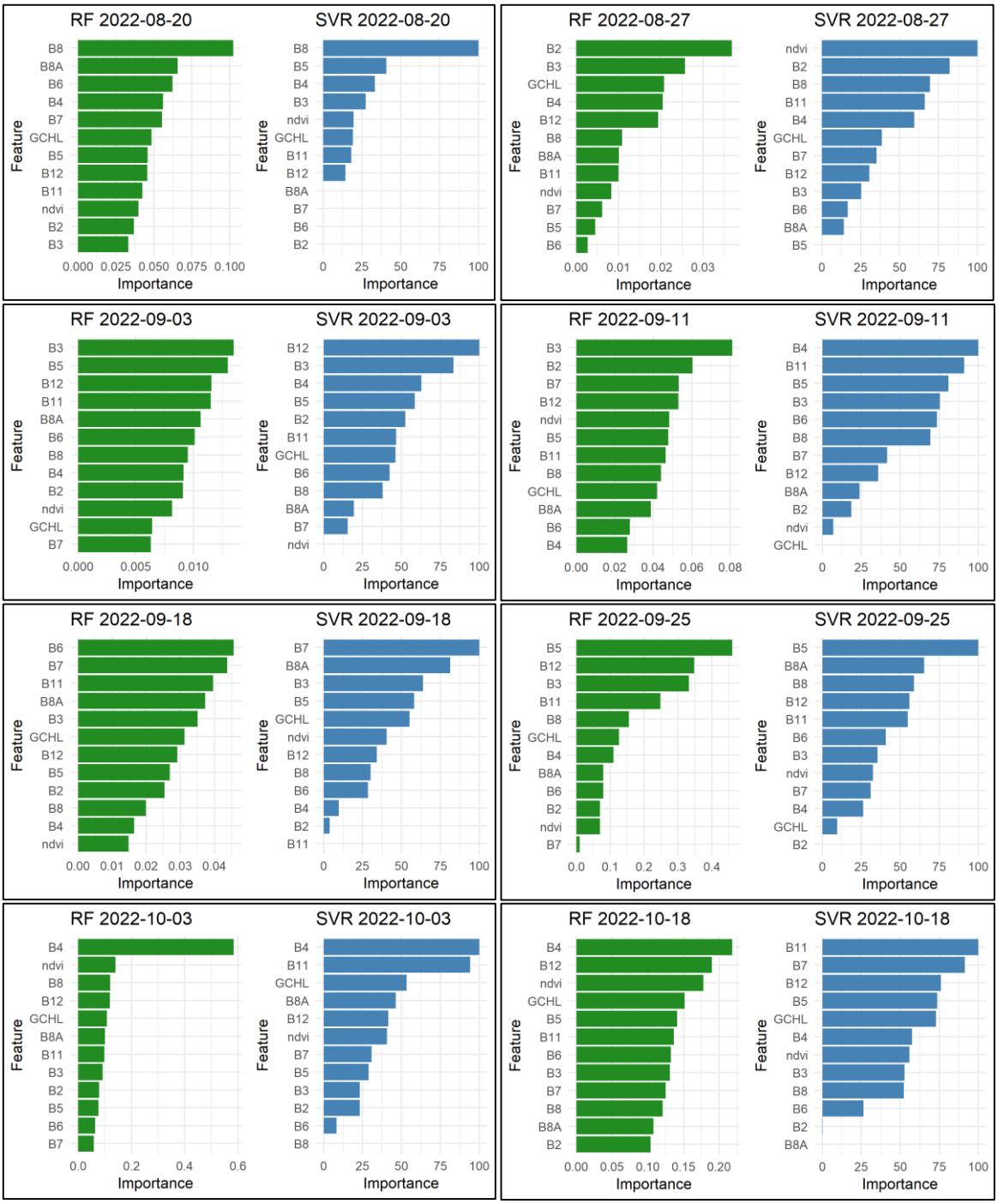

RF 2022-08-20
B8
B8A
B6
B4
B7
GCHL
B5
B12
B11
ndvi
B2
B3
0.000 0.025 0.050 0.075 0.100
Importance
Feature
SVR 2022-08-20
B8
B5
B4
B3
ndvi
GCHL
B11
B12
B8A
B7
B6
B2
0 25 50 75 100
Importance
Feature
RF 2022-08-27
B2
B3
GCHL
B4
B12
B8
B8A
B11
ndvi
B7
B5
B6
0.00 0.01 0.02 0.03
Importance
Feature
SVR 2022-08-27
ndvi
B2
B8
B11
B4
GCHL
B7
B12
B3
B6
B8A
B5
0 25 50 75 100
Importance
Feature
RF 2022-09-03
B3
B5
B12
B11
B8A
B6
B8
B4
B2
ndvi
GCHL
B7
0.000 0.005 0.010
Importance
Feature
SVR 2022-09-03
B12
B3
B4
B5
B2
B11
GCHL
B6
B8
B8A
B7
ndvi
0 25 50 75 100
Importance
Feature
RF 2022-09-11
B3
B2
B7
B12
ndvi
B5
B11
B8
GCHL
B8A
B6
B4
0.00 0.02 0.04 0.06 0.08
Importance
Feature
SVR 2022-09-11
B4
B11
B5
B3
B6
B8
B7
B12
B8A
B2
ndvi
GCHL
0 25 50 75 100
Importance
Feature
RF 2022-09-18
B6
B7
B11
B8A
B3
GCHL
B12
B5
B2
B8
B4
ndvi
0.00 0.01 0.02 0.03 0.04
Importance
Feature
SVR 2022-09-18
B7
B8A
B3
B5
GCHL
ndvi
B12
B8
B6
B4
B2
B11
0 25 50 75 100
Importance
Feature
RF 2022-09-25
B5
B12
B3
B11
B8
GCHL
B4
B8A
B6
B2
ndvi
B7
0.0 0.1 0.2 0.3 0.4
Importance
Feature
SVR 2022-09-25
B5
B8A
B8
B12
B11
B6
B3
ndvi
B7
B4
GCHL
B2
0 25 50 75 100
Importance
Feature
RF 2022-10-03
B4
ndvi
B8
B12
GCHL
B8A
B11
B3
B2
B5
B6
B7
0.0 0.2 0.4 0.6
Importance
Feature
SVR 2022-10-03
B4
B11
GCHL
B8A
B12
ndvi
B7
B5
B3
B2
B6
B8
0 25 50 75 100
Importance
Feature
RF 2022-10-18
B4
B12
ndvi
GCHL
B5
B11
B6
B3
B7
B8
B8A
B2
0.00 0.05 0.10 0.15 0.20
Importance
Feature
SVR 2022-10-18
B11
B7
B12
B5
GCHL
B4
ndvi
B3
B8
B6
B2
B8A
0 25 50 75 100
Importance
Feature



Figure 8: The relative feature importance and contribution of each feature in the Random Forest (RF) and Support Vector Regression (SVR) estimation of the winter wheat crop bio-parameter, namely LAI at Depaleseng Local Municipality.

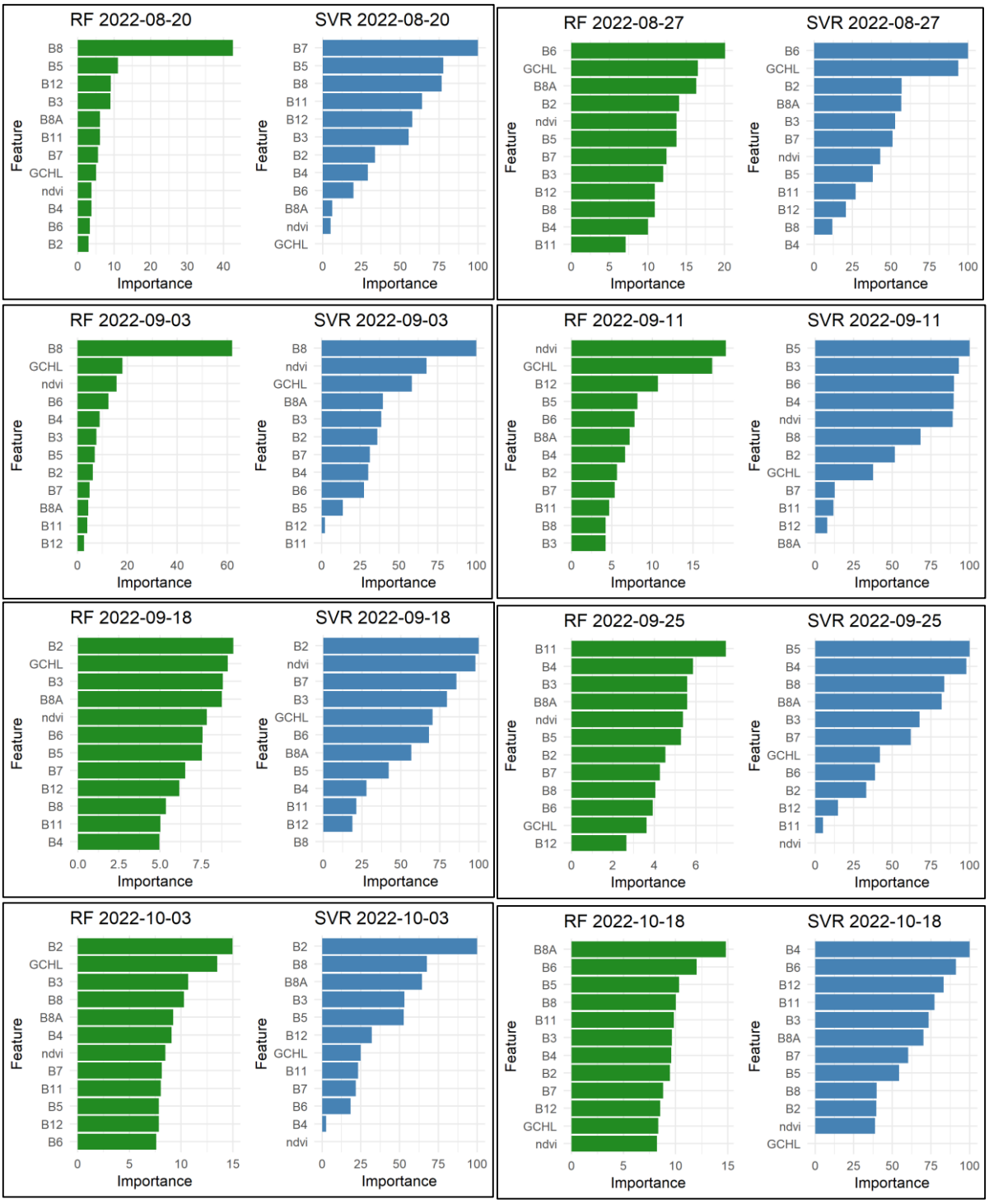
4.4.2.2. Cab Feature Importance

The results below show which Sentinel-2 features, including the indices, contributed the most or the least to the prediction of Cab by the RF and SVR models (Figure 9). The features that contributed the most to the prediction model when it comes to predicting Cab seem to vary across time and at different growth stages (Figure 9). For example, on 20 and 27 August 2022, the most important features for the RF Cab model were band 8 and band 6, respectively; meanwhile, the most important features for the SVR Cab model were band 7 and band 6, respectively. Additionally, the least important features for the RF Cab model on the 20th and 27th of August 2022 were band 2 and band 11, respectively, while the least important features for the SVR Cab model were green chlorophyll index (GCHL) and band 4, respectively (Figure 9).

Moreover, on the 3rd, 11th, 18th and 25th of September 2022, the most important features for the RF Cab model were band 8, NDVI, band 2 and band 11, respectively, while the most important

features for the SVR Cab model were band 8, band 5, band 2 and band 5, respectively (Figure 9). The least important features for the RF Cab model on the 3rd, 11th, 18th and 25th of September 2022 were band 12, band 3, band 4 and band 12, respectively, while the least important features for the SVR Cab model were band 11, band 8A, band 8 and NDVI, respectively (Figure 9). Moreover, on the 3rd, 18th and 25th of October 2022, the most important features for the RF Cab model were band 2, band 8A, and NDVI, respectively; meanwhile, the most important features for the SVR Cab model were band 2, band 4, and green chlorophyll index (GCHL), respectively (Figure 9). On the 3rd, 18th and 25th of October 2022, the least important features for the RF Cab model were band 6, NDVI and band 3, while the least important features for SVR were NDVI, the green chlorophyll index (GCHL) and band 5 (Figure 9).

Furthermore, on the 3rd and 9th of November 2022, the most important features for the RF Cab model were band 4 and NDVI, respectively; while the most important features for the SVR Cab model were band 8 and NDVI, respectively (Figure 9). Additionally, on the 3rd and 9th of November 2022, the least important features for the RF Cab model were band 6 and band 2, respectively, while the least important features for the SVR Cab model were NDVI and band 12, respectively (Figure 9). However, since the important features in terms of predicting the Cab were different throughout the growing season the feature selection was not performed in this study. However, if one was focusing on a single growth stage at a time, then the most important features were going to be ultimately used for the estimation.


RF 2022-08-20
B8 B5 B12 B3 B8A B11 B7 GCHL ndvi B4 B6 B2
0 10 20 30 40
Importance
Feature
SVR 2022-08-20
B7 B5 B8 B11 B12 B3 B2 B4 B6 B8A ndvi GCHL
0 25 50 75 100
Importance
Feature
RF 2022-08-27
B6 GCHL B8A B2 ndvi B5 B7 B3 B12 B8 B4 B11
0 5 10 15 20
Importance
Feature
SVR 2022-08-27
B6 GCHL B2 B8A B3 B7 ndvi B5 B11 B12 B8 B4
0 25 50 75 100
Importance
Feature
RF 2022-09-03
B8 GCHL ndvi B6 B4 B3 B5 B2 B7 B8A B11 B12
0 20 40 60
Importance
Feature
SVR 2022-09-03
B8 ndvi GCHL B8A B3 B2 B7 B4 B6 B5 B12 B11
0 25 50 75 100
Importance
Feature
RF 2022-09-11
ndvi GCHL B12 B5 B6 B8A B4 B2 B7 B11 B8 B3
0 5 10 15
Importance
Feature
SVR 2022-09-11
B5 B3 B6 B4 ndvi B8 B2 GCHL B7 B11 B12 B8A
0 25 50 75 100
Importance
Feature
RF 2022-09-18
B2 GCHL B3 B8A ndvi B6 B5 B7 B12 B8 B11 B4
0.0 2.5 5.0 7.5
Importance
Feature
SVR 2022-09-18
B2 ndvi B7 B3 GCHL B6 B8A B5 B4 B11 B12 B8
0 25 50 75 100
Importance
Feature
RF 2022-09-25
B11 B4 B3 B8A ndvi B5 B2 B7 B8 B6 GCHL B12
0 2 4 6
Importance
Feature
SVR 2022-09-25
B5 B4 B8 B8A B3 B7 GCHL B6 B2 B12 B11 ndvi
0 25 50 75 100
Importance
Feature
RF 2022-10-03
B2 GCHL B3 B8 B8A B4 ndvi B7 B11 B5 B12 B6
0 5 10 15
Importance
Feature
SVR 2022-10-03
B2 B8 B8A B3 B5 B12 GCHL B11 B7 B6 B4 ndvi
0 25 50 75 100
Importance
Feature
RF 2022-10-18
B8A B6 B5 B8 B11 B3 B4 B2 B7 B12 GCHL ndvi
0 5 10 15
Importance
Feature
SVR 2022-10-18
B4 B6 B12 B11 B3 B8A B7 B5 B8 B2 ndvi GCHL
0 25 50 75 100
Importance
Feature

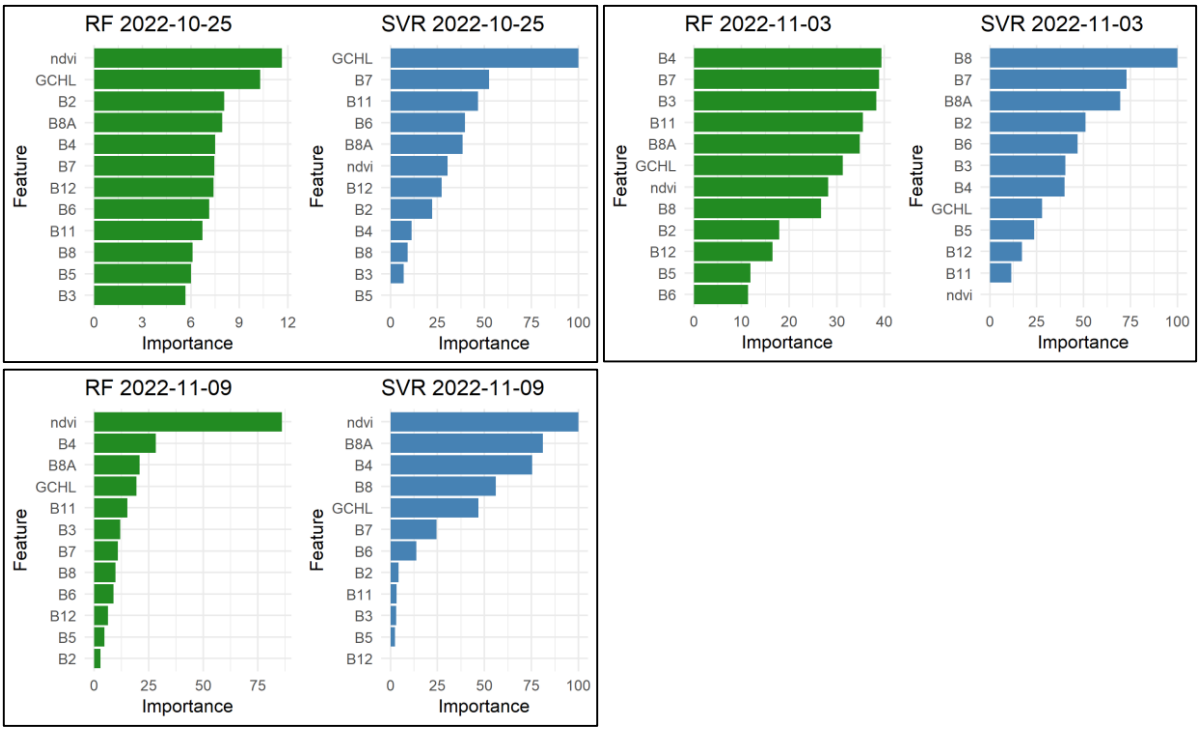


Figure 9: The relative feature importance and contribution of each feature in Random Forest (RF) and Support Vector Regression (SVR) estimation of the winter wheat crop bio-parameter, namely Cab at Depaleseng Local Municipality.

4.5. Prediction Maps

4.5.1. RF LAI Prediction Maps

The RF model was chosen as the best model in terms of the RMSE, MAE, and R^2 of the predicted against the actual values of LAI. Additionally, the Wilcoxon test was performed to assess whether the difference between RF and SVR was statistically significant for RMSE and R^2. Based on this, RF was chosen as the best and was used to predict LAI across the plot at different periods and growth stages. The LAI values varied across the plot, with the majority of the highest highlighted by the green colour being observed in the Southern and Northern regions of the plot on 20 August 2022 (Figure 10). Furthermore, the LAI values seem to vary across time, for example, on the 20th and 27th of August 2022 the predicted LAI values ranged from around 1 to 1.3 and 1.4 to 1.55 respectively, then on the 3rd, 11th, 18th and 25th of September 2022 LAI values were ranging from around 1.7 to 1.8, 2.1 to 2.4, 2.55 to 2.75 and 3.0 to 3.6 respectively (Figure 10).

Moreover, on the 3rd, 18th and 25th of October 2022, the LAI values ranged from around 4.6 to 5.1, 5.6 to 6.1, and 4.2 to 5.0, respectively. Furthermore, on 3 and 9 November 2022, the LAI values ranged from around 3.3 to 3.7 and 2.2 to 2.7, respectively (Figure 10). Finally, the LAI values increase from the early stages to the late middle growth stages of winter wheat; moreover, these values then decrease during the late stages of growth across the plot (Figure 10).

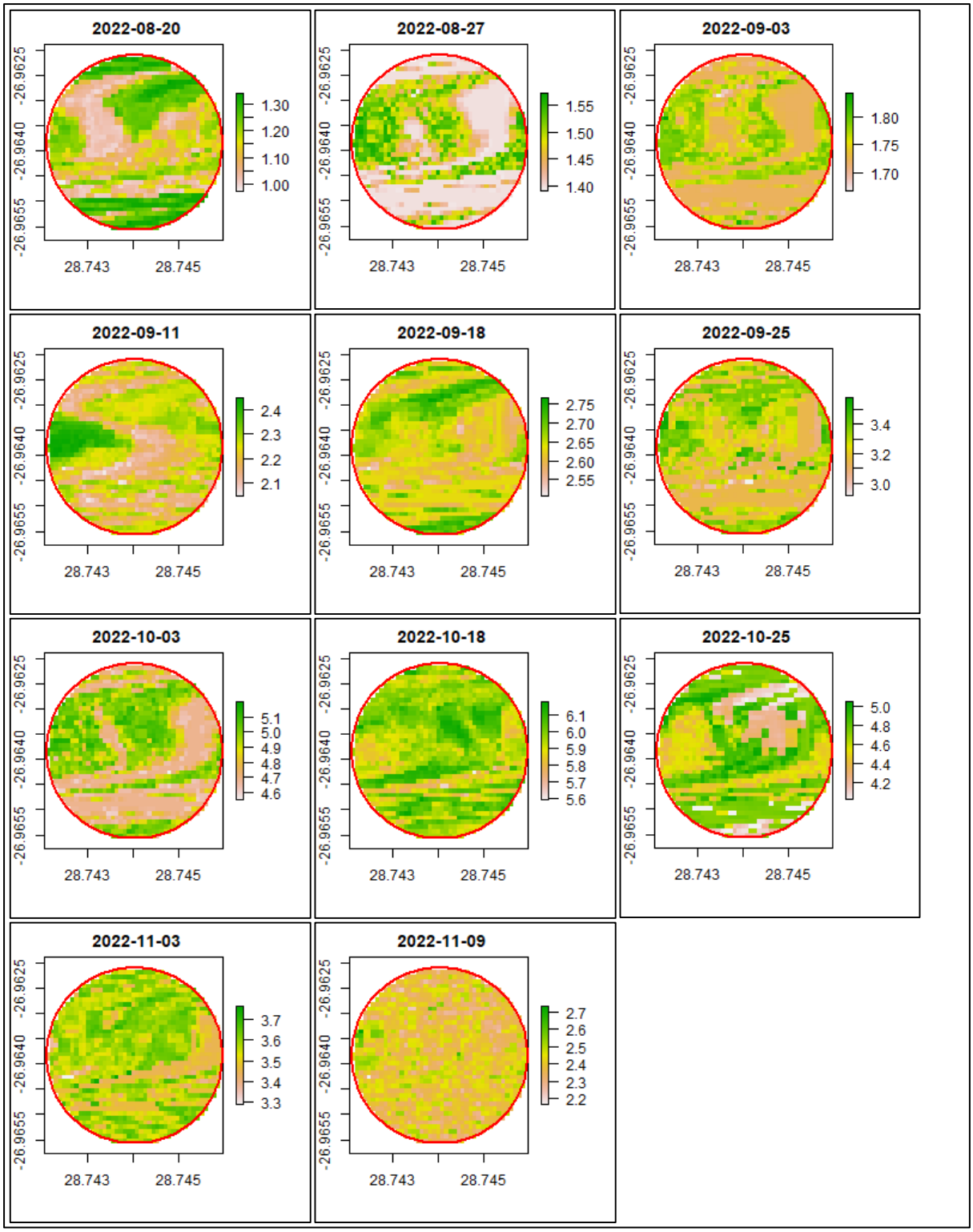


Figure 10: Maps of winter wheat crop bio-parameter, namely LAI, produced by the best model, the Random Forest (RF), in conjunction with Sentinel-2 imagery in a farm plot at Depaleseng Local Municipality.

4.5.2. RF Cab Prediction Maps

The RF model was chosen as the best model in terms of RMSE, MAE, and R^2 of the predicted against the actual values of Cab. Additionally, the Wilcoxon test was performed to assess whether the difference between RF and SVR was statistically significant at a 95% level across all three metrics. Based on this, RF was used to predict Cab across the plot at different periods and growth stages. The Cab values vary across the plot, with the majority of the highest highlighted by a green colour being observed in the southwestern regions of the plot on 20 August 2022 (Figure 11). Furthermore, the RF predicted Cab values seem to vary across time, for example, on the 20 and 27 August 2022 the Cab values ranged from around 35 to 40 and 38 to 42 respectively, then on the 3rd, 11th, 18th and 25th of September 2022, the predicted Cab values were ranging from around 40 to 45, 44 to 47, 44 to 48 and 46.0 to 49.0 respectively (Figure 11).

Moreover, on the 3rd, 18th and 25th of October 2022, the Cab values ranged from around 49 to 53, 48 to 54, and 47 to 50, respectively. Furthermore, on 3 and 9 November 2022, the Cab values ranged from around 38 to 44 and 22 to 28, respectively (Figure 11). Finally, the Cab values increase over time from the early stages to the late middle growth stages of winter wheat; moreover, these values then decrease during the late growth stages across the plot (Figure 11).

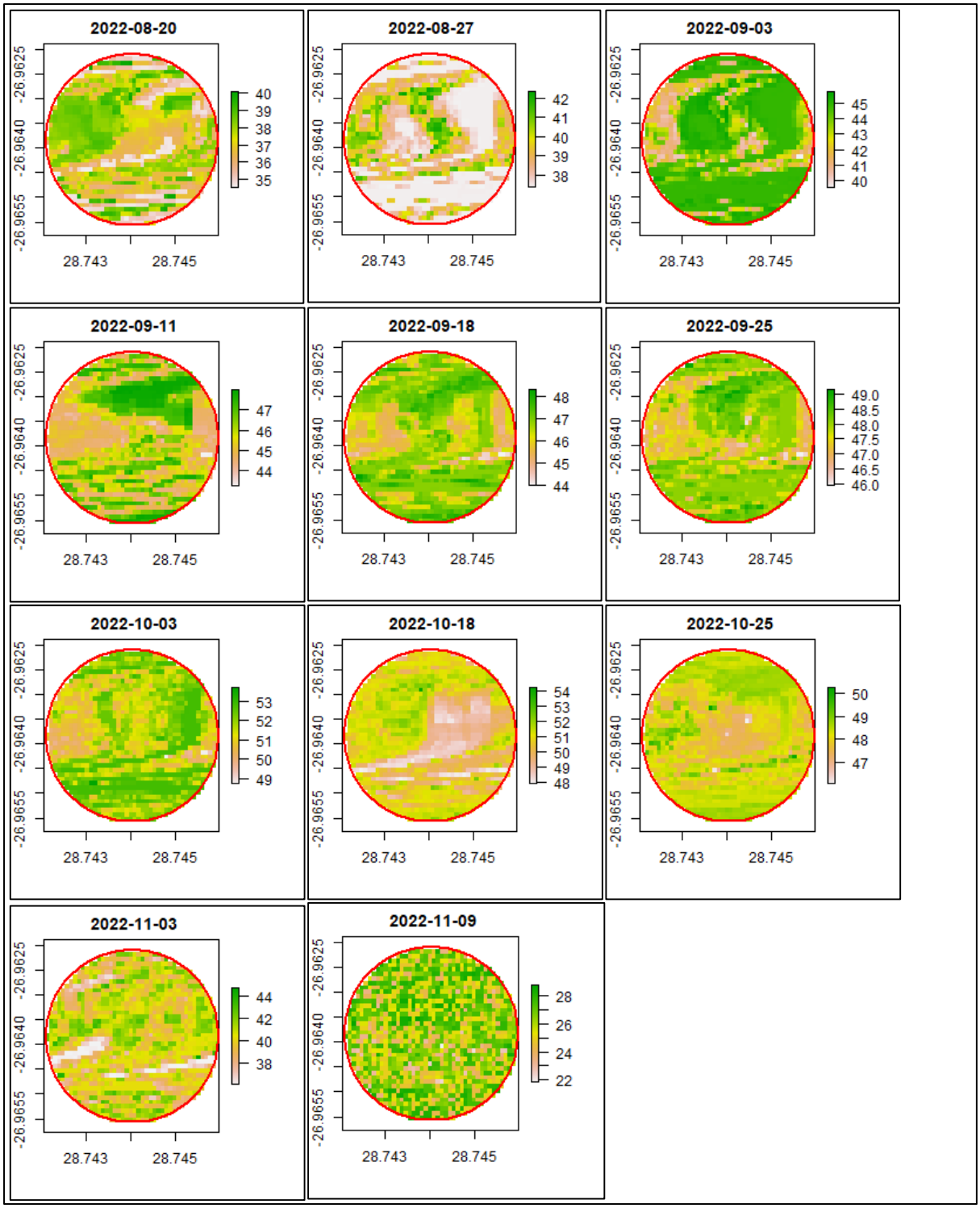


Figure 11: Maps of winter wheat crop bio-parameter, namely Cab, produced by the best model, the Random Forest (RF), in conjunction with Sentinel-2 imagery in a farm plot at Depaleseng Local Municipality.

Chapter 5: Discussions

5.1. Introduction

This segment provides a detailed synthesis of the findings above and puts them into context by comparing them to the findings from previous studies. Additionally, it looks at whether the results above were satisfactory to meet the aims and goals of the study. The first section looks at the RF and SVR, LAI and Cab models' training accuracy and how they relate to the previous research, while the section that follows looks at the predictive accuracies of the RF and SVR models in terms of predicting the LAI and Cab values from Sentinel-2. This is followed by the section looking at the variation of the winter wheat LAI and Cab across time, this is followed by a section looking at the most and least important features in the estimation of LAI and Cab of winter wheat. The last section looks at the LAI and Cab maps that were predicted by the best-performing model in the study, the RF model, and how this relates to the previous studies.

5.2. RF and SVR Model's Prediction Performance

Plant growth and productivity are closely linked to the LAI and chlorophyll concentration; additionally, spatiotemporal estimates of these parameters are crucial for efficient and accurate crop management (Revill et al., 2019). Thus, the first goal of the study was to use SVR and RF machine learning algorithms to predict LAI and Cab at different growth stages from Sentinel-2 multispectral imagery.

5.2.1. Training Accuracy

Both the SVR and RF models performed well during training; however, the SVR model showed slightly better performance than the RF for both LAI and Cab estimations in terms of RMSE, MAE, and R^2. For LAI, SVR achieved the lowest average RMSE and MAE with values of 0.202 and 0.175, respectively, while RF recorded slightly higher averages of 0.221 and 0.196. Similarly, for Cab, SVR maintained its advantage with average RMSE and MAE values of 2.270 and 2.019, respectively, compared to RF's 2.412 and 2.145. The R^2 values for both models varied across different growth stages, with RF ranging from 0.619 to 0.958 for LAI and 0.62 to 0.863 for Cab, while SVR ranged from 0.617 to 0.872 for LAI and 0.664 to 0.865 for

Cab (Tables 1 and 2). These results align with findings from C. Zhang et al. (2024), who reported that SVR combined with sequential forward selection (SFS) performed better than RF and other regression models in terms of RMSE across the four growth stages of winter wheat. Similarly, L. Zhang et al. (2024) found that support vector machines (SVM) outperformed RF and multiple linear regression (MLR) in training accuracy, which matches the current study's results, even though SVR was used instead of SVM. The variation in R^2 values across growth stages also corresponds with studies such as Madonsela et al. (2023) and H. Zhang et al. (2023), which reported different accuracies across growth stages. However, the lower range of R^2 values in the current study, for both the SVR and RF models, is slightly below the range reported by L. Zhang et al. (2024), who obtained R^2 values between 0.932 and 0.944 for the three models used in their study, namely RF, SVM, and multiple linear regression (MLR).

Additionally, the findings in Tables 1 and 2 above demonstrate that the training accuracy of the models can be affected by the crop's growth stages, which might be caused by the changes in the spectral signatures or response of the crop across different growth stages. Moreover, the LAI and Cab levels vary across growth stages, which may contribute to variations in model accuracy at these stages (Figure 7). This corresponds with the results from L. Zhang et al. (2024), which found that the growth stages of winter wheat tend to have more influence than nitrogen application and wheat species on the generalisation ability of leaf chlorophyll content (LCC) prediction models. Finally, the findings from previous studies exhibit the same trend as the results in the current study, as the accuracy and robustness of the models seem to vary across different growth stages, which is the same trend observed in the current study (Tables 1 and 2). Thus, the variation in model accuracy across the different crop stages suggests that the hyperparameters must be tuned specifically for each growth stage. Moreover, approaches such as grid search might assist in finding the best model combination for each growth stage, improving prediction accuracy and model robustness.

5.2.2. Predictive Accuracy

However, when these two models were used to estimate the LAI and Cab values on new data, the RF model outperformed the SVR model for almost all the collection periods (Figures 3 and 5). These results show a contrasting trend from the training accuracy, as the SVR model seemed to have a slight edge over RF during training. Moreover, the RF LAI and Cab models were

more stable and consistent in terms of predictive accuracy across different crop stages, while the SVR LAI and Cab models seemed to be unstable with fluctuating R^2, RMSE, and MAE values at different growth stages (Figures 4 and 6).

Furthermore, results from the Wilcoxon signed-rank test showed that there was a significant difference between RF and SVR across all accuracy metrics, namely RMSE, MAE, and R^2; however, there was no significant difference in MAE between RF LAI and SVR LAI (Tables 3 and 4). Thus, it seems like when one is assessing crop health across the growing season, the RF model is the best to use as it seems to perform consistently through the winter wheat growing season, since it had the lowest RMSE and MAE values and the highest R^2 values when compared to SVR (Figures 4 and 6). This corresponds with the results from Yuan et al. (2017) even though they were using UAV data, their results showed that RF can be more effective for predicting LAI when the sample plots are big and vary greatly, which usually occurs when one is concentrating on the entire growth cycle of the crop from emergence to the harvesting stage, which is the case in the current study. Additionally, the RF LAI models had an average predictive R^2 of 0.869, with RMSE of 0.110 and MAE of 0.087, while RF Cab models had an average predictive R^2 of 0.892, with RMSE of 1.237, and MAE values of 0.949 (Tables 3 and 4) and (Figures 3 and 4). These results also align with findings from Kganyago et al. (2021), which also found that RF with an RMSE of 0.5 and R^2 of 0.76 usually has better predictive accuracy when compared to other algorithms such as gradient boosting machines (GBM) and Partial Least Squares (PLS).

Furthermore, the findings in (Tables 1 and 2) and (Figures 3 and 5) indicate that the SVR model might have overfitted, resulting in the model doing well with training data; however, it has poor accuracy when it comes to predicting signalling weak generalisation on unseen data or observations and limited transferability across conditions or growth stages of winter wheat. As a result, it has a high RMSE and MAE with a lower predictive R^2 for the predicted versus actual LAI and Cab values when compared to RF for almost all collection periods (Tables 3 and 4). For example, when nonparametric approaches are employed, flexible model definitions may lead to the issue of overfitting training data obtained from field experiments (Zhou et al., 2020).

Additionally, this might have been caused by the limited number of training samples used in this study and by using almost all the Sentinel-2 features, including the NDVI and green

chlorophyll indices, as this might have affected the ability of the model to generalise. For example, reducing model complexity and the likelihood of overfitting from redundant features depends heavily on the comparatively tiny feature subsets (Kganyago et al., 2024). However, in the case of RF, it is usually less sensitive to the number of inputs; additionally, it is generally robust to noise and overtraining (Lambert et al., 2018), which might explain why it had a good balance between training accuracy and prediction accuracy when new data was used.

Moreover, these findings further indicate that one must pay special attention to ensure that there is no overfitting in their models, as this might lead to misleading results. Additionally, various studies, including the study by Naidoo et al. (2022) and Kganyago et al. (2024) have shown that approaches such as Recursive Feature Elimination (RFE), which start by using all predictor variables or features, and then eliminate or remove features until the desired or required number of predictors remains, can be used to limit the issues of collinearity and thus avoid overfitting. Furthermore, the effectiveness of the learning algorithm and the stability of the model may be enhanced by appropriately selecting features, which can also efficiently eliminate redundant and unnecessary features (Liu et al., 2024). Thus, future studies can utilise this approach to ensure that only features that contribute meaningfully to the model remain, resulting in models that are not overfit, as this reduces their usability and transferability. Finally, the first goal of the study was successfully achieved, as the RF and SVR winter wheat crop bio-parameters were predicted, and the prediction models were created and contrasted.

5.3. Temporal Variation and Feature Importance of LAI and Cab

Understanding the behaviour of crop bio-parameters at different growth stages and determining which features contribute the most in predicting these variables is crucial, and it is a requirement or prerequisite for accurate crop yield estimation (Xiao et al., 2024). To put it simply, an explanatory variable with a high significance score indicates a strong correlation with the response variable (Kganyago et al., 2024). Thus, the second goal of the study was to evaluate the variations in LAI and Cab across time at various growth stages and to find the features that contributed most or least to the LAI and Cab prediction models.

5.3.1. Temporal variation in LAI and Cab

The LAI and Cab values fluctuated over time; however, there was only a small difference between the lowest and highest Cab values, so the change was not very significant (Figure 7). In contrast, the difference between the lowest and highest values appeared to be more noticeable for LAI values (Figure 7). These actual LAI and Cab values change over time and show an increase in both from the early growth stages to the mid-growth stages, followed by a decrease as the crop approaches harvest and senescence in winter wheat around early to mid-November (Figure 7). These findings align with those from (Verma et al., 2022), which indicated that crop bio-parameters, such as LAI and LCC, vary at different growth stages, with the lowest values occurring in the early and late stages, and the highest during the middle stages. This pattern is consistent for both LAI and Cab results in this study. The variation may be linked to changes in leaf development and photosynthetic activity. Similarly, (Han et al., 2021) observed that LAI increases during the vegetative growth stage and declines during senescence. This is also true for the current study; for instance, the lowest LAI and Cab values, recorded in August, were approximately 1.188 and 37.400, respectively, while the highest, recorded in October, were around 5.932 and 51.144. The final LAI and Cab values, obtained in November, were 2.392 and 26.216, respectively (Figure 7).

5.3.2. Feature importance

Furthermore, clarifying the significance of input traits or factors is necessary for accurate crop yield estimation (Xiao et al., 2024). Thus, the least and most important features in predicting the LAI and Cab of winter wheat were identified and, these features seemed to change over time (Figures 8 and 9), this might be due to the changes in levels of Cab and LAI at various growth stages, for example, the LAI and Cab levels show an increasing then a decreasing trend (Figure 7). Furthermore, the variation in the most and least important features is more prevalent when predicting LAI, as there seems to be no dominant feature which played the most crucial role at different periods (Figure 8). Additionally, even though the variations on the most and least important features are there when predicting Cab, green chlorophyl index, NDVI, band 2 and band 4 were shown to be important features in predicting Cab more than once; however, these were not the most dominant feature for all periods (Figure 9).

Moreover, Xiao et al. (2024) found that the B4 (Red) and B5 (Vegetation Red Edge) were the most significant or relevant bands or features in crop monitoring (i.e., estimating winter wheat yield). Furthermore, Zhou et al. (2020) found that the red band and NIR bands are very crucial when retrieving crop biophysical variables such as LCC. Additionally, Xu et al. (2022) and Madonsela et al. (2023), also found that the red edge (RE) bands play a significant role in evaluating willow health status; thus, these studies demonstrate that the red edge bands usually contribute a lot to crop monitoring.

Nonetheless, the feature importance results in the current study showed that the most significant features varied across time and growth stages for both LAI and Cab predictions. However, even though there is no constant fixed most important feature, bands 4 and 5 seemed to be amongst the most important features for most periods in LAI; meanwhile, in terms of Cab predictions, the green chlorophyll index can be singled out as one of the most significant features across different growth stages. Thus, in terms of Cab prediction green chlorophyll index seemed to be one of the most significant features at various growth stages of winter wheat, however, in general, the NIR and Red Edge (RE) features seem to play an important role in retrieving the bio-variables for LAI and Cab (Figures 8 and 9), this is also in line with the results from the studies, such the one by Zhou et al. (2020) and Madonsela et al. (2023) above, which found that NIR and RE bands play a significant role in crop yield and health prediction. This might be because the NIR bands are extremely sensitive to leaf structure and biomass; meanwhile, the Red Edge bands are highly sensitive to chlorophyll concentration, making them a powerful indicator of Cab (Imran et al., 2020).

These findings emphasise the importance of evaluating which features most significantly contribute to predicting crop bio-variables, such as winter wheat, at various growth stages, as these features differ over time and across stages. This indicates that one cannot assume that features important at one stage will necessarily be important at another. Moreover, this variation in key features highlights that one should not select features in advance for estimating bio-variables; instead, feature importance should be assessed during prediction, followed by backwards elimination. For instance, Naidoo et al. (2022) demonstrated that techniques such as Recursive Feature Elimination (RFE) can be used to identify key predictor features or variables, ensuring only the most relevant features remain in the model and reducing collinearity, which can lead to overfitting. Lastly, the second aim of the study, to evaluate the

temporal variation of winter wheat crop bio-parameters and identify the features most influential in predicting these bio-parameters, was successfully achieved.

5.4. RF LAI and Cab Prediction Maps

The ultimate goal of creating the LAI and Cab or bio-variables prediction models is to gain perspective about what is happening in the real world, or on the ground. In this case, we are trying to find out what is happening in a farm plot. Thus, the third goal of the study was to apply the best-performing model to the multispectral remote sensing images to create estimation maps for Cab and LAI.

The RF estimated LAI and Cab maps show variation across the plot and across time; moreover, the estimated minimum and maximum LAI and Cab values vary over time (Figures 10 and 11). For example, the LAI and Cab levels seemed to increase across the field with the progression into the new growth stage, with peak LAI and Cab values observed during the early nodding stage to late middle growth stages (Figures 10 and 11). This might be because the size of the leaves and the pigments within crops, such as chlorophyll content, change over time (Han et al., 2021), resulting in varying LAI and Cab levels. Moreover, this might be because of the application of fertilisers and different irrigation treatments. For example, Adeluyi et al. (2021) found that LAI values changed with water and nitrogen application, which might explain the variation in LAI and Cab across time in the current study. Furthermore, as the winter wheat crop progresses to the later stages, the LAI and Cab values seem to decrease across the plot.

These results correspond with findings from Verma et al. (2022), even though they were focusing on a different crop, the Lentils and other crops; their findings showed the LAI and LCC increased during the nodding stage and started decreasing after the flowering stage, which was almost the same trend observed in the current study. Moreover, both the LAI and Cab prediction maps follow the same trend as the actual LAI and Cab measured in the field; the actual LAI and Cab values increase over time from early (nodding stage) to late stages and then start decreasing (Figures 7, 10, and 11). This similarity between the actual LAI and Cab measurements and the predicted LAI and Cab maps shows that the RF can be used to make inferences on the condition of crops from Sentinel-2 imagery, because it produced estimated results that were similar to the actual values. In addition, it also predicted results which followed the same trend as the actual values at different growth stages (Figures 7, 10 and 11).

Finally, the third goal of the study was successfully achieved as the best-performing model was ultimately used to create LAI and Cab prediction maps.

Chapter 6: Conclusions and Recommendations

6.1. Introduction

This part offers an overview of the conclusions that may be made from the study's findings, as well as an analysis of its limitations and suggestions for how future research projects might build on the findings.

6.2. Conclusion

The primary goal of the study was to examine the utility of freely accessible under open access Sentinel-2 images, along with various machine learning algorithms, to estimate crop LAI and Cab of winter wheat across the growing season. The first objective of the study, which was to use SVR and RF machine learning algorithms to estimate the LAI and Cab at different growth stages from Sentinel-2 multispectral imagery, was successfully achieved. The second objective of the study, which was to assess the variations in LAI and Cab at different growth stages, and to find the features that contributed the most or least to the LAI and Cab prediction models, was also successfully achieved. Furthermore, the final goal of the study, which was to apply the best-performing model to multispectral remote sensing images to create estimation maps for Cab and LAI, was successfully achieved.

The results show that LAI and Cab of winter wheat can be accurately estimated at the plot level across different growth stages using machine learning regression models, such as RF, along with Sentinel-2 imagery. This method delivers timely and relatively precise information on crop status, empowering farmers to make informed management decisions. For example, they can apply fertilisers and water based on actual need, rather than uniformly irrigating and fertilising the entire plot when it might not be necessary. However, the findings also highlight the importance of identifying which features contribute most or least to estimating bio-parameters, specifically LAI and Cab, as these vary at different growth stages of winter wheat. Furthermore, the study emphasises the necessity of properly tuning hyperparameters and selecting only the features that significantly impact the estimates, since redundancy and correlated features can lead to overfitting models that perform well on training data but fail to generalise on new data, as seen with SVR in this study.

6.3. Limitations

The limited number of samples was one of the major limitations of the study because all the samples were used for training purposes, and the k-fold cross-validation method was used for validation purposes. Moreover, the spatial resolution of the images used for prediction was one of the factors that might have affected the model accuracy; thus, a finer spatial resolution might improve the prediction accuracy even further, resulting in better results that might improve decision-making. Additionally, getting the imagery that corresponded with the collection dates of the in-situ samples was also a challenge, as some images were unavailable, resulting in the closest date with imagery available being used. Finally, the study area was relatively small and lacked sufficient complexity to fairly evaluate the model's performance and flexibility.

6.4. Recommendations

The first recommendation will be to increase the sample size when collecting the LAI and Cab from the field, thus allowing for increased accuracy and robust models, unlike in this study, where all the in-situ samples were used for training. Moreover, using multispectral UAVs with a similar bandwidth to the satellite, such as Sentinel-2, can be helpful. For example, these can then account for days when imagery is not available due to cloud cover or revisit time, and also provide red edge bands with finer spatial resolution.

References

Adeluyi, O., Harris, A., Verrelst, J., Foster, T., & Clay, G. D. (2021). Estimating the phenological dynamics of irrigated rice leaf area index using the combination of PROSAIL and Gaussian Process Regression. *International Journal of Applied Earth Observation and Geoinformation*, *102*. https://doi.org/10.1016/j.jag.2021.102454

Bajocco, S., Ferrara, C., Savian, F., Ginaldi, F., Puletti, N., Crecco, L., Bregaglio, S., & Chianucci, F. (2024). LAIr: an R package to estimate LAI from Normalized Difference Vegetation Index. *SoftwareX*, *27*. https://doi.org/10.1016/j.softx.2024.101776

Bhadra, S., Sagan, V., Sarkar, S., Braud, M., Mockler, T. C., & Eveland, A. L. (2024). PROSAIL-Net: A transfer learning-based dual stream neural network to estimate leaf chlorophyll and leaf angle of crops from UAV hyperspectral images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *210*, 1–24. https://doi.org/10.1016/j.isprsjprs.2024.02.020

Chemura, A., Mutanga, O., & Odindi, J. (2017). Empirical Modeling of Leaf Chlorophyll Content in Coffee (Coffea Arabica) Plantations with Sentinel-2 MSI Data: Effects of Spectral Settings, Spatial Resolution, and Crop Canopy Cover. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(12), 5541–5550. https://doi.org/10.1109/JSTARS.2017.2750325

Chen, Z., Jia, K., Wei, X., Liu, Y., Zhan, Y., Xia, M., Yao, Y., & Zhang, X. (2022). Improving leaf area index estimation accuracy of wheat by involving leaf chlorophyll content information. *Computers and Electronics in Agriculture*, *196*. https://doi.org/10.1016/j.compag.2022.106902

Cheng, Z., Gu, X., Du, Y., Zhou, Z., Li, W., Zheng, X., Cai, W., & Chang, T. (2024). Spectral purification improves monitoring accuracy of the comprehensive growth evaluation index for film-mulched winter wheat. *Journal of Integrative Agriculture*, *23*(5), 1523–1540. https://doi.org/10.1016/j.jia.2023.05.036

Choudhury, B. U., Webster, R., Sharma, V., Goswami, J., Meetei, T. T., Krishnappa, R., & Raju, P. L. N. (2019). Frost damage to maize in northeast India: assessment and estimated loss of yield by hyperspectral proximal remote sensing. *Journal of Applied Remote Sensing*, *13*(04), 1. https://doi.org/10.1117/1.jrs.13.044527

Couch, S., Kazan, Z., Shi, K., Bray, A., & Groce, A. (2018). *A Differentially Private Wilcoxon Signed-Rank Test*. http://arxiv.org/abs/1809.01635

Dipaleseng Local Municipality Local Economic Development Strategy Final Report Dipaleseng Local Economic Development Strategy. (2011).

Dipaleseng Local Municipality topographic map, elevation, terrain. (n.d.). Retrieved August 16, 2025, from https://en-gb.topographic-map.com/map-c9l6cz/Dipaleseng-Local-Municipality/?center=-27.11048%2C29.32251&zoom=9

Ekwe, M. C., Adeluyi, O., Verrelst, J., Kross, A., & Odiji, C. A. (2024). Estimating rainfed groundnut's leaf area index using Sentinel-2 based on Machine Learning Regression Algorithms and Empirical Models. *Precision Agriculture*, *25*(3), 1404–1428. https://doi.org/10.1007/s11119-024-10117-0

Gao, D., Qiao, L., An, L., Zhao, R., Sun, H., Li, M., Tang, W., & Wang, N. (2022). Estimation of spectral responses and chlorophyll based on growth stage effects explored by machine learning methods. *Crop Journal*, *10*(5), 1292–1302. https://doi.org/10.1016/j.cj.2022.07.011

Han, X., Wei, Z., Chen, H., Zhang, B., Li, Y., & Du, T. (2021). Inversion of Winter Wheat Growth Parameters and Yield Under Different Water Treatments Based on UAV Multispectral Remote Sensing. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.609876

Ienco, D., Interdonato, R., Gaetano, R., & Ho Tong Minh, D. (2019). Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, *158*, 11–22. https://doi.org/10.1016/j.isprsjprs.2019.09.016

Imran, H. A., Gianelle, D., Rocchini, D., Dalponte, M., Martín, M. P., Sakowska, K., Wohlfahrt, G., & Vescovo, L. (2020). VIS-NIR, red-edge and NIR-shoulder based normalized vegetation indices response to co-varying leaf and Canopy structural traits in heterogeneous grasslands. *Remote Sensing*, *12*(14). https://doi.org/10.3390/rs12142254

Jin, Z., Liu, H., Cao, H., Li, S., Yu, F., & Xu, T. (2025). Hyperspectral Remote Sensing Estimation of Rice Canopy LAI and LCC by UAV Coupled RTM and Machine Learning. *Agriculture (Switzerland)*, *15*(1). https://doi.org/10.3390/agriculture15010011

Kanning, M., Kühling, I., Trautz, D., & Jarmer, T. (2018). High-resolution UAV-based hyperspectral imagery for LAI and chlorophyll estimations from wheat for yield prediction. *Remote Sensing*, *10*(12). https://doi.org/10.3390/rs10122000

Kganyago, M., Adjorlolo, C., & Mhangara, P. (2024). Optimizing Sentinel-2 feature space for improved crop biophysical and biochemical variables retrieval using the novel spectral triad feature selection algorithm. *Geocarto International*, *39*(1). https://doi.org/10.1080/10106049.2024.2309174

Kganyago, M., Adjorlolo, C., Sibanda, M., Mhangara, P., Laneve, G., & Alexandridis, T. (2022). Testing Sentinel-2 spectral configurations for estimating relevant crop biophysical and biochemical parameters for precision agriculture using tree-based and kernel-based algorithms. *Geocarto International*. https://doi.org/10.1080/10106049.2022.2146764

Kganyago, M., Mhangara, P., & Adjorlolo, C. (2021). Estimating crop biophysical parameters using machine learning algorithms and sentinel-2 imagery. *Remote Sensing*, *13*(21). https://doi.org/10.3390/rs13214314

Křížová, K., Kadeřábek, J., Novák, V., Linda, R., Kurešová, G., & Šařec, P. (2022). Using a single-board computer as a low-cost instrument for SPAD value estimation through colour images and chlorophyll-related spectral indices. *Ecological Informatics*, *67*. https://doi.org/10.1016/j.ecoinf.2021.101496

Lambert, M. J., Traoré, P. C. S., Blaes, X., Baret, P., & Defourny, P. (2018). Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sensing of Environment*, *216*, 647–657. https://doi.org/10.1016/j.rse.2018.06.036

LI, H., CHEN, Z. xin, JIANG, Z. wei, WU, W. bin, REN, J. qiang, LIU, B., & Tuya, H. (2017). Comparative analysis of GF-1, HJ-1, and Landsat-8 data for estimating the leaf area index

of winter wheat. *Journal of Integrative Agriculture*, *16*(2), 266–285. https://doi.org/10.1016/S2095-3119(15)61293-X

Liu, Y., Liu, M., Liu, G., Sun, H., An, L., Zhao, R., Tang, W., Zhao, F., Yan, X., Ma, Y., & Li, M. (2024). Estimating canopy chlorophyll content of powdery mildew stressed winter wheat by different spatial resolutions of UAV-imagery. *Computers and Electronics in Agriculture*, *227*. https://doi.org/10.1016/j.compag.2024.109621

Madonsela, S., Cho, M. A., Naidoo, L., Main, R., & Majozi, N. (2023). Exploring the utility of Sentinel-2 for estimating maize chlorophyll content and leaf area index across different growth stages. *Journal of Spatial Science*, *68*(2), 339–351. https://doi.org/10.1080/14498596.2021.2000898

Mercier, A., Betbeder, J., Rumiano, F., Baudry, J., Gond, V., Blanc, L., Bourgoin, C., Cornu, G., Ciudad, C., Marchamalo, M., Poccard-Chapuis, R., & Hubert-Moy, L. (2019). Evaluation of Sentinel-1 and 2 Time Series for Land Cover Classification of Forest–Agriculture Mosaics in Temperate and Tropical Landscapes. *Remote Sensing*, *11*(8), 979. https://doi.org/10.3390/rs11080979

Naidoo, L., Main, R., Cho, M. A., Madonsela, S., & Majozi, N. (2022). Machine learning modelling of crop structure within the Maize Triangle of South Africa. *International Journal of Remote Sensing*, *43*(1), 27–51. https://doi.org/10.1080/01431161.2021.1998714

Peng, Y., Zhu, T., Li, Y., Dai, C., Fang, S., Gong, Y., Wu, X., Zhu, R., & Liu, K. (2019). Remote prediction of yield based on LAI estimation in oilseed rape under different planting methods and nitrogen fertilizer applications. *Agricultural and Forest Meteorology*, *271*, 116–125. https://doi.org/10.1016/j.agrformet.2019.02.032

PHASE 2: SPATIAL CHALLENGES AND OPPORTUNITIES REPORT Review of the Municipal Spatial Development Framework (MSDF) for Dipaleseng Local Municipality within a Period of 12 Months. (2019).

Revill, A., Florence, A., MacArthur, A., Hoad, S. P., Rees, R. M., & Williams, M. (2019). The value of Sentinel-2 spectral bands for the assessment of winter wheat growth and development. *Remote Sensing*, *11*(17). https://doi.org/10.3390/rs11172050

Risbøl, O., & Gustavsen, L. (2018). LiDAR from drones employed for mapping archaeology – Potential, benefits and challenges. *Archaeological Prospection*, *25*(4), 329–338. https://doi.org/10.1002/arp.1712

Sahoo, R. N., Rejith, R. G., Gakhar, S., Verrelst, J., Ranjan, R., Kondraju, T., Meena, M. C., Mukherjee, J., Dass, A., Kumar, S., Kumar, M., Dhandapani, R., & Chinnusamy, V. (2024). Estimation of wheat biophysical variables through UAV hyperspectral remote sensing using machine learning and radiative transfer models. *Computers and Electronics in Agriculture*, *221*. https://doi.org/10.1016/j.compag.2024.108942

Shahrooz, M., Talaeizadeh, A., & Alasty, A. (2020, November 23). Agricultural Spraying Drones: Advantages and Disadvantages. *2020 Virtual Symposium in Plant Omics Sciences, OMICAS 2020 - Conference Proceedings*. https://doi.org/10.1109/OMICAS52284.2020.9535527

Shen, J., Wang, Q., Zhao, M., Hu, J., Wang, J., Shu, M., Liu, Y., Guo, W., Qiao, H., Niu, Q., & Yue, J. (2024). Mapping Maize Planting Densities Using Unmanned Aerial Vehicles,

Multispectral Remote Sensing, and Deep Learning Technology. *Drones*, *8*(4). https://doi.org/10.3390/drones8040140

Shiferaw, B., Smale, M., Braun, H. J., Duveiller, E., Reynolds, M., & Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security*, *5*(3), 291–317. https://doi.org/10.1007/s12571-013-0263-y

Shoko, C., Mutanga, O., & Dube, T. (2020). Optimal season for discriminating C3 and C4 grass functional types using multi-date Sentinel 2 data. *GIScience and Remote Sensing*, *57*(1), 127–139. https://doi.org/10.1080/15481603.2019.1675286

Sibanda, M., Ndlovu, H. S., Brewer, K., Buthelezi, S., Matongera, T. N., Mutanga, O., Odidndi, J., Clulow, A. D., Chimonyo, V. G. P., & Mabhaudhi, T. (2023). Remote sensing hail damage on maize crops in smallholder farms using data acquired by remotely piloted aircraft system. *Smart Agricultural Technology*, *6*. https://doi.org/10.1016/j.atech.2023.100325

Singh, U., Kumar, N., & Chauhan, P. (n.d.). *Enhancing Autonomous Drones with Self-Designed Payload Dropping Mechanism*. www.ijfmr.com

Singhal, G., Choudhury, B. U., Singh, N., & Goswami, J. (2024). An enhanced chlorophyll estimation model with a canopy structural trait in maize crops: Use of multi-spectral UAV images and machine learning algorithm. *Ecological Informatics*, *83*. https://doi.org/10.1016/j.ecoinf.2024.102811

Srinet, R., Nandy, S., & Patel, N. R. (2019). Estimating leaf area index and light extinction coefficient using Random Forest regression algorithm in a tropical moist deciduous forest, India. *Ecological Informatics*, *52*, 94–102. https://doi.org/10.1016/j.ecoinf.2019.05.008

Tomíček, J., Mišurec, J., Lukeš, P., & Potůčková, M. (2022). Retrieval of Harmonized LAI Product of Agricultural Crops from Landsat OLI and Sentinel-2 MSI Time Series. *Agriculture (Switzerland)*, *12*(12). https://doi.org/10.3390/agriculture12122080

Verma, B., Prasad, R., Srivastava, P. K., Yadav, S. A., Singh, P., & Singh, R. K. (2022). Investigation of optimal vegetation indices for retrieval of leaf chlorophyll and leaf area index using enhanced learning algorithms. *Computers and Electronics in Agriculture*, *192*. https://doi.org/10.1016/j.compag.2021.106581

Wang, T., Gao, M., Cao, C., You, J., Zhang, X., & Shen, L. (2022). Winter wheat chlorophyll content retrieval based on machine learning using in situ hyperspectral data. *Computers and Electronics in Agriculture*, *193*. https://doi.org/10.1016/j.compag.2022.106728

Xiao, G., Zhang, X., Niu, Q., Li, X., Li, X., Zhong, L., & Huang, J. (2024). Winter wheat yield estimation at the field scale using sentinel-2 data and deep learning. *Computers and Electronics in Agriculture*, *216*. https://doi.org/10.1016/j.compag.2023.108555

Xie, Q., Dash, J., Huete, A., Jiang, A., Yin, G., Ding, Y., Peng, D., Hall, C. C., Brown, L., Shi, Y., Ye, H., Dong, Y., & Huang, W. (2019). Retrieval of crop biophysical parameters from Sentinel-2 remote sensing imagery. *International Journal of Applied Earth Observation and Geoinformation*, *80*, 187–195. https://doi.org/10.1016/j.jag.2019.04.019

Xu, J., Quackenbush, L. J., Volk, T. A., & Im, J. (2022). Estimation of shrub willow biophysical parameters across time and space from Sentinel-2 and unmanned aerial system (UAS) data. *Field Crops Research*, *287*. https://doi.org/10.1016/j.fcr.2022.108655

Yan, P., Feng, Y., Han, Q., Hu, Z., Huang, X., Su, K., & Kang, S. (2024). Enhanced cotton chlorophyll content estimation with UAV multispectral and LiDAR constrained SCOPE model. *International Journal of Applied Earth Observation and Geoinformation*, *132*. https://doi.org/10.1016/j.jag.2024.104052

Yao, X., Wang, N., Liu, Y., Cheng, T., Tian, Y., Chen, Q., & Zhu, Y. (2017). Estimation of wheat LAI at middle to high levels using unmanned aerial vehicle narrowband multispectral imagery. In *Remote Sensing* (Vol. 9, Issue 12). MDPI AG. https://doi.org/10.3390/rs9121304

Yuan, H., Yang, G., Li, C., Wang, Y., Liu, J., Yu, H., Feng, H., Xu, B., Zhao, X., & Yang, X. (2017). Retrieving soybean leaf area index from unmanned aerial vehicle hyperspectral remote sensing: Analysis of RF, ANN, and SVM regression models. *Remote Sensing*, *9*(4). https://doi.org/10.3390/rs9040309

Yue, J., Wang, J., Zhang, Z., Li, C., Yang, H., Feng, H., & Guo, W. (2024). Estimating crop leaf area index and chlorophyll content using a deep learning-based hyperspectral analysis method. *Computers and Electronics in Agriculture*, *227*, 109653. https://doi.org/10.1016/j.compag.2024.109653

Zama, S. B. (2014). *POLICY & PRACTICE Service Delivery: Focus on Dipaleseng Local Municipality: Mpumalanga*. http://epress.lib.uts.edu.au/ojs/index.php/cjlg

Zhang, B., Gu, L., Dai, M., Bao, X., Sun, Q., Zhang, M., Qu, X., Li, Z., Zhen, W., & Gu, X. (2024). Estimation of grain filling rate of winter wheat using leaf chlorophyll and LAI extracted from UAV images. *Field Crops Research*, *306*. https://doi.org/10.1016/j.fcr.2023.109198

Zhang, C., Yi, Y., Wang, L., Zhang, X., Chen, S., Su, Z., Zhang, S., & Xue, Y. (2024). Estimation of the Bio-Parameters of Winter Wheat by Combining Feature Selection with Machine Learning Using Multi-Temporal Unmanned Aerial Vehicle Multispectral Images. *Remote Sensing*, *16*(3). https://doi.org/10.3390/rs16030469

Zhang, H., Zhang, Y., Liu, K., Lan, S., Gao, T., & Li, M. (2023). Winter wheat yield prediction using integrated Landsat 8 and Sentinel-2 vegetation index time-series data and machine learning algorithms. *Computers and Electronics in Agriculture*, *213*. https://doi.org/10.1016/j.compag.2023.108250

Zhang, L., Wang, A., Zhang, H., Zhu, Q., Zhang, H., Sun, W., & Niu, Y. (2024). Estimating Leaf Chlorophyll Content of Winter Wheat from UAV Multispectral Images Using Machine Learning Algorithms under Different Species, Growth Stages, and Nitrogen Stress Conditions. *Agriculture (Switzerland)*, *14*(7). https://doi.org/10.3390/agriculture14071064

Zhou, X., Zhang, J., Chen, D., Huang, Y., Kong, W., Yuan, L., Ye, H., & Huang, W. (2020). Assessment of leaf chlorophyll content models for winter wheat using Landsat-8 multispectral remote sensing data. *Remote Sensing*, *12*(16). https://doi.org/10.3390/RS12162574

Zhu, W., Sun, Z., Yang, T., Li, J., Peng, J., Zhu, K., Li, S., Gong, H., Lyu, Y., Li, B., & Liao, X. (2020). Estimating leaf chlorophyll content of crops via optimal unmanned aerial vehicle

hyperspectral data at multi-scales. *Computers and Electronics in Agriculture*, *178*. https://doi.org/10.1016/j.compag.2020.105786